



19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 341 217**

51 Int. Cl.:  
**G06F 19/00** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **00909923 .5**  
96 Fecha de presentación : **18.01.2000**  
97 Número de publicación de la solicitud: **1072010**  
97 Fecha de publicación de la solicitud: **31.01.2001**

54 Título: **Recombinación de ácidos nucleicos mediada por oligonucleótidos.**

30 Prioridad: **19.01.1999 US 116447 P**  
**05.02.1999 US 118813 P**  
**05.02.1999 US 118854 P**  
**24.06.1999 US 141049 P**  
**28.09.1999 US 408392**  
**28.09.1999 US 408393**  
**12.10.1999 US 416375**  
**12.10.1999 US 416837**

45 Fecha de publicación de la mención BOPI:  
**17.06.2010**

45 Fecha de la publicación del folleto de la patente:  
**17.06.2010**

73 Titular/es: **Maxygen, Inc.**  
**515, Galveston Drive**  
**Redwood City, California 94063, US**

72 Inventor/es: **Crameri, Andreas;**  
**Stemmer, Willem, P., C.;**  
**Minshull, Jeremy;**  
**Bass, Steven, H.;**  
**Welch, Mark;**  
**Ness, Jon, E.;**  
**Gustafsson, Claes y**  
**Patten, Phillip, A.**

74 Agente: **Isern Jara, Jorge**

ES 2 341 217 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

Recombinación de ácidos nucleicos mediada por oligonucleótidos.

5 **Antecedentes de la invención**

La trasposición de secuencias de ADN ha proporcionado un cambio de paradigma en la generación, manipulación y selección de ácidos nucleicos recombinantes. Los inventores y sus colaboradores han desarrollado metodologías de evolución artificial rápida para generar genes industriales, agrícolas y terapéuticos, y proteínas codificadas por los mismos, mejorados. Estos métodos, y composiciones relacionadas y un aparato para la práctica de estos métodos representan un trabajo pionero de los inventores y colaboradores.

Varias publicaciones de los inventores y sus colaboradores describen la trasposición de secuencias de ADN. Por ejemplo, Stemmer *et al.*, “Rapid Evolution of a Protein”, Nature 370:389-391, 1994; Stemmer, “DNA Shuffling by Random Fragmentation and Reassembly: *in vitro* Recombination for Molecular Evolution”, Proc. Natl. Acad. Sci. USA 91:10747-10751, 1994; Stemmer, patente US n° 5.603.793, titulada “Methods for *in vitro* recombination”; Stemmer *et al.*, patente US n° 5.830.721, titulada “DNA mutagenesis by random fragmentation and reassembly”; Stemmer *et al.*, patente US n° 5.811.238, titulada “Methods for generating polynucleotides having desired characteristics by iterative selection and recombination”, describe, por ejemplo, la trasposición *in vitro* e *in vivo* de ácidos nucleicos, ADN y proteínas, en una diversidad de formatos, por ejemplo mediante ciclos repetidos de mutagénesis, trasposición de secuencias y selección, así como métodos para generar bibliotecas de péptidos y anticuerpos expresados.

Los inventores y sus colaboradores también han desarrollado aplicaciones de la tecnología de trasposición de secuencias de ADN. Además de las publicaciones indicadas anteriormente, Minshull *et al.*, patente US n° 5.837.458, titulada “Methods and compositions for cellular and metabolic engineering” proporciona, por ejemplo, la evolución de rutas metabólicas y la potenciación del bioprocesamiento mediante técnicas de trasposición recursiva de secuencias. Cramer *et al.*, “Construction And Evolution Of Antibody-Phage Libraries By DNA Shuffling”, Nature Medicine 2 (1):100-103, 1996, describen, por ejemplo, la trasposición de anticuerpos para bibliotecas fágicas de anticuerpos. Pueden encontrarse detalles adicionales sobre la trasposición de secuencias de ADN en las patentes WO n° 95/22625, n° 97/20078, n° 96/33207, n° 97/33957, n° 98/27230, n° 97/35966, n° 98/31837, n° 98/13487, n° 98/13485 y n° 98/42832, así como varias otras publicaciones de los inventores y sus colaboradores.

Varias publicaciones de los inventores y sus colaboradores, así como de otros investigadores de la técnica también describen técnicas que facilitan la trasposición de ADN, por ejemplo que permiten el reensamblaje de genes a partir de fragmentos pequeños, o incluso de oligonucleótidos. Por ejemplo, además de las publicaciones anteriormente indicadas, Stemmer *et al.*, la patente US n° 5.834.252, titulada “End complementary polymerase reaction”, 1998, describe procedimientos para amplificación y detectar una secuencia diana (por ejemplo en una mezcla de ácidos nucleicos), así como para ensamblar grandes polinucleótidos a partir de fragmentos de ácidos nucleicos.

Una revisión de las publicaciones anteriormente indicadas revela que la evolución forzada mediante trasposición génica es una importante técnica nueva con muchas aplicaciones prácticas y potentes. De esta manera, resultan altamente deseables nuevas técnicas que faciliten la trasposición génica. La presente invención proporciona nuevos protocolos significativos de trasposición génica, así como muchas otras características que resultarán evidentes tras la revisión completa de la presente exposición.

45 **Descripción resumida de la invención**

La invención proporciona un método de recombinación de ácidos nucleicos homólogos, comprendiendo el método:

- 50 (i) alinear secuencias de ácidos nucleicos diana homólogos utilizando programas informáticos para seleccionar regiones conservadas de identidad de secuencia y regiones de diversidad de secuencia,
- 55 (ii) sintetizar un conjunto de oligonucleótidos solapantes de trasposición génica familiar, en donde el conjunto de oligonucleótidos corresponde a por lo menos una región de diversidad de secuencias en los ácidos nucleicos diana, y en donde el conjunto de oligonucleótidos corresponde colectivamente a 50% o más de la longitud completa de cada uno de los ácidos nucleicos diana homólogos,
- (iii) hibridar el conjunto de oligonucleótidos de trasposición génica familiar,
- 60 (iv) elongar el conjunto de oligonucleótidos de trasposición génica familiar, proporcionando de esta manera una población de ácidos nucleicos recombinados,
- (v) desnaturalizar la población de ácidos nucleicos recombinados, proporcionando de esta manera ácidos nucleicos recombinados desnaturalizados,
- 65 (vi) hibridar nuevamente los ácidos nucleicos desnaturalizados,

- (vii) extender los ácidos nucleicos recombinados nuevamente hibridados resultantes utilizando una polimerasa,
- (viii) repetir las etapas (v) a (vii) por lo menos una vez, y

- 5 (ix) seleccionar uno o más de los ácidos nucleicos recombinados nuevamente hibridados resultantes para un carácter o propiedad deseada,

en donde se utilizan proporciones no equimolares de oligonucleótidos de trasposición familiar para sesgar la re-combinación.

10

Se proporcionan otros aspectos de la invención en las reivindicaciones adjuntas.

### Descripción resumida de la exposición

15 Se dan a conocer en la presente memoria la trasposición de ácidos nucleicos asistida por oligonucleótidos. Estos enfoques asistidos por oligonucleótidos facilitan particularmente los procedimientos de trasposición génica familiar, proporcionando protocolos de trasposición sustancialmente simplificados que pueden utilizarse para producir ácidos nucleicos de trasposición familiar sin aislar ni clonar los ácidos nucleicos homólogos de longitud completa. Además, los enfoques asistidos por oligonucleótidos incluso pueden extenderse a ácidos nucleicos de trasposición no homólogos, accediendo de esta manera a un mayor espacio de secuencia en las moléculas recombinantes resultantes y, de esta manera, a una mayor diversidad molecular. Las técnicas también pueden combinarse con protocolos clásicos de trasposición de ADN, tales como los métodos mediados por ADNasa, o con otros procedimientos de generación de diversidad, tales como la mutagénesis clásica, para incrementar la versatilidad y rendimiento de estos métodos.

25 Se dan a conocer varios métodos que resultan aplicables a procedimientos de trasposición familiar. En los métodos reivindicados, se hibridan y elongan conjuntos de oligonucleótidos solapantes de trasposición génica familiar, proporcionando una población de ácidos nucleicos recombinados que pueden seleccionarse para un carácter o propiedad deseada. Típicamente el conjunto de oligonucleótidos solapantes de trasposición génica familiar incluye una pluralidad de tipos de oligonucleótido-miembro que presentan subsecuencias de región de consenso derivadas de una pluralidad de ácidos nucleicos diana homólogos. Los conjuntos de oligos opcionalmente proporcionan otras características distintivas, incluyendo la capacidad de entrecruzamiento, la modificación o la selección de codones, y similares.

30 La población de ácidos nucleicos recombinados puede desnaturalizarse e hibridarse nuevamente, proporcionando ácidos nucleicos recombinados desnaturalizados que después pueden hibridarse nuevamente. Los ácidos nucleicos recombinante resultantes también pueden seleccionarse. Cualquiera o la totalidad de dichas etapas puede repetirse reiteradamente, permitiendo múltiples sucesos de recombinación y selección para producir un ácido nucleico con un carácter o propiedad deseada.

40 En un aspecto relacionado de la exposición, los métodos para introducir diversidad de ácidos nucleicos familiar durante la recombinación de ácidos nucleicos se llevan a cabo proporcionando una composición que presenta por lo menos un conjunto de fragmentos de ácido nucleico que incluye una población de oligonucleótidos de trasposición génica familiar y que recombina por lo menos uno de los ácidos nucleicos fragmentados con por lo menos uno de los oligonucleótidos de trasposición génica familiar. A continuación, se regenera un ácido nucleico recombinante que presenta una subsecuencia de ácidos nucleicos correspondiente a dicho por lo menos un oligonucleótido de trasposición génica familiar, típicamente codificante de una molécula de longitud completa (por ejemplo una proteína de longitud completa).

50 En los métodos reivindicados, los oligonucleótidos de trasposición génica familiar se proporcionan mediante la alineación de secuencias de ácidos nucleicos homólogos para seleccionar regiones conservadas de identidad de secuencia y regiones de diversidad de secuencia. Se sintetiza (en serie o en paralelo) una pluralidad de oligonucleótidos de trasposición génica familiar que corresponden a por lo menos una región de diversidad de secuencias. En contraste, en otros métodos se proporcionan conjuntos de fragmentos mediante el corte de uno o más ácidos nucleicos homólogos (por ejemplo con una ADNasa) o mediante la síntesis de un conjunto de oligonucleótidos correspondiente a una pluralidad de regiones de por lo menos un ácido nucleico (típicamente se proporcionan oligonucleótidos correspondientes un ácido nucleico de longitud completa como miembros de un conjunto de fragmentos de ácido nucleico). En los procedimientos de trasposición dados a conocer en la presente memoria, dichos fragmentos cortados pueden utilizarse conjuntamente con oligonucleótidos de trasposición génica familiar, por ejemplo en una o más reacciones de recombinación.

60 Se dan a conocer métodos recursivos de trasposición de oligonucleótidos. Tal como se indica en la presente memoria, los ácidos nucleicos recombinantes generados sintéticamente utilizando oligonucleótidos pueden cortarse y trasponerse mediante metodologías estándares de trasposición de ácidos nucleicos, o los ácidos nucleicos pueden secuenciarse y utilizarse para diseñar un segundo conjunto de oligonucleótidos de trasposición familiar que se utiliza para recombinar los ácidos nucleicos recombinantes. Puede utilizarse una cualquiera de estas técnicas recursivas, o ambas, para rondas posteriores de recombinación, y también pueden utilizarse conjuntamente con rondas de selección de productos recombinantes. Las etapas de selección pueden realizarse tras una o varias rondas de recombinación, dependiendo de la diversidad deseada de los ácidos nucleicos recombinantes (cuantas más rondas de recombinación se lleven a cabo, más diversa será la población resultante de ácidos nucleicos recombinantes).

## ES 2 341 217 T3

La utilización de oligonucleótidos de trasposición génica familiar en reacciones de recombinación en la presente memoria permite el cambio de dominios de identidad o de diversidad de secuencias entre ácidos nucleicos homólogos, por ejemplo en el caso de que los recombinantes resultantes de la reacción de recombinación proporcionen ácidos nucleicos recombinantes con un dominio de secuencia de un primer ácido nucleico incluido dentro de una secuencia correspondiente a un segundo ácido nucleico, por ejemplo en el caso de que la región más similar a la región incluida en el segundo ácido nucleico no se encuentra presente en el ácido nucleico recombinante.

También se da a conocer una diversidad de composiciones para la práctica de los métodos anteriormente indicados y que resultan de la práctica de los mismos. Un ejemplo son las composiciones que incluyen una biblioteca de oligonucleótidos que presenta una pluralidad de tipos de oligonucleótido-miembro. La biblioteca puede incluir por lo menos aproximadamente 2, 3, 5, 10, 20, 30, 40, 50, 100 ó más oligonucleótidos-miembro diferentes. Los tipos de oligonucleótido-miembro corresponden a una pluralidad de regiones subsecuencia de una pluralidad de miembros pertenecientes a un conjunto seleccionado de entre una pluralidad de secuencias diana homólogas. Entre la pluralidad de regiones subsecuencias puede incluirse, por ejemplo, una pluralidad de regiones de secuencia solapante o no solapante perteneciente al conjunto seleccionado de secuencias diana homólogas. Cada uno de los tipos de oligonucleótido-miembro típicamente presenta una secuencia idéntica a por lo menos una subsecuencia de por lo menos uno de los conjuntos seleccionados de secuencias diana homólogas. Cualquiera de los tipos y conjuntos de oligonucleótidos indicados anteriormente, o en otros sitios en la presente memoria, puede incluirse en las composiciones dadas a conocer en la presente memoria (por ejemplo oligonucleótidos de trasposición familiar, oligonucleótidos entrecruzados, oligonucleótidos de cambio de dominio, etc.). Entre los tipos de oligonucleótido miembro se puede incluir una pluralidad de oligonucleótidos homólogos correspondientes una región homóloga de la pluralidad de secuencias diana homólogas. En este aspecto, cada uno de entre la pluralidad de oligonucleótidos homólogos presenta por lo menos una subsecuencia variante. Las bibliotecas de ácidos nucleicos y proteínas codificadas que resultan de la práctica de la recombinación mediada por oligonucleótidos tal como se indica en la presente memoria también son una característica de la exposición.

Las composiciones opcionalmente incluyen componentes que facilitan las reacciones de recombinación, por ejemplo una polimerasa, tal como una ADN polimerasa termoestable (por ejemplo *taq*, *vent* o cualquiera de las muchas otras polimerasas disponibles comercialmente), una recombinasa, un reactivo de síntesis de ácidos nucleicos, tampones, sales, magnesio, uno o más ácidos nucleicos que presenta uno o más de entre la pluralidad de miembros del conjunto seleccionado de secuencias diana homólogas, y similares.

También se dan a conocer kits que comprenden las composiciones dadas a conocer en la presente memoria, por ejemplo en recipientes o en otros materiales de empaquetamiento, por ejemplo con materiales instruccionales para la práctica de los métodos de la invención. También se dan a conocer en la presente invención usos para las composiciones y kits para la práctica de los métodos.

### Breve descripción de las figuras

La figura 1 es un esquema que muestra la trasposición *in vivo* dirigida por oligonucleótidos utilizando quimeroplastos.

La figura 2 es un esquema de un procedimiento de trasposición de baja homología que proporciona la mezcla génica sintética.

La figura 3 es un esquema de una biblioteca modular por delección/inserción de exones.

### Definiciones

A menos que se indique lo contrario, las definiciones siguientes suplementan aquéllas de la técnica.

Los ácidos nucleicos son “homólogos” en el caso de que sean derivados, natural o artificialmente, de una secuencia ancestral común. Durante la evolución natural, lo anterior se produce cuando dos o más secuencias descendientes divergen de una secuencia parental a lo largo del tiempo, es decir, debido a mutación y selección natural. Bajo condiciones artificiales, se produce divergencia, por ejemplo en una de entre dos maneras básicas. En primer lugar, una secuencia dada puede recombinarse artificialmente con otra secuencia, tal como ocurre, por ejemplo durante la clonación típica, para producir un ácido nucleico descendiente, o una secuencia dada puede modificarse químicamente, o manipularse de otro modo para modificar la molécula resultante. Alternativamente, puede sintetizarse un ácido nucleico *de novo*, mediante la síntesis de un ácido nucleico que varíe en secuencia respecto a una secuencia de ácidos nucleicos parental seleccionada. En el caso de que no se disponga de conocimiento explícito de los ancestros de dos ácidos nucleicos, la homología típicamente se infiere mediante comparación de las dos secuencias. En el caso de que dos secuencias de ácidos nucleicos muestren similitud de secuencias en una parte significativa de cada uno de los ácidos nucleicos, se infiere que los dos ácidos nucleicos comparten un ancestro común. El nivel exacto de similitud de secuencia que establece la homología varía en la técnica dependiendo de una diversidad de factores. Para los fines de la presente invención, los intermediarios cladísticos (secuencias propuestas que comparten características de dos o más ácidos nucleicos relacionados) son ácidos nucleicos homólogos.

## ES 2 341 217 T3

Para los fines de la presente exposición, dos ácidos nucleicos se consideran homólogos si comparten suficiente identidad de secuencia para permitir que se produzca la recombinación directa entre las dos moléculas de ácido nucleico. Típicamente los ácidos nucleicos utilizan regiones de estrecha similitud espaciadas aproximadamente la misma distancia para permitir que se produzca la recombinación. La recombinación puede ser *in vitro* o *in vivo*.

5

Sin embargo, debe apreciarse que una ventaja de determinadas características de la exposición es la capacidad de recombinar ácidos nucleicos relacionados entre sí en un nivel más distante que el permitido por las técnicas estándares de recombinación. En particular, pueden recombinarse secuencias de dos ácidos nucleicos relacionados distantemente, o incluso que no se encuentran detectablemente relacionados, utilizando oligonucleótidos entrecruzados que presentan subsecuencias de dos o más ácidos nucleicos diana no homólogos diferentes, o dos o más ácidos nucleicos relacionados distantemente. Sin embargo, en el caso de que los dos ácidos nucleicos sólo puedan recombinarse indirectamente, utilizando oligonucleótidos intermediarios tal como se proporciona en la presente memoria, se consideran “no homólogos” para los fines de la presente exposición.

15 El término “conjunto” tal como se utiliza en la presente memoria se refiere a una colección de por lo menos dos tipos de molécula, y típicamente incluye por lo menos aproximadamente, por ejemplo, 5, 10, 50, 100, 500, 1.000 ó más miembros, dependiendo del uso pretendido exacto del conjunto.

20 Un conjunto de “oligonucleótidos de trasposición génica familiar” es un conjunto de oligonucleótidos sintetizados derivados de un conjunto seleccionado de ácidos nucleicos homólogos. Los oligonucleótidos se derivan de un conjunto seleccionado de ácidos nucleicos homólogos en el caso de que presenten (individual o colectivamente) regiones de identidad de secuencia (y, opcionalmente, regiones de diversidad de secuencia) con más de un ácido nucleico homólogo. Colectivamente, los oligonucleótidos típicamente corresponden a una parte sustancial de la longitud completa de los ácidos nucleicos homólogos del conjunto de ácidos nucleicos homólogos, por ejemplo los oligonucleótidos se corresponden en una parte sustancial de la longitud de los ácidos nucleicos homólogos (por ejemplo, los oligonucleótidos del conjunto corresponden colectivamente a, por ejemplo, 25% o más, con frecuencia 35% o más, generalmente 50% o más, típicamente 60% o más, más típicamente 70% o más, y en algunas aplicaciones, 80%, 90% o 100% de la longitud completa de cada uno de los ácidos nucleicos homólogos). Más comúnmente, entre los oligonucleótidos de trasposición génica familiar se incluyen múltiples tipos de miembro, presentando cada uno regiones de identidad de secuencia con por lo menos un miembro del conjunto seleccionado de ácidos nucleicos homólogos (por ejemplo aproximadamente 2, 3, 5, 10, 50 ó más tipos de miembro).

30 Un oligonucleótido “entrecruzado” presenta regiones de identidad de secuencia con por lo menos dos miembros diferentes de un conjunto seleccionado de ácidos nucleicos, que opcionalmente son homólogos o no homólogos.

35

Los ácidos nucleicos se “hibridan” cuando se asocian, típicamente en solución. Los ácidos nucleicos se hibridan debido a una diversidad de fuerzas físicoquímicas caracterizadas, tales como los puentes de hidrógeno, la exclusión de solvente, el apilamiento de bases y similares. Se encuentra una guía extensa de la hibridación de ácidos nucleicos en Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology - Hybridization with Nucleic Acid Probes, parte 1, capítulo 2, “Overview of principles of hybridization and the strategy of nucleic acid probe assays”, Elsevier, New York, 1993, así como en Ausubel, *supra*.

40

Dos ácidos nucleicos “se corresponden” en el caso de que presentan la misma secuencia o secuencias complementarias, o en el caso de que un ácido nucleico sea una subsecuencia de la otra, o en el caso de que una secuencia se derive, mediante manipulación natural o artificial, de la otra.

45

Los ácidos nucleicos se “elongan” cuando se incorporan nucleótidos (u otras moléculas análogas) adicionales al ácido nucleico. Más comúnmente, lo anterior se lleva a cabo con una polimerasa (por ejemplo una ADN polimerasa), por ejemplo una polimerasa que añade secuencias en el extremo 3' terminal del ácido nucleico.

50

Dos ácidos nucleicos se han “recombinado” cuando secuencias de cada uno de los dos ácidos nucleicos se han combinado en un ácido nucleico progenie. Dos secuencias se han recombinado “directamente” cuando ambos ácidos nucleicos son sustratos para la recombinación. Dos secuencias se han “recombinado indirectamente” cuando las secuencias se han recombinado utilizando un intermediario, tal como un oligonucleótido entrecruzado. Para la recombinación indirecta, no más de una de las secuencias es un sustrato real para la recombinación, y en algunos casos, ninguna de las secuencias es un sustrato para la recombinación (es decir, el proceso en el que uno o más oligonucleótidos de secuencia correspondiente a los ácidos nucleicos se hibridan y se elongan).

55

Una colección de “ácidos nucleicos fragmentados” es una colección de ácidos nucleicos derivados mediante corte de uno o más ácidos nucleicos parentales (por ejemplo con una nucleasa, o mediante corte químico) o mediante la producción de subsecuencias de las secuencias parentales de cualquier otra manera, tal como la elongación parcial de cadena de un ácido nucleico complementario.

60

Una “proteína de longitud completa” es una proteína que presenta sustancialmente los mismos dominios de secuencia que la proteína correspondiente codificada por un gen natural. La proteína puede presentar secuencias modificadas respecto al gen correspondiente naturalmente codificado (por ejemplo debido a recombinación y selección), pero presenta una longitud de por lo menos 95% de la del gen naturalmente codificado.

65

## ES 2 341 217 T3

Un “enzima ADNasa” es un enzima tal como la ADNasa I que cataliza el corte de un ADN, *in vitro* o *in vivo*. Una amplia diversidad de enzimas ADNasa son bien conocidos y se encuentran perfectamente descritos en, por ejemplo, Sambrook, Berger y Ausubel (todos *supra*) y muchos se encuentran disponibles comercialmente.

5 Un “dominio de ácidos nucleicos” es una región o subsecuencia de ácidos nucleicos. El dominio puede encontrarse conservado o no conservado entre una pluralidad de ácidos nucleicos homólogos. Típicamente, se delinea un dominio mediante la comparación entre dos o más secuencias, es decir, una región de diversidad de secuencia entre secuencias es un “dominio de diversidad de secuencia”, mientras que una región de similitud es un “dominio de similitud de secuencia”. El “cambio de dominio” se refiere a la capacidad de intercambiar una región ácido nucleico de un ácido nucleico por un segundo dominio de un segundo ácido nucleico.

10 Una región de “elevada similitud de secuencia” se refiere a una región que es 90% o más idéntica a una segunda región seleccionada tras el alineamiento para la máxima correspondencia (por ejemplo manualmente o utilizando el programa común BLAST ajustado a los parámetros por defecto). Una región de “baja similitud de secuencia” es 60% o menos idéntica, más preferentemente 40% o menos idéntica a una segunda región seleccionada, tras el alineamiento para la máxima correspondencia (por ejemplo manualmente o utilizando BLAST ajustado a los parámetros por defecto).

20 Un “amplión de PCR” es un ácido nucleico construido utilizando la reacción en cadena de la polimerasa (PCR). Típicamente, el ácido nucleico es una copia de un ácido nucleico seleccionado. Un “cebador de PCR” es un ácido nucleico que se hibrida a un ácido nucleico molde y permite la elongación de la cadena utilizando una polimerasa termoestable bajo condiciones de reacción apropiadas.

25 Una “biblioteca de oligonucleótidos” es un conjunto de oligonucleótidos. El conjunto puede encontrarse agrupado, o puede ser individualmente accesible. Los oligonucleótidos pueden ser de ADN, de ARN o combinaciones de ARN y ADN (por ejemplo quimeroplastos).

### *Comentario detallado*

30 La presente exposición se refiere a formatos mejorados de trasposición de ácidos nucleicos. En particular, mediante la utilización de conjuntos seleccionados de oligonucleótidos como sustratos para la recombinación y/o síntesis génica, resulta posible acelerar drásticamente el procedimiento de trasposición. Además, resulta posible utilizar oligonucleótidos intermediarios para recombinar indirectamente ácidos nucleicos que de otra manera no se recombinarían. No resulta necesario acceder directamente a los ácidos nucleicos físicos correspondientes a las secuencias que deben combinarse, debido a que las secuencias pueden recombinarse indirectamente mediante oligonucleótidos intermediarios.

40 Brevemente, en primer lugar se alinea una familia de secuencias de ácidos nucleicos homólogos, por ejemplo utilizando programas informáticos disponibles, con el fin de seleccionar regiones de identidad/similitud y regiones de diversidad. Se sintetiza una pluralidad (por ejemplo 2, 5, 10, 20, 50, 75 ó 100 ó más) de oligonucleótidos correspondientes a por lo menos una región de diversidad (y ordinariamente a por lo menos una región de similitud). Estos oligonucleótidos pueden trasponerse directamente, o pueden recombinarse con uno o más de la familia de ácidos nucleicos.

45 Esta recombinación basada en oligonucleótidos de ácidos nucleicos relacionados puede combinarse con varios métodos estándares de trasposición disponibles. Por ejemplo, existen varios procedimientos ahora disponibles para trasponer ácidos nucleicos homólogos, por ejemplo mediante la digestión de ácidos nucleicos con ADNasa, dejando que se produzca la recombinación y después regenerando moldes de longitud completa, por ejemplo tal como se describe en Stemmer, DNA mutagenesis by random fragmentation and reassembly, 1998, patente US nº 5.830.721. De esta manera, en un aspecto de la exposición, se proporciona un ácido nucleico de longitud completa que es idéntico, o por lo menos homólogo, a por lo menos uno de los ácidos nucleicos homólogos, se corta con una ADNasa, y el conjunto resultante de fragmentos de ácidos nucleicos se recombina con la pluralidad de oligonucleótidos de trasposición génica familiar. Esta combinación de métodos puede resultar ventajosa, debido a que los fragmentos cortados por ADNasa forman un “andamiaje” que puede reconstituirse para formar una secuencia de longitud completa; una ventaja en el caso de que uno o más oligos sintetizados en el conjunto sintetizado sea defectuoso.

60 Sin embargo, una ventaja de la presente invención es la capacidad de recombinar varias regiones de diversidad entre ácidos nucleicos homólogos, incluso sin que los ácidos nucleicos homólogos, o los fragmentos cortados de los mismos, se encuentren presentes en la mezcla de recombinación. Los ácidos nucleicos traspuestos resultantes pueden incluir regiones de diversidad procedentes de diferentes ácidos nucleicos, proporcionando la capacidad de combinar diferentes dominios de diversidad en un único ácido nucleico. Esto proporciona un método muy potente para acceder a la diversidad de secuencia natural.

65 En general, los métodos en la presente memoria permiten la “trasposición mediada por oligonucleótidos”, en la que oligonucleótidos correspondientes a una familia de ácidos nucleicos homólogos relacionados se recombinan para producir ácidos nucleicos seleccionables. La técnica puede utilizarse para recombinar secuencias de ácidos nucleicos homólogos e incluso no homólogos. Al recombinar ácidos nucleicos homólogos, se hibridan y elongan (mediante PCR de reensamblaje) conjuntos de oligonucleótidos solapantes de trasposición génica familiar (que se derivan, por

ejemplo, mediante comparación entre ácidos nucleicos homólogos y la síntesis de fragmentos de oligonucleótido), proporcionando una población de ácidos nucleicos recombinados, que pueden seleccionarse para un carácter o propiedad deseado. Típicamente, el conjunto de oligonucleótidos solapantes de trasposición génica familiar incluye una pluralidad de tipos de miembro oligonucleótido que presenta subsecuencias de región de consenso derivadas de una pluralidad de ácidos nucleicos diana homólogos.

Típicamente, se proporciona la trasposición génica de oligonucleótidos familiar mediante la alineación de secuencias homólogas de ácidos nucleicos para seleccionar las regiones conservadas de identidad de secuencia y las regiones de diversidad de secuencia. Se sintetiza (en serie o en paralelo) una pluralidad de oligonucleótidos de trasposición génica familiar que corresponde a por lo menos una región de diversidad de secuencia.

Pueden proporcionarse parcialmente conjuntos de fragmentos, o subconjuntos de fragmentos utilizados en enfoques de trasposición de oligonucleótidos, mediante el corte de uno o más ácidos nucleicos homólogos (por ejemplo con una ADNasa), así como mediante la síntesis de un conjunto de oligonucleótidos correspondiente a una pluralidad de regiones de por lo menos un ácido nucleico (típicamente se proporcionan oligonucleótidos correspondientes a un ácido nucleico de longitud parcial o completa como miembros del conjunto de "fragmentos" de ácido nucleico, un término que comprende tanto los fragmentos cortados como los oligonucleótidos sintetizados). Durante los procedimientos de trasposición en la presente memoria, dichos fragmentos cortados pueden utilizarse conjuntamente con oligonucleótidos de trasposición génica familiar, por ejemplo en una o más reacciones de recombinación, produciendo ácidos nucleicos recombinantes.

A continuación se proporcionan detalles y ejemplos referentes a la alineación de secuencias, la construcción de oligonucleótidos y la generación de bibliotecas, procedimientos de trasposición y otros aspectos de la presente exposición.

*Alineación de secuencias homólogas de ácidos nucleicos para seleccionar regiones conservadas de identidad de secuencia y regiones de diversidad de secuencia*

La invención permite la alineación de secuencias de ácidos nucleicos para determinar las regiones de identidad o similitud de secuencia y las regiones de diversidad. El conjunto de oligonucleótidos solapantes de trasposición génica familiar puede comprender una pluralidad de tipos de miembro oligonucleótido que comprende subsecuencias de región de consenso derivadas de una pluralidad de ácidos nucleicos diana homólogos. Estas subsecuencias de región de consenso se determinan mediante la alineación de ácidos nucleicos homólogos y la identificación de regiones de identidad o similitud.

En una realización, se alinean secuencias homólogas de ácidos nucleicos, y se selecciona por lo menos una región conservada de identidad de secuencia y una pluralidad de regiones de diversidad de secuencia. La pluralidad de regiones de diversidad de secuencia proporciona una pluralidad de dominios de diversidad de secuencia. Típicamente, se sintetiza una pluralidad de oligonucleótidos de trasposición génica familiar correspondiente a la pluralidad de dominios de diversidad de secuencia y se utiliza en los diversos protocolos de recombinación indicados en la presente memoria o que se encuentran disponibles de otra manera. Los genes sintetizados mediante estos métodos de recombinación opcionalmente se criban adicionalmente o se diversifican adicionalmente mediante cualquier método disponible, incluyendo la recombinación y/o la mutagénesis.

*Alineación de ácidos nucleicos homólogos*

Típicamente, la invención comprende en primer lugar alinear ácidos nucleicos idénticos, o regiones de similitud de ácidos nucleicos, por ejemplo para secuencias disponibles de cualquiera de las bases de datos públicamente disponibles o propietarias. Entre las bases de datos/servicios de búsqueda públicos se incluyen Genbank<sup>®</sup>, Entrez<sup>®</sup>, EMBL, DDBJ y aquellos proporcionados por el NCBI. Muchas bases de datos adicionales de secuencias se encuentran disponibles en internet o bajo contrato de una diversidad de compañías especializadas en la generación y/o almacenamiento de información genómica.

Los términos "identidad" o porcentaje de "identidad", en el contexto de dos o más ácidos nucleicos o secuencias polipeptídicas se refieren a dos o más secuencias o subsecuencias que son iguales o que presentan un porcentaje especificado de residuos aminoácidos o de nucleótidos que son iguales en la comparación y alineación para la máxima correspondencia, medida utilizando uno de los algoritmos de comparación de secuencias indicados posteriormente (u otros algoritmos disponibles para el experto en la materia) o mediante inspección visual. Dichas secuencias "sustancialmente idénticas" típicamente se consideran homólogas.

Para la comparación de secuencias y la determinación de homología, típicamente una secuencia actúa de secuencia de referencia con la que se comparan las secuencias de ensayo. Al utilizar un algoritmo de comparación de secuencias, se introducen las secuencias de ensayo y de referencia en un ordenador, se designan coordenadas de subsecuencias, en caso necesario, y se fijan los parámetros del programa de algoritmo de secuencias. El algoritmo de comparación de secuencias seguidamente calcula el porcentaje de identidad de secuencias para la secuencia o secuencias de ensayo respecto a la secuencia de referencia, basándose en los parámetros de programa fijados.

## ES 2 341 217 T3

Puede llevarse a cabo una alineación óptima de las secuencias para la comparación, por ejemplo mediante el algoritmo de homología local de Smith y Waterman, Adv. Appl. Math. 2:482, 1981, mediante el algoritmo de alineación de homología de Needleman y Wunsch, J. Mol. Biol. 48:443, 1970, mediante el método de búsqueda de similitud de Pearson y Lipman, Proc. Natl. Acad. Sci. USA 85:2444, 1988, mediante implementaciones computerizadas de estos algoritmos (GAP, BESTFIT, FAST y TFASTA en el paquete informático Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI) o mediante inspección visual (ver de manera general Ausubel *et al.*, *supra*).

Un algoritmo de ejemplo que resulta adecuado para determinar el porcentaje de identidad de secuencias y la similitud de secuencias es el algoritmo BLAST, que se describe en Altschul *et al.*, J. Mol. Biol. 215:403-410, 1990. Los programas para llevar a cabo análisis BLAST se encuentran disponibles públicamente a través del National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). Este algoritmo implica en primer lugar identificar pares de secuencias de alta puntuación (HSPs) mediante la identificación de palabras cortas de longitud W en la secuencia pregunta, que alcanzan o satisfacen cierta puntuación T umbral de valor positivo al alinearla con una palabra de la misma longitud en una secuencia de la base de datos. T se denomina puntuación umbral de palabra vecina (Altschul *et al.*, *supra*). Estos aciertos iniciales de palabras vecinas actúan como semillas para iniciar búsquedas para encontrar HSPs más largas que las contengan. Los aciertos de palabra seguidamente se extienden en ambas direcciones a lo largo de cada secuencia durante la longitud que permite incrementar la puntuación acumulada de alineación. Las puntuaciones acumuladas se calculan utilizando, para secuencias de nucleótidos, los parámetros M (puntuación de premio para una pareja de residuos correspondientes; siempre >0) y N (puntuación de penalización para residuos no correspondientes, siempre >0). Para las secuencias de aminoácidos, se utiliza una matriz de puntuaciones para calcular la puntuación acumulada. La extensión de los aciertos de palabra en cada dirección se detiene cuando la puntuación acumulada de alineación cae en una cantidad X respecto al valor máximo alcanzado; la puntuación acumulada baja hasta cero o menos debido a la acumulación de una o más alineaciones de residuos de puntuación negativa, o cuando se alcanza el extremo de cualquiera de las secuencias. Los parámetros W, T y X del algoritmo BLAST determinan la sensibilidad y velocidad de la alineación. El programa BLASTN (para las secuencias de nucleótidos) utiliza como valores por defecto una longitud de palabra (W) de 11, un valor esperado (E) de 10, un valor de corte de 100, M=5, N=-4 y una comparación de ambas cadenas. Para las secuencias de aminoácidos, el programa BLASTP utiliza como valores por defecto una longitud de palabra (W) de 3, un valor esperado (E) de 10 y una matriz de puntuaciones BLOSUM62 (ver Henikoff y Henikoff, Proc. Natl. Acad. Sci. USA 89:10915, 1989).

Además de calcular el porcentaje de identidad de secuencia, el algoritmo BLAST también lleva a cabo un análisis estadístico de la similitud entre dos secuencias (ver, por ejemplo, Karlin y Altschul, Proc. Natl. Acad. Sci. USA 90:5873-5787, 1993). Una medida de similitud proporcionada por el algoritmo BLAST es la probabilidad de suma mínima (P(N)), que proporciona una indicación de la probabilidad con la que se produciría aleatoriamente una correspondencia entre dos secuencias de nucleótidos o de aminoácidos. Por ejemplo, un ácido nucleico se considera similar a una secuencia de referencia (y, por lo tanto, probablemente homóloga) en el caso de que la probabilidad de suma mínima en una comparación entre el ácido nucleico de ensayo y el ácido nucleico de referencia sea inferior a aproximadamente 0,1, más preferentemente inferior a aproximadamente 0,01, y todavía más preferentemente inferior a aproximadamente 0,001. Entre otros programas de alineación de secuencias disponibles se incluyen, por ejemplo, PILEUP.

### *Síntesis de oligonucleótidos*

La invención comprende sintetizar una pluralidad de oligonucleótidos de trasposición génica familiar, que corresponde a por lo menos una región de diversidad de secuencia. Típicamente se producen conjuntos de oligonucleótidos de trasposición génica familiar, por ejemplo, mediante protocolos de síntesis de oligonucleótidos secuenciales o en paralelo.

Los oligonucleótidos, por ejemplo, sean para la utilización en métodos de amplificación *in vitro*/reconstrucción génica/reensamblaje, o para proporcionar conjuntos de oligonucleótidos de trasposición génica familiar, típicamente se sintetizan químicamente siguiendo el método en fase sólida de triéster de fosforamida descrito por Beaucage y Caruthers, Tetrahedron Letts. 22(20):1859-1862, 1981, por ejemplo utilizando un sintetizador automático, tal como se describe en Needham-VanDevanter *et al.*, Nucleic Acids Res. 12:6159-6168, 1984. Se encuentra comercialmente disponible una amplia diversidad de equipos para la síntesis automática de oligonucleótidos. Los enfoques de síntesis de multinucleótidos (por ejemplo la síntesis de trinucleótidos), tal como se ha comentado *supra*, también resultan útiles.

Además, puede encargarse al efecto esencialmente cualquier ácido nucleico de cualquiera de entre una diversidad de proveedores comerciales, tales como The Midland Certified Reagent Company (mrcr@oligos.com), The Great American Gene Company (<http://www.genco.com>), ExpressGen Inc. ([www.expressgen.com](http://www.expressgen.com)), Operon Technologies Inc. (Alameda, CA) y muchos otros.

### *Ensamblaje de bibliotecas sintéticas*

Se dan a conocer bibliotecas de oligonucleótidos de trasposición génica familiar. Por ejemplo, se alinean genes de interés homólogos utilizando un programa de alineación de secuencias, tal como BLAST, tal como se ha descrito anteriormente. Se indican los nucleótidos correspondientes a variaciones de aminoácidos entre los homólogos. Se diseñan oligos para la trasposición génica sintética que comprenden una (o más) diferencias de nucleótidos respecto

a cualquiera de las secuencias homólogas alineadas, es decir, se diseñan oligos que son idénticos a un primer ácido nucleico, pero que incorporan un residuo en una posición que corresponde a un residuo de un ácido nucleico homólogo, aunque no idéntico, al primer ácido nucleico.

5 Preferentemente, se construyen todos los oligonucleótidos de una longitud seleccionada (por ejemplo, aproximadamente 20, 30, 40, 50, 60, 70, 80, 90 ó 100 ó más nucleótidos) que incorporan todas las posibles variantes de ácidos nucleicos. Entre ellos se incluyen los oligonucleótidos X por cada X variaciones de secuencia, en donde X es el número de secuencias diferentes en un locus. Los X oligonucleótidos son mayoritariamente idénticos en su secuencia, excepto por el nucleótido o nucleótidos que representan el nucleótido o nucleótidos variantes. Debido a esta similitud, puede resultar ventajoso utilizar estrategias de síntesis en paralelo o agrupadas en las que se utiliza una única reacción de síntesis o conjunto de reactivos para construir partes comunes de cada oligonucleótido. Esto puede llevarse a cabo mediante, por ejemplo, técnicas de síntesis en fase sólida de ácidos nucleicos bien conocidas o utilizando, por ejemplo, métodos sintéticos de oligonucleótidos basadas en matrices (ver, por ejemplo, Fodor *et al.*, Science 251:767-777, 1991; Fodor, "Genes, Chips and the Human Genome", FASEB Journal 11:121-121, 1997; Fodor, "Massively Parallel Genomics", Science 277:393-395, 1997, y Chee *et al.*, "Accessing Genetic Information with High-Density DNA Arrays", Science 274:610-614, 1996).

En un aspecto, se seleccionan los oligonucleótidos de manera que únicamente se consideran las alteraciones de aminoácidos codificadas en la estrategia de síntesis. En esta estrategia, tras alinear una familia de ácidos nucleicos homólogos, se sintetizan oligos de trasposición familiar para que sean degenerados únicamente en aquellas posiciones en las que un cambio de base resulta en una alteración de una secuencia polipeptídica codificada. Esto presenta la ventaja de requerir menos oligonucleótidos degenerados para conseguir el mismo grado de diversidad en los productos codificados, simplificando de esta manera la síntesis del conjunto de oligonucleótidos de trasposición génica familiar.

En las estrategias de síntesis en general, los oligonucleótidos presentan por lo menos aproximadamente 10 bases de identidad de secuencia con cada lado de una región de varianza, para garantizar una hibridación y ensamblaje razonablemente eficientes. Sin embargo, las regiones flanqueantes con bases idénticas pueden presentar menos bases idénticas (por ejemplo 5, 6, 7, 8 ó 9) y pueden, evidentemente, presentar regiones más grandes de identidad (por ejemplo 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 50 ó más).

Durante el ensamblaje génico, los oligonucleótidos pueden incubarse conjuntamente y reensamblarse utilizando cualquiera de entre una diversidad de métodos de reensamblaje mediado por polimerasa, por ejemplo tal como se ha descrito en la presente memoria y es conocido por el experto en la materia. La mezcla de reacción puede "pulsarse" con oligonucleótidos seleccionados a cualquier concentración seleccionada, provocando de esta manera la incorporación preferente de las modificaciones deseables.

Por ejemplo, durante la elongación de oligonucleótidos, se incuban oligonucleótidos hibridados en presencia de una ácido nucleico polimerasa, por ejemplo Taq, Klenow, o similar, y dNTPs (es decir, dATP, dCTP, dGTP y dTTP). En el caso de que las regiones de identidad de secuencia sean grandes, puede utilizarse una polimerasa Taq u otra polimerasa de alta temperatura utilizando una temperatura de hibridación de entre aproximadamente la temperatura ambiente y, por ejemplo, aproximadamente 65°C. Si las áreas de identidad son pequeñas, puede utilizarse Klenow, Taq o polimerasas con una temperatura de hibridación inferior a la temperatura ambiente. La polimerasa puede añadirse a los fragmentos de ácido nucleico (oligonucleótidos más cualquier ácido nucleico adicional que forme una mezcla de recombinación) antes, simultáneamente o después de la hibridación de los oligonucleótidos y otros componentes de recombinación. Tal como se ha indicado en otros sitios en la presente exposición, la invención reivindicada implica desnaturalizar las secuencias de ácido nucleico de doble cadena elongadas resultantes y después hibridar y elongar dichas secuencias nuevamente. Este ciclo puede repetirse cualquier número deseado de veces. El ciclo se repite, por ejemplo, entre aproximadamente 2 y aproximadamente 100 veces.

#### 50 Pulsado de bibliotecas

También pueden utilizarse oligonucleótidos familiar para variar los ácidos nucleicos presentes en una mezcla de trasposición típica, por ejemplo una mezcla de fragmentos de ADNasa de uno o más genes de un conjunto homólogo de genes. En un aspecto, la totalidad de los ácidos nucleicos que deben trasponerse se alinean tal como se ha indicado anteriormente. Se registran y/o marcan las variaciones de aminoácidos (por ejemplo en un sistema integrado que comprenda un ordenador con los programas de alineación de secuencias apropiados, o manualmente, por ejemplo sobre una impresión de las secuencias o alienaciones de secuencias. Tal como anteriormente, los oligos de trasposición familiar se diseñan para que incorporen algunas o todas las variaciones de aminoácidos codificadas por la diversidad natural de secuencia de los ácidos nucleicos alineados. Uno o más ácidos nucleicos correspondientes al conjunto homólogo de ácidos nucleicos alineados se cortan (por ejemplo utilizando una ADNasa, o mediante corte químico). Se pulsa una mezcla de ácidos nucleicos cortados con los oligos de trasposición familiar, que seguidamente se recombinan y se reensamblan para formar secuencias de longitud completa utilizando técnicas estándares.

Para determinar el grado de incorporación de oligonucleótidos, puede utilizarse cualquier enfoque que distinga entre ácidos nucleicos similares. Por ejemplo, los ácidos nucleicos reensamblados pueden clonarse y secuenciarse, o amplificarse (*in vitro* o mediante clonación, por ejemplo en un vector de clonación estándar) y cortarse con un enzima de restricción que reconoce específicamente una secuencia polimórfica particular presente en los oligos de trasposición familiar, pero no presentes en la misma posición en el ácido o ácidos nucleicos cortados originales.

En otra realización, se seleccionan oligonucleótidos que incorporan una o más variaciones de secuencia correspondientes a un polimorfismo de aminoácidos, pero que eliminan las variaciones de nucleótidos polimórficos entre secuencias de ácidos nucleicos que corresponden a sustituciones silenciosas. Una ventaja de esta estrategia es que la eliminación de sustituciones silenciosas puede provocar que una secuencia dada sea más similar a un sustrato dado para la recombinación (por ejemplo un ácido nucleico diana seleccionado). Esta similitud incrementada permite la recombinación de ácidos nucleicos entre secuencias que de otra manera podrían ser excesivamente diversas para la recombinación eficiente.

Por ejemplo, puede amplificarse por PCR un ácido nucleico seleccionado utilizando métodos estándares. El ácido nucleico seleccionado se corta y se mezcla con una biblioteca de oligonucleótidos de trasposición génica familiar que se construyen para que resulten lo más similares posible a las secuencias correspondientes del ácido nucleico seleccionado haciendo que los oligonucleótidos incluyan el mismo conjunto de sustituciones silenciosas presentes en el ácido nucleico seleccionado. La mezcla de corte se pulsa con los oligonucleótidos a una concentración seleccionada, que después se reensambla para formar secuencias de longitud completa. La calidad de la biblioteca resultante (por ejemplo la frecuencia a la que los oligos se incorporan en las secuencias reensambladas) se comprueba, tal como se ha indicado anteriormente, mediante clonación (o, de otra manera, mediante amplificación) y secuenciación y/o digestión de restricción de las secuencias reensambladas.

También pueden utilizarse estrategias de elongación por PCR para preparar bibliotecas utilizando diferentes proporciones molares de oligonucleótidos en las mezclas de recombinación (ver también, por ejemplo, las patentes WO n° 97/20078, n° 98/42832 y n° 98/01581).

#### *Formatos iterativos de oligonucleótidos*

En un aspecto, se dan a conocer formatos iterativos de recombinación mediada por oligonucleótidos. Estos formatos pueden combinarse con métodos estándares de recombinación, también, opcionalmente, en formato iterativo.

En particular, pueden cribarse para actividad y secuenciarse ácidos nucleicos recombinantes producidos mediante recombinación mediada por oligonucleótidos. Los ácidos nucleicos recombinantes secuenciados se alinean y se identifican las regiones de identidad y de diversidad. A continuación, se seleccionan los oligonucleótidos de trasposición familiar para la recombinación de los ácidos nucleicos recombinantes secuenciados. Este procedimiento de cribado, secuenciación de ácidos nucleicos recombinantes y recombinación de los ácidos nucleicos recombinantes activos puede repetirse iterativamente hasta obtener una molécula con una propiedad deseada.

Además, pueden cortarse y trasponerse utilizando métodos estándares de recombinación ácidos nucleicos recombinantes construidos utilizando oligonucleótidos de trasposición familiar, que son, opcionalmente, reiterativos. Puede utilizarse la recombinación estándar conjuntamente con la trasposición de oligonucleótidos y una de las dos etapas o ambas opcionalmente pueden repetirse reiterativamente.

Un ejemplo útil de trasposición iterativa mediante recombinación mediada por oligonucleótidos de los oligonucleótidos familiar se produce en el caso de que se desee trasposición de grano extremadamente fino. Por ejemplo, los genes pequeños codificantes de proteínas pequeñas, tales como las defensinas (proteínas antifúngicas de aproximadamente 50 aminoácidos) EF40 (una familia de proteínas antifúngicas de aproximadamente 28 aminoácidos), antibióticos peptídicos, proteínas peptídicas insecticidas, hormonas peptídicas, muchas citoquinas y muchas otras proteínas pequeñas, son difíciles de recombinar mediante métodos estándares de recombinación debido a que la recombinación con frecuencia se produce a una frecuencia que es aproximadamente igual al tamaño del gen que debe recombinarse, limitando la diversidad resultante de la recombinación. En contraste, los métodos de recombinación mediada por oligonucleótidos pueden recombinar esencialmente cualquier región de diversidad en cualquier conjunto de secuencias, con sucesos de recombinación (por ejemplo entrecruzamientos) en cualquier par de bases seleccionado.

De esta manera, algunas bibliotecas de secuencias preparadas mediante recombinación recursiva mediada por oligonucleótidos opcionalmente se criban y se seleccionan para una propiedad deseada, y se secuencian (o, de otra manera, se deconvolutan, por ejemplo mediante análisis de PCR en tiempo real, tal como FRET o TaqMan, o utilizando análisis de enzimas de restricción) clones mejorados (o, de otra manera, deseables) repitiendo iterativamente el procedimiento para generar bibliotecas adicionales de ácidos nucleicos. De esta manera, se llevan a cabo rondas adicionales de recombinación utilizando métodos estándares de recombinación basados en fragmentación, o mediante la secuenciación de clones positivos, diseñando oligonucleótidos apropiados de trasposición familiar y llevando a cabo una segunda ronda de recombinación/selección para producir una biblioteca adicional (que puede recombinarse tal como se ha indicado). Además, también pueden recombinarse bibliotecas construidas a partir de diferentes rondas de recombinación, mediante secuenciación/recombinación de oligonucleótidos o mediante métodos estándares de recombinación.

#### *Trasposición de oligonucleótidos entrecruzados de PCR*

En un aspecto, se da a conocer en la presente memoria la trasposición de secuencias distantemente relacionadas o incluso no homólogas. En este aspecto, se diseñan oligonucleótidos entrecruzados de PCR con una primera región derivada de un primer ácido nucleico y una segunda región correspondiente a un segundo ácido nucleico. Se diseñan oligos adicionales que corresponden al primer o segundo ácido nucleico, y que presentan secuencias que son comple-

mentarias (o idénticas) a los oligos entrecruzados. Mediante la recombinación de estos oligos (es decir, hibridándolos y después elongando los oligonucleótidos hibridados en reacciones sucesivas de elongación mediadas por polimerasa), se proporciona un sustrato que puede recombinarse con el primer o segundo ácido nucleico y que incorporará, simultáneamente, secuencias del otro ácido nucleico.

5

#### *Oligonucleótidos de uno o más codones modificados*

Los oligonucleótidos de uno o más codones modificados son oligonucleótidos de secuencia similar aunque con una o más variaciones de bases en las que las variaciones corresponden a por lo menos una diferencia de aminoácido codificado. Pueden sintetizarse utilizando una reacción de trinucleótido, es decir una reacción de acoplamiento de fosforamiditas basada en codones, en la que se utilizan fosforamiditas trinucleótido que representan codones de la totalidad de los 20 aminoácidos para introducir codones enteros en secuencias de oligonucleótidos sintetizadas mediante esta técnica de fase sólida. Preferentemente, se sintetiza la totalidad de los oligonucleótidos de una longitud seleccionada (por ejemplo aproximadamente 20, 30, 40, 50, 60, 70, 80, 90 ó 100 ó más nucleótidos) que incorporan las secuencias de ácido nucleico seleccionadas. En la presente invención, las secuencias de oligonucleótido de uno o más codones modificados pueden basarse en secuencias de un conjunto seleccionado de ácidos nucleicos homólogos.

15

La síntesis de fosforamiditas trinucleótido, su utilización posterior en la síntesis de oligonucleótidos, y cuestiones relacionadas se describen en, por ejemplo, Virnekäs B. *et al.*, *Nucleic Acids Res.* 22:5600-5607, 1994, Kayushin A.L. *et al.*, *Nucleic Acids Res.* 24:3748-3755, 1996, Huse *et al.*, patente US nº 5.264.563, "Process for synthesizing oligonucleotides with random codons", Lyttle *et al.*, patente US nº 5.717.085, "Process for preparing codon amidites", Shortle *et al.*, patente US nº 5.869.644, "Synthesis of diverse and useful collections of oligonucleotides", Greyson, patente US nº 5.789.577, "Method for the controlled synthesis of polynucleotide mixtures which encode desired mixtures of peptides", y Huse, patente WO nº 92/06176, "Surface expression libraries of randomized peptides".

25

Los oligonucleótidos de uno o más codones modificados pueden sintetizarse utilizando diversas técnicas relacionadas con trinucleótidos, por ejemplo, el formato de síntesis de trinucleótidos y el formato de síntesis split-pool. La química implicada en los métodos sintéticos de oligonucleótidos de uno o más codones modificados, tanto el método del trinucleótido como el de división-agrupación ("split-pool"), son bien conocidos por el experto en la materia. En general, ambos métodos utilizan la síntesis química en fase sólida de fosforamidita en la que los extremos 3' de las secuencias de sustrato de ácidos nucleicos se unen covalentemente a un soporte sólido, por ejemplo vidrio de poro controlado. Los grupos 5'-protectores pueden ser, por ejemplo, un grupo trifenilmetilo, tal como dimetoxiltrilito (DMT) o monometioxiltrilito, un grupo que contenga carbonilo, tal como 9-fluorenilmetiloxycarbonilo (Fmoc) o levulinilo; un grupo cortable con ácido, tal como terc-butil-dimetilsililo (T-BDMSi), trisopropilsililo o trimetilsililo. Los grupos 3'-protectores pueden ser, por ejemplo, grupos  $\beta$ -cianoetilo.

35

El formato de síntesis de trinucleótido incluye proporcionar una secuencia de sustrato que presenta un extremo 5'-terminal y por lo menos una base, presentando ambos, grupos protectores en los mismos. El grupo 5'-protector de la secuencia de sustrato seguidamente se elimina para proporcionar una secuencia de sustrato de extremo 5' desprotegido que seguidamente se acopla con una secuencia fosforamidita trinucleótido seleccionada. El trinucleótido presenta un extremo 3'-terminal, un extremo 5'-terminal y tres bases, cada una de las cuales presenta grupos protectores en las mismas. La etapa de acoplamiento proporciona una secuencia extendida de oligonucleótido. A continuación, opcionalmente se repiten las etapas de eliminación y acoplamiento. Al repetir estas etapas, la secuencia extendida de oligonucleótido proporcionada por cada etapa de acoplamiento repetida se convierte en la secuencia de sustrato de la siguiente etapa repetida de eliminación hasta obtener un oligonucleótido deseado con uno o más codones modificados. Este formato básico de la síntesis puede incluir opcionalmente el acoplamiento de uno o más de entre: mononucleótidos, secuencias de fosforamidita trinucleótido, y oligonucleótidos.

40

45

El formato de síntesis de división-agrupación incluye proporcionar secuencias de sustrato, presentando cada una un extremo 5'-terminal y por lo menos una base, presentando ambos, grupos protectores en los mismos. Los grupos 5'-protectores de las secuencias de sustrato se eliminan para proporcionar secuencias de sustrato de extremo 5' desprotegido que seguidamente se acopla con secuencias de fosforamidita trinucleótido seleccionadas. Cada trinucleótido presenta un extremo 3'-terminal, un extremo 5'-terminal y tres bases, la totalidad de las cuales presenta grupos protectores en las mismas. La etapa de acoplamiento proporciona secuencias extendidas de oligonucleótido. A continuación, opcionalmente se repiten las etapas de eliminación y acoplamiento. Al repetir estas etapas, las secuencias extendidas de oligonucleótido proporcionadas por cada etapa repetida de acoplamiento se convierten en las secuencias de sustrato de la etapa de eliminación repetida siguiente, hasta producir las secuencias extendidas de oligonucleótido intermediario.

55

Entre las etapas adicionales del formato de división-agrupación opcionalmente se incluyen la división de las secuencias oligonucleótidas intermediarias extendidas en dos o más agrupaciones separadas. A continuación, se eliminan los grupos 5'-protectores de las secuencias oligonucleótidas intermediarias extendidas para proporcionar en dos o más agrupaciones separadas secuencias oligonucleótidas intermediarias extendidas con el extremo 5' desprotegido. Después, estos intermediarios con extremo 5' desprotegido se acoplan con uno o más mononucleótidos seleccionados, secuencias fosforamidita trinucleótido u oligonucleótidos en las dos o más agrupaciones separadas para proporcionar secuencias oligonucleótidas intermediarias extendidas adicionales. A su vez, estas secuencias extendidas adicionales se agrupan en una única agrupación. A continuación, las etapas a partir de la eliminación de los grupos 5'-protectores de las secuencias de sustrato opcionalmente se repiten. Durante la repetición de estas etapas, las secuencias oligo-

60

65

nucleótidas más extendidas, proporcionadas por cada etapa repetida de acoplamiento que genera aquellas secuencias específicas, se convierten en las secuencias de sustrato de la etapa repetida de eliminación siguiente, que incluye aquellas secuencias específicas, hasta obtener los oligonucleótidos de uno o más codones modificados.

5 Ambos protocolos sintéticos descritos, *supra*, pueden llevarse a cabo opcionalmente en un sintetizador automático que lleva a cabo automáticamente las etapas. Este aspecto comprende la introducción de información de cadena de caracteres en un ordenador, la producción del cual seguidamente dirige el sintetizador automático a la realización de las etapas necesarias para sintetizar los oligonucleótidos deseados de codón o codones modificados.

10 Pueden encontrarse detalles adicionales sobre la síntesis de trinucleótido en "Use of codon varied oligonucleotide synthesis for synthetic shuffling", por Welch *et al.*, documento USSN nº 09/408.393, presentado el 28 de septiembre de 1999.

#### *Ajuste fino de la recombinación de ácidos nucleicos utilizando la mezcla mediada por oligonucleótidos*

15 En los métodos reivindicados, se utilizan proporciones no equimolares de oligonucleótidos de trasposición familiar para sesgar la recombinación. En este enfoque, *no* se utilizan proporciones equimolares de oligonucleótidos de trasposición familiar en un conjunto de oligonucleótidos de trasposición familiar para producir una biblioteca de ácidos nucleicos recombinantes, como en ciertos otros métodos proporcionados en la presente memoria. Por el contrario, el profesional selecciona proporciones de oligonucleótidos particulares que corresponden a las secuencias de un miembro seleccionado o conjunto seleccionado de miembros de la familia de ácidos nucleicos a partir de los que se derivan los oligonucleótidos de trasposición de la familia.

25 De esta manera, en un ejemplo ilustrativo simple, se utiliza la recombinación mediada por oligonucleótidos tal como se describe en la presente memoria para recombinar, por ejemplo, un gen de rana y un gen humano que son idénticos al 50%. Se sintetizan oligonucleótidos familiar que codifican las secuencias tanto humanas como de rana en todas las posiciones polimórficas. Sin embargo, en lugar de utilizar una proporción equimolar de los oligonucleótidos derivados de ser humano y de rana, se sesga la proporción en favor del gen que el usuario desea emular más estrechamente. Por ejemplo, al generar un gen de tipo humano, la proporción de oligonucleótidos que corresponde a la secuencia humana en posición polimórficas puede sesgarse a una proporción superior a 50% (por ejemplo aproximadamente 60%, 70%, 80% o 90% o más de los oligos puede corresponder a la secuencia humana, con, por ejemplo, aproximadamente 40%, 30%, 20%, 10% o menos de los oligos correspondiente a la secuencia de rana). De manera similar, si se desea un gen de tipo rana, la proporción de oligonucleótidos que corresponde a la secuencia de rana en sitios polimórficos puede sesgarse a más del 50%. En cualquiera de los dos casos, el gen "mezclado" resultante (es decir, el gen recombinante resultante con características de más de un gen parental) seguidamente puede recombinarse con miembros de la familia génica que presentan una secuencia estrechamente relacionada con la del gen mezclado. De esta manera, en el caso anteriormente indicado, en el caso de que se seleccione la proporción de oligonucleótidos para producir un gen mezclado más similar al humano, el gen mezclado opcionalmente se recombina adicionalmente con genes más estrechamente similares al gen humano original. De manera similar, en el caso de que se seleccione la proporción de oligonucleótidos para producir un gen mezclado más similar al de rana, el gen mezclado opcionalmente se recombina adicionalmente con genes más estrechamente similares al gen de rana original. Se proporciona esta estrategia en la figura 2. La estrategia resulta generalmente aplicable a la recombinación de dos o más ácidos nucleicos cualesquiera mediante recombinación mediada por oligonucleótidos.

45 El sesgo puede conseguirse de una diversidad de maneras, incluyendo la síntesis de cantidades desproporcionadas de los oligonucleótidos relevantes, o simplemente suministrando cantidades desproporcionadas al método de síntesis génica relevante (por ejemplo a un método sintético de PCR tal como se ha indicado *supra*).

50 Tal como se ha indicado, el enfoque de sesgado puede aplicarse a la recombinación de cualquier conjunto de dos o más ácidos nucleico relacionados. Las secuencias no necesitar se estrechamente similares para realizar la selección. De hecho, ni siquiera resulta necesario que las secuencias sean detectablemente homólogas para que se produzca el sesgo. En este caso, se sustituyen oligonucleótidos "de la familia" por conjuntos de oligonucleótidos de secuencia no homóloga derivados de la consideración de la similitud estructural de las proteínas codificadas. Por ejemplo, la superfamilia de inmunoglobulinas incluye miembros estructuralmente similares que muestran una homología de secuencias reducida o no detectable (especialmente al nivel de los ácidos nucleicos). En estos casos, las secuencias no homólogas se "alinean" mediante consideración de la homología estructural (por ejemplo mediante alineación de residuos peptídicos funcionalmente similares). Puede definirse un espacio de recombinación de interés que incluya todas las permutaciones de la diversidad de aminoácidos representada por la alineación. El método de sesgado anteriormente indicado opcionalmente se utiliza para mezclar las secuencias con proporciones deseadas de los nucleótidos codificantes de secuencias relevantes de aminoácidos estructuralmente similares.

65 Puede alinearse dos o más secuencias cualesquiera mediante cualquier algoritmo o criterios de interés y el método de sesgado utilizado para mezclar las secuencias puede basarse en cualesquiera criterios deseados. Entre estos se incluyen la homología de las secuencias, la similitud estructural, la similitud estructural predicha (a partir de cualesquiera criterios de similitud que se especifiquen), o similares. Puede aplicarse a situaciones en las que existe un núcleo estructural que es constante, pero que presenta muchas variaciones estructurales construidas en torno al núcleo (por ejemplo, un dominio Ig puede ser un núcleo estructural que presente muchas longitudes de bucle y conformaciones que se unen al núcleo).

Una ventaja general de dicho enfoque en comparación con los métodos estándares de recombinación génica es que la identidad de secuencia global de dos secuencias que deben mezclarse puede ser inferior a la identidad necesaria para que se produzca la recombinación mediante métodos más estándares. Además, en ocasiones únicamente se re-combinan regiones seleccionadas, permitiendo tener en cuenta cualquier dato estructural o funcional que se encuentre disponible al especificar cómo se construye el gen mezclado. De esta manera, se accede a espacio de secuencia que no se produce mediante algunos otros protocolos de traslocación mediante el enfoque de mezcla génica y en ocasiones puede obtenerse un porcentaje más alto de clones activos si se tiene en cuenta información estructural.

La estrategia general anteriormente indicada resulta aplicable, por ejemplo, a cualquier conjunto de genes con baja similitud de secuencia. Por ejemplo, existe una gran familia de homólogos de TNF cuya identidad de secuencia es del orden de aproximadamente 30%, provocando que resulten difíciles de seguir los protocolos estándares de transposición. Evidentemente el ajuste fino de la recombinación mediante la selección de las proporciones de oligonucleótidos también resulta generalmente aplicable a la recombinación de dos ácidos nucleicos cualesquiera, incluyendo tanto los homólogos de elevada similitud como los homólogos de similitud reducida. Puede seleccionarse cualquier protocolo de alineación para alinear dos o más secuencias, y la alineación resultante puede utilizarse para crear oligonucleótidos apropiados para llevar a cabo la recombinación, y puede llevarse a cabo cualquier sesgado de las frecuencias relativas de las secuencias en comparación con las secuencias parentales.

#### *Dianas de transposición de oligonucleótidos*

Puede trasponerse esencialmente cualquier ácido nucleico utilizando los métodos mediados por oligonucleótidos dados a conocer en la presente memoria. No se realiza ningún intento para identificar los cientos de miles de ácidos nucleicos conocidos. Tal como se ha indicado anteriormente, entre los repositorios de secuencias comunes para las proteínas conocidas se incluyen GenBank, EMBL, DDBJ y el NCBI. Pueden identificarse fácilmente otros repositorios mediante la búsqueda en internet.

Una clase de dianas preferentes para la activación incluye ácidos nucleicos codificantes de proteínas terapéuticas, tales como la eritropoyetina (EPO), la insulina, las hormonas peptídicas, tales como la hormona del crecimiento humana; factores de crecimiento y citoquinas, tales como el péptido 78 activador de neutrófilos epiteliales, GRO $\alpha$ /MGSA, GRO $\beta$ , GRO $\gamma$ , MIP-1 $\alpha$ , MIP-1 $\beta$ , MCP-1, factor de crecimiento epidérmico, factor de crecimiento fibroblástico, factor de crecimiento de hepatocitos, factor de crecimiento similar a la insulina, los interferones, las interleuquinas, el factor de crecimiento de queratinocitos, el factor inhibidor de la leucemia, la oncostatina M, PD-ECSF, PDGF, pleiotropina, SCF, ligando de c-kit, VEGF, G-CSF, etc. Muchas de estas proteínas se encuentran disponibles comercialmente (ver, por ejemplo, el catálogo 1997 de Sigma BioSciences y la lista de precios) y los genes correspondientes son bien conocidos.

Otra clase de dianas preferentes son los activadores transcripcionales y de expresión. Entre los ejemplos de activadores transcripcionales y de expresión se incluyen genes y proteínas que modulan el crecimiento, la diferenciación y la regulación celular, y similares. Se encuentran activadores de expresión y transcripcionales en procariotas, virus y eucariotas, incluyendo hongos, plantas y animales, entre ellos mamíferos, proporcionando un amplio abanico de dianas terapéuticas. Se apreciará que los activadores de expresión y de transcripción regulan la transcripción mediante muchos mecanismos, por ejemplo mediante la unión a receptores, la estimulación de una cascada de transducción de señales, la regulación de la expresión de factores transcripcionales, la unión a promotores e intensificadores, la unión a proteínas que se unen a promotores e intensificadores, el desenrollamiento del ADN, el procesamiento previo al ARNm, la poliadenilación del ARN, y la degradación del ARN. Entre los activadores de expresión se incluyen citoquinas, moléculas inflamatorias, factores de crecimiento, receptores de los mismos y productos oncogénicos, por ejemplo interleuquinas (por ejemplo IL-1, IL-2, IL-8, etc.), interferones, FGF, IGF-I, IGF-II, FGF, PDGF, TNF, TGF- $\alpha$ , TGF- $\beta$ , EGF, KGF, SCF/c-Kit, CD40L/CD40, VLA-4/VCAM-1, ICAM-1/LFA-1 e hialurina/CD44; moléculas de transducción de señales y los productos oncogénicos correspondientes, por ejemplo Mos, Ras, Raf y Met, y activadores y supresores transcripcionales, por ejemplo p53, Tat, Fos, Myc, Jun, Myb, Rel, y receptores de hormona esteroidea, tales como los receptores de estrógeno, progesterona, testosterona, aldosterona, ligando del receptor LDL y corticoesterona.

Las ARNasas tales como la onconasa y el EDN son dianas preferentes para los métodos sintéticos en la presente memoria, particularmente aquellos métodos que utilizan la mezcla de genes. El experto en la materia apreciará que se conocen ARNasas tanto de rana como humanas que presentan varias actividades farmacológicas importantes. Debido a la divergencia evolutiva entre estos genes, los métodos de recombinación mediada por oligonucleótidos resultan particularmente útiles para recombinar los ácidos nucleicos.

De manera similar, proteínas de organismos infecciosos para posibles aplicaciones de vacuna, descritas en más detalle posteriormente, incluyendo hongos infecciosos, por ejemplo *Aspergillus*, especies de *Candida*; bacterias, particularmente *E. coli*, que sirve como modelo para bacterias patógenas, así como bacterias médicamente importantes, tales como estafilocos (por ejemplo *Staphylococcus aureus*), estreptococos (por ejemplo *Streptococcus pneumoniae*), *Clostridia* (por ejemplo *C. perfringens*), *Neisseria* (por ejemplo *N. gonorrhoea*), enterobacteriáceas (por ejemplo *Enterobacter coli*), *Helicobacter* (por ejemplo *H. pylori*), *Vibrio* (por ejemplo *V. cholerae*), *Campylobacter* (por ejemplo *C. jejuni*), *Pseudomonas* (por ejemplo *P. aeruginosa*), *Haemophilus* (por ejemplo *H. influenzae*), *Bordetella* (por ejemplo *B. pertussis*), *Mycoplasma* (por ejemplo *M. pneumoniae*), *Ureaplasma* (por ejemplo *U. urealyticum*), *Legionella* (por ejemplo *L. pneumophila*), espiroquetas (por ejemplo *Treponema*, *Leptospira* y *Borrelia*), micobacterias (por

ejemplo *Mycobacterium tuberculosis*, *M. smegmatis*), *Actinomyces* (por ejemplo *A. israelii*), *Nocardia* (por ejemplo *N. asteroides*), *Chlamydia* (por ejemplo *C. trachomatis*), *Rickettsia*, *Coxiella*, *Ehrlichia*, *Rocholima*, *Brucella*, *Yersinia*, *Francisella* y *Pasteurella*; protozoos, tales como esporozoos (por ejemplo *Plasmodia*), rizópodos (por ejemplo *Entamoeba*) y flagelados (*Trypanosoma*, *Leishmania*, *Trichomonas*, *Giardia*, etc.); virus, tales como virus ARN(+) (entre los ejemplos se incluyen Poxvirus, por ejemplo *Vaccinia*; Picornavirus, por ejemplo HCV; y Coronavirus), virus ARN(-) (entre los ejemplos se incluyen Rabdovirus, por ejemplo VSV; Paramixovirus, por ejemplo RSV; Ortomixovirus, por ejemplo virus influenza, Bunyavirus y Arenavirus), virus ADNdc (por ejemplo Reovirus), virus ARN a ADN, es decir, retrovirus, por ejemplo especialmente VIH y HTLV, y determinados virus ADN a ARN, tales como el virus de la hepatitis B.

Otras proteínas relevantes para usos no médicos, tales como inhibidores de transcripción o toxinas de plagas de cultivos, por ejemplo insectos, hongos, malas hierbas, y similares, también son dianas preferentes para la trasposición de oligonucleótidos. Algunos enzimas industrialmente importantes, tales como las monooxigenasas (por ejemplo p450S), proteasas, nucleasas y lipasas también son dianas preferentes. A título de ejemplo, puede hacerse evolucionar la subtilisina mediante la trasposición de oligonucleótidos familiar para formas homólogas del gen de la subtilisina. Von der Osten *et al.*, J. Biotechnol. 28:55-68, 1993, proporcionan una subtilisina de ejemplo codificante de ácidos nucleicos, y se encuentran ácidos nucleicos adicionales en GENBANK®. Las proteínas adyuvantes del plegamiento, tales como las chaperoninas, también son dianas preferentes.

Entre los genes conocidos preferentes adecuados para la trasposición mediada por oligonucleótidos también se incluyen los siguientes: antitripsina alfa-1, angiostatina, factor antihemolítico, apolipoproteína, apoproteína, factor natriurético atrial, polipéptido natriurético atrial, péptidos atriales, quimoquinas C-X-C (por ejemplo T39765, NAP-2, ENA-78, Gro-a, Gro-b, Gro-c, IP-10, GCP-2, NAP-4, SDF-1, PF4, MIG), calcitonina, quimoquinas CC (por ejemplo la proteína quimoatrayente 1 de monocitos, la proteína quimoatrayente 2 de monocitos, la proteína quimoatrayente 3 de monocitos, la proteína 1 alfa inflamatoria de monocitos, la proteína 1 beta inflamatoria de monocitos, RANTES, I309, R83915, R91733, HCC1, T58847, D31065, T64262), ligando CD40, colágeno, factor estimulante de colonias (CSF), factor del complemento 5a, inhibidor del complemento, receptor 1 del complemento, factor IX, factor VII, factor VIII, factor X, fibrinógeno, fibronectina, glucocerebrosidasa, gonadotropina, proteínas de erizo (por ejemplo Sonic, Indian, Desert), hemoglobina (para sustitutivo de sangre; para radiosensibilización), hirudina, albúmina sérica humana, lactoferrina, luciferasa, neuriturina, factor inhibidor de neutrófilos (NIF), proteína osteogénica, hormona paratiroidea, proteína A, proteína G, relaxina, renina, calcitonina del salmón, hormona de crecimiento del salmón, receptor I del complemento soluble, I-CAM 1 soluble, receptores solubles de interleuquina (IL-1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15), receptor soluble de TNF, somatomedina, somatostatina, somatotropina, estreptoquinasa, superantígenos, es decir, enterotoxinas estafilocócicas (SEA, SEB, SEC1, SEC2, SEC3, SED, SEE), toxina del síndrome del choque tóxico (TSST-1), toxinas A y B exfoliantes, exotoxinas pirogénicas A, B y C, y mitógeno de *M. arthritides*, superóxido dismutasa, timosina alfa 1, activador del plasminógeno tisular, factor beta de necrosis tumoral (TNF beta), receptor del factor de necrosis tumoral (TNFR), factor alfa de necrosis tumoral (TNF alfa) y uroquinasa.

Algunas proteínas pequeñas, tales como defensinas (proteínas antifúngicas de aproximadamente 50 aminoácidos, EF40 (una proteína antifúngica de 28 aminoácidos), péptidos antibióticos, y proteínas insecticidas también son dianas preferentes y existen en forma de familias de proteínas relacionadas. Los ácidos nucleicos codificantes de proteínas pequeñas son dianas particularmente preferentes, debido a que los métodos convencionales de recombinación proporcionan únicamente una diversidad de secuencia del producto limitada. Esto se debe a que la metodología convencional de recombinación produce entrecruzamientos entre secuencias homólogas aproximadamente cada 50 a 100 pares de bases. Esto significa que para dianas de recombinación muy cortas, los entrecruzamientos mediante técnicas estándares se producen aproximadamente una vez por molécula. En contraste, los formatos de trasposición de oligonucleótidos en la presente memoria permiten la recombinación de ácidos nucleicos pequeños, debido a que el profesional selecciona cualquier "entrecruzamiento" deseado.

#### 50 *Trasposición del ADN y reensamblaje génico. métodos de trasposición de híbridos sintéticos*

Un aspecto de la presente exposición es la capacidad de utilizar la trasposición de oligonucleótidos familiar y oligonucleótidos entrecruzados como moldes/intermediarios de recombinación en diversos métodos de trasposición de ADN. Además, los ácidos nucleicos construidos con las nuevas técnicas sintéticas en la presente memoria pueden trasponerse nuevamente utilizando otras metodologías de trasposición disponibles.

Se conoce una diversidad de dichos métodos, incluyendo aquellos que enseñan los inventores y sus colaboradores. Las publicaciones siguientes describen una diversidad de procedimientos recursivos de recombinación y/o métodos relacionados que pueden ponerse en práctica conjuntamente con los procedimientos de la invención: Stemmer *et al.*, "Molecular breeding of viruses for targeting and other clinical properties. Tumor Targeting" 4:1-4, 1999; Nasset *et al.*, "DNA Shuffling of subgenomic sequences of subtilisin", Nature Biotechnology 17:893-896, 1999; Chang *et al.*, "Evolution of a cytokine using DNA family shuffling", Nature Biotechnology 17:793-797, 1999; Minshull y Stemmer, "Protein evolution by molecular breeding", Current Opinion in Chemical Biology 3:284-290, 1999; Christians *et al.*, "Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling", Nature Biotechnology 17:259-264, 1999; Crameriet *et al.*, "DNA shuffling of a family of genes from diverse species accelerates directed evolution", Nature 391:288-291, 1998; Crameri *et al.*, "Molecular evolution of an arsenate detoxification pathway by DNA shuffling", Nature Biotechnology 15:436-438, 1997; Zhang *et al.*, "Directed evolution of an effective fucosidase from a galactosidase by DNA shuffling and screening", Proceedings of the National Academic of Sciences, U.S.A.

94:4504-4509, 1997; Patten *et al.*, “Applications of DNA Shuffling to Pharmaceuticals and Vaccines”, *Current Opinion in Biotechnology* 8:724-733, 1997; Crameri *et al.*, “Construction and evolution of antibody-phage libraries by DNA shuffling”, *Nature Medicine* 2:100-103, 1996; Crameri *et al.*, “Improved green fluorescent protein by molecular evolution using DNA shuffling”, *Nature Biotechnology* 14:315-319, 1996; Gates *et al.*, “Affinity selective isolation of ligands from peptide libraries through display on a lac repressor’s head-piece dimer”, *Journal of Molecular Biology* 255:373-386, 1996; Stemmer, “Sexual PCR and Assembly PCR”, en: *The Encyclopedia of Molecular Biology*, VCH Publishers, New York, páginas 447 a 457, 1996; Crameri y Stemmer, “Combinatorial multiple cassette mutagenesis creates all the permutations of mutante and wildtype cassettes”, *BioTechniques* 18:194-195, 1995; Stemmer *et al.*, “Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides”, *Gene* 164:49-53, 1995; Stemmer, “The Evolution of Molecular Computation”, *Science* 270:1510, 1995; Stemmer, “Searching Sequence Space”, *Bio/Technology* 13:549-553, 1995; Stemmer, “Rapid evolution of a protein *in vitro* by DNA shuffling”, *Nature* 370:389-391, 1994, y Stemmer, “DNA shuffling by random fragmentation and reassembly: *In vitro* recombination for molecular evolution”, *Proceedings of the National Academy of Sciences, U.S.A.* 91:10747-10751, 1994.

Pueden encontrarse detalles adicionales sobre los métodos de trasposición de ADN en las patentes US de los inventores y sus colaboradores, incluyendo: patente U.S. n° 5.605.793, de Stemmer (25 de febrero de 1997), titulada “Methods for *in vitro* recombination”; patente U.S. n° 5.811.238, de Stemmer *et al.* (22 de septiembre de 1998), titulada “Methods for generating polynucleotides having desired characteristics by iterative selection and recombination”; patente U.S. n° 5.830.721, de Stemmer *et al.* (3 de noviembre de 1998), titulada “DNA mutagenesis by random fragmentation and reassembly”; patente U.S. n° 5.834.252, de Stemmer *et al.* (10 de noviembre de 1998), “End-complementary polymerase reaction”, y la patente U.S. n° 5.837.458, de Minshull *et al.* (17 de noviembre de 1998), titulada “Methods and compositions for cellular and metabolic engineering”.

Además, pueden encontrarse detalles y formatos para la trasposición de ácidos nucleicos en una diversidad de solicitudes publicadas de patentes extranjeras, incluyendo: Stemmer y Crameri, “DNA mutagenesis by random fragmentation and reassembly”, patente WO n° 95/22625; Stemmer y Lipschutz, “End complementary polymerase chain reaction”, patente WO n° 96/33207; Stemmer y Crameri, “Methods for generating polynucleotides having desired characteristics by iterative selection and recombination”, patente WO n° 97/0078; Minshull y Stemmer, “Methods and compositions for cellular and metabolic engineering”, patente WO n° 97/35966; Punnonen *et al.*, “Targeting of genetic vaccine vectors”, patente WO n° 99/41402; Punnonen *et al.*, “Antigen library immunization”, patente WO n° 99/41383; Punnonen *et al.*, “Genetic vaccine vector engineering”, patente WO n° 99/41369; Punnonen *et al.*, “Optimization of immunomodulatory properties of genetic vaccines”, patente WO n° 99/41368; Stemmer y Crameri, “DNA mutagenesis by random fragmentation and reassembly”, patente EP n° 0934999; Stemmer, “Evolving cellular DNA uptake by recursive sequence recombination”, patente EP n° 0932670; Stemmer *et al.*, “Modification of virus tropism and host range by viral genome shuffling”, patente WO n° 99/23107; Apt *et al.*, “Human papillomavirus vectores”, patente WO n° 99/21979; Del Cardayre *et al.*, “Evolution of whole cells and organisms by recursive sequence recombination”, patente WO n° 98/31837; Patten y Stemmer, “Methods and compositions for polypeptide engineering”, patente WO n° 98/27230; Stemmer *et al.*, “Methods for optimization of gene therapy by recursive sequence shuffling and selection”, patente WO n° 98/13487.

Determinadas solicitudes de patentes U.S. proporcionan detalles adicionales sobre la trasposición de ADN y técnicas relacionadas, incluyendo: “Shuffling of codon altered genes”, de Patten *et al.*, presentada el 29 de septiembre de 1998 (documento USSN n° 60/102.362), el 29 de enero de 1999 (documento USSN n° 60/117.729) y el 28 de septiembre de 1999, documento USSN n° PCT/US99/22588, titulada “Evolution of whole cells and organisms by recursive sequence recombination”, de Del Cardayre *et al.*, presentada el 5 de febrero de 1999 (documento USSN n° 60/118.813) y el 4 de junio de 1999 (documento USSN n° 09/408.392) y “Use of codon-based oligonucleotide synthesis for synthetic shuffling”, de Welch *et al.*, presentada el 28 de septiembre de 1999 (documento USSN n° 09/408.393).

Las referencias anteriormente indicadas también proporcionan detalles adicionales sobre el procedimiento de hibridación y elongación de ácidos nucleicos para llevar a cabo la recombinación de ácidos nucleicos.

En un aspecto, se utiliza un método híbrido que utiliza la trasposición génica familiar en combinación con métodos más tradicionales de recombinación basados en la trasposición. Por ejemplo, un ácido nucleico activo puede reensamblarse a partir de oligonucleótidos para que presente unas cuantas o ninguna sustitución homóloga respecto a un gen diana dado. El ácido nucleico “de esqueleto” reensamblado se trata con ADNAsas tal como en métodos estándares, y los fragmentos resultantes del corte con ADNasa se pulsan con oligonucleótidos familiar que comprenden secuencias correspondientes a regiones de identidad y diversidad de secuencia en un ácido nucleico dado. A continuación, se reensamblan los ácidos nucleicos en una biblioteca de secuencias homólogas mediante los métodos proporcionados posteriormente (por ejemplo el reensamblaje de PCR u otros métodos de reensamblaje). Este procedimiento puede resultar en un incremento del porcentaje de clones activos encontrados en comparación con los métodos sintéticos de oligonucleótidos que no incorporan la utilización de un ácido nucleico de esqueleto.

Varias publicaciones de los inventores y sus colaboradores, así como de otros investigadores de la técnica, describen técnicas que facilitan la trasposición del ADN, por ejemplo proporcionando el reensamblaje de genes a partir de fragmentos pequeños, incluyendo oligonucleótidos, según resulte relevante para la presente invención. Por ejemplo, Stemmer *et al.*, patente US n° 5.834.252, titulada “End complementary polymerasa reaction”, 1998, describen proce-

dimientos para amplificar y detectar una secuencia diana (por ejemplo en una mezcla de ácidos nucleicos), así como para ensamblar polinucleótidos grandes a partir de fragmentos. Crameri *et al.*, Nature 391:288-291, 1998, proporcionan metodologías básicas para el reensamblaje génico, al igual que Crameri *et al.*, Bio/Techniques 18(2):194-196, 1998.

5 También pueden utilizarse otros enfoques de generación de diversidad para modificar los ácidos nucleicos producidos mediante los métodos en la presente memoria, o para utilizarlos como moldes para los métodos en la presente memoria. Por ejemplo, puede introducirse diversidad adicional mediante métodos que resulten en la alteración de nucleótidos individuales o de grupos de nucleótidos contiguos o no contiguos, es decir, métodos de mutagénesis. Entre los métodos de mutagénesis se incluyen, por ejemplo, recombinación (patente PCT n° US98/05223, publicación WO n° 98/42727); mutagénesis dirigida por oligonucleótidos (para una revisión ver Smith, Ann. Rev. Genet. 19:423-462, 1985; Botstein y Shortle, Science 229:1193-1201, 1985; Carter, Biochem. J. 237:1-7, 1986; Kunkel, "The efficiency of oligonucleotide directed mutagenesis", en: Nucleic Acids & Molecular Biology, Eckstein and Lilley, editores, Springer Verlag, Berlín, 1987. Se encuentran incluidos en estos métodos la mutagénesis dirigida por oligonucleótidos (Zoller y Smith, Nucl. Acids Res. 10:6487-6500, 1982; Methods in Enzymol. 100:468-500, 1983, y Methods in Enzymol. 154:329-350, 1987), la mutagénesis de ADN modificado con fosforotioato (Taylor *et al.*, Nucl. Acids Res. 13:8749-8764, 1985; Taylor *et al.*, Nucl. Acids Res. 13:8765-8787, 1985; Nakamaye y Eckstein, Nucl. Acids Res. 14:9679-9698, 1986; Sayers *et al.*, Nucl. Acids Res. 16:803-814, 1988), la mutagénesis utilizando moldes que contienen uracilo (Kunkel, Proc. Nat'l. Acad. Sci. USA 82:488-492, 1985), la mutagénesis utilizando ADN dúplex con huecos (Kramer *et al.*, Nucl. Acids Res. 16:7207, 1988), y Fritz *et al.*, Nucl. Acids Res. 16:6987-6999, 1988). Entre los métodos adicionales se incluyen la reparación de desapareamientos puntuales (Kramer *et al.*, Cell 38:879-887, 1984), la mutagénesis utilizando cepas huésped de reparación deficiente (Carter *et al.*, Nucl. Acids Res. 13:4431-4443, 1985; Carter, Methods in Enzymol. 154:382-403, 1987), la mutagénesis por delección (Eghedarzadeh y Henikoff, Nucl. Acids Res. 14:5115, 1986), selección de restricción y purificación de restricción (Wells *et al.*, Phil. Trans. R. Soc. Lond. A317:415-423, 1986), la mutagénesis mediante síntesis de genes totales (Nambiar *et al.*, Science 223:1299-1301, 1984; Sakamar y Khorana, Nucl. Acids Res. 14:6361-6372, 1988; Wells *et al.*, Gene 34:315-323, 1984, y Grundström *et al.*, Nucl. Acids Res. 13:3305-3316, 1985). Se encuentran disponibles comercialmente los kits para la mutagénesis (por ejemplo Bio-Rad, Amersham International, Anglian Biotechnology).

30 Se proponen otros procedimientos de generación de diversidad en la patente US n° 5.756.316, en la patente US n° 5.965.408; Ostermeier *et al.*, "A combinatorial approach to hybrid enzymes independent of DNA homology", Nature Biotech. 17:1205, 1999; patentes US n° 5.783.431 y n° 5.958.672; Jirholt *et al.*, "Exploiting sequence space: shuffling *in vivo* formed complementarity determining regions into a master framework", Gene 215:471, 1998; patente US n° 5.939.250, patentes WO n° 99/10539, n° 98/58085, n° 99/10539 y otras. Los métodos de generación de diversidad pueden combinarse entre sí o con reacciones de trasposición o métodos de trasposición de oligos, en cualquier combinación seleccionada por el usuario, para producir diversidad de ácidos nucleicos, que puede cribarse para utilizar cualquier método de cribado disponible.

40 Tras la recombinación u otras reacciones de diversificación, cualesquiera ácidos nucleicos producidos pueden seleccionarse para una actividad deseada. En el contexto de la presente exposición, ésta puede incluir el ensayo e identificación de cualquier actividad detectable o ensayable, mediante cualquier ensayo relevante de la técnica. Pueden someterse a ensayo una diversidad propiedades relacionadas (o incluso no relacionadas) utilizando cualquier ensayo disponible.

#### 45 *Trasposición de ADN sin utilización de PCR*

Aunque un formato preferente de la invención para el reensamblaje génico utiliza PCR, también resultan útiles otros formatos. Por ejemplo, los métodos de mutagénesis dirigida a sitio o dirigida por oligonucleótidos pueden utilizarse para generar quimeras entre 2 ó más genes parentales (homólogos o no homólogos). A este respecto, un aspecto de la presente exposición se refiere a un nuevo método para llevar a cabo la recombinación entre ácidos nucleicos mediante ligación de bibliotecas de oligonucleótidos correspondientes a los ácidos nucleicos que deben recombinarse.

55 En este formato, se liga un conjunto de una pluralidad de oligonucleótidos que incluye una pluralidad de secuencias de ácidos nucleicos de una pluralidad de ácidos nucleicos parentales, para producir uno o más ácidos nucleicos recombinantes, típicamente codificantes de una proteína de longitud completa (aunque la ligación también puede utilizarse para construir bibliotecas de secuencias parciales de ácidos nucleicos que después pueden recombinarse, por ejemplo para producir un ácido nucleico recombinante de longitud parcial o completa). El conjunto de oligonucleótidos típicamente incluye por lo menos un primer oligonucleótido que es complementario a por lo menos al primero de los ácidos nucleicos parentales en una primera región de diversidad de secuencia, y por lo menos un segundo oligonucleótido que es complementario a por lo menos el segundo de los ácidos nucleicos parentales en una segunda región de diversidad. Los ácidos nucleicos parentales pueden ser homólogos o no homólogos.

65 Con frecuencia, se ligan ácidos nucleicos, tales como oligos, con una ligasa. En un formato típico, los oligonucleótidos se hibridan con un primer ácido nucleico parental que actúa como molde y se ligan con una ligasa. Los oligos también pueden extenderse con una polimerasa y ligarse. La polimerasa puede ser, por ejemplo, una ADN polimerasa ordinaria u otra ADN polimerasa termoestable. La ligasa también puede ser una ADN ligasa ordinaria, o una ADN ligasa termoestable. Muchas de dichas polimerasas y ligasas se encuentran comercialmente disponibles.

En un conjunto de enfoques, un elemento común para métodos de recombinación no basados en PCR es la preparación de un molde de una cadena al que se hibridan cebadores y después se elonga con una ADN polimerasa en presencia de dNTPs y un tampón apropiado. El dúplex con huecos puede sellarse con ligasa antes de la transformación o electroporación en *E. coli*. La cadena recién sintetizada se replica y genera un gen quimérico con contribuciones del oligo en el contexto de la cadena parental de una cadena (ss).

Por ejemplo, el molde ss puede prepararse mediante la incorporación de la región fágica IG en un plásmido y la utilización de un fago ayudante, tal como M13KO7 (Farmacia Biotech) o R408 para empaquetar plásmidos ss en partículas fágicas filamentosas. El molde ss también puede generarse mediante desnaturalización de un molde de doble cadena y la hibridación en presencia de los cebadores. Los métodos varían en los métodos de enriquecimiento para el aislamiento de la nueva cadena quimérica sintetizada respecto a la cadena molde parental. El aislamiento y selección de moldes de doble cadena pueden llevarse a cabo utilizando métodos disponibles. Ver, por ejemplo, Ling *et al.*, "Approaches to DNA mutagenesis: an overview", *Anal. Biochem.* 254(2):157-78, 15 de diciembre, 1997. Dale *et al.*, "Oligonucleotide-directed random mutagenesis using the phosphorothioate method", *Methods Mol. Biol.* 57:369-74, 1996; Smith, "In vitro mutagenesis", *Ann. Rev. Genet.* 19:423-462, 1985; Botstein y Shortle, "Strategies and applications of in vitro mutagenesis", *Science* 229:1193-1201, 1985, y Carter, "Site-directed mutagenesis", *Biochem. J.* 237:1-7, 1986; Kunkel, "The efficiency of oligonucleotide directed mutagenesis", *Nucleic Acids & Molecular Biology*, 1987; Eckstein F. y Lilley D.M.J. (editores), Springer-Verlag, Berlín.

Por ejemplo, en un aspecto, un método de "estilo Kunkel" utiliza moldes que contienen uracilo. De manera similar, el método "Eckstein" utiliza ADN modificado con fosforotioato (Taylor *et al.*, "The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA", *Nucleic Acids Res.* 13:8749-8764, 1985; Taylor *et al.*, "The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA", *Nucleic Acids Res.* 13:8765-8787, 1985; Nakamaye y Eckstein, "Inhibition of restriction endonuclease NciI cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis", *Nucleic Acids Res.* 14:9679-9698, 1986; Sayers *et al.*, "Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis", *Nucleic Acids Res.* 16:791-802, 1988; Sayers *et al.*, "5'-3' Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide", *Nucleic Acids Res.* 16:803-814, 1988). La utilización de selección de restricción o, por ejemplo, la purificación, puede utilizarse conjuntamente con cepas de reparación deficiente por desapareamiento (ver, por ejemplo, Carter *et al.*, "Improved oligonucleotide site directed mutagenesis using M13 vectores", *Nucleic Acids Res.* 13:4431-4443, 1985; Carter, "Improved oligonucleotide-directed mutagenesis using M13 vectores", *Methods in Enzymol.* 154:382-403, 1987; Wells, "Importance of hydrogen bond formation in stabilizing the transition state of subtilisin", *Trans. R. Soc. Lond.* A317:415-423, 1986).

El cebador "mutagénico" utilizado en estos métodos puede ser un oligonucleótido sintético codificante de cualquier tipo de oligonucleótido por aleatorización, inserción, delección, trasposición génica familiar, basada en la diversidad de secuencia de genes homólogos, etc. El cebador o cebadores también pueden ser fragmentos de genes homólogos que se hibridan con el molde ss parental. De esta manera pueden generarse quimeras entre 2 ó más genes parentales.

Pueden hibridarse múltiples cebadores a un molde dado y ser extendidos para crear genes múltiplemente quiméricos. La utilización de una ADN polimerasa tal como la del fago T4 o del fago T7 resulta adecuada para este fin debido a que no se degradan ni desplazan del molde los cebadores situados cadena abajo.

Por ejemplo, en un aspecto, la trasposición del ADN se lleva a cabo utilizando moldes que contienen uracilo. En esta realización, el gen de interés se clona en un plásmido de *E. coli* que contiene la región intergénica de fago filamentosos (IG, ori). Se empaqueta ADN de una cadena (ss) de plásmido en partículas fágicas tras la infección con un fago ayudante, tal como M13KO7 (Farmacia) o R408 y puede purificarse fácilmente mediante métodos tales como la extracción con fenol/cloroformo y la precipitación con etanol. Si este ADN se prepara en una cepa dutung- de *E. coli*, se incorpora un número pequeño de residuos uracilo en lugar de los residuos normales de timina. Se hibrida uno o más cebadores, u otros oligos tal como se ha indicado anteriormente, con el molde ss que contiene uracilos mediante el calentamiento a 90°C y enfriamiento lento hasta la temperatura ambiente. Se añade un tampón adecuado que contiene los 4 desoxirribonucleótidos, ADN polimerasa de T7 y ADN ligasa de T4 a la mezcla de molde/cebador hibridados y se incuba a una temperatura entre la temperatura ambiente y, por ejemplo, aproximadamente 37°C, durante  $\geq 1$  hora. La ADN polimerasa de T7 extiende a partir del extremo 3' del cebador y sintetiza una cadena complementaria al molde que incorpora el cebador. La ADN ligasa sella el hueco entre el extremo 3' de la cadena recién sintetizada y el extremo 5' del cebador.

En el caso de que se utilicen múltiples cebadores, la polimerasa extenderá hasta el siguiente cebador, se parará y la ligasa sellará el hueco. A continuación, esta reacción se transforma en una cepa ung<sup>+</sup> de *E. coli* y se aplica selección con antibiótico para el plásmido. El enzima uracil-N-glucosilasas (producto del gen *ung*) en la célula huésped reconoce el uracilo en la cadena de molde y la elimina, creando sitios apirimidínicos que no se replican o que los sistemas de reparación del huésped corrigen mediante la utilización de la cadena recién sintetizada como molde. Los plásmidos resultantes predominantemente contienen el cambio deseado en el gen de interés. En el caso de que se utilicen múltiples genes, resulta posible introducir simultáneamente numerosos cambios en una única reacción. En el caso de que los cebadores se deriven o correspondan a fragmentos de genes homólogos, pueden generarse genes múltiplemente quiméricos.

*Modificación de codones*

En un aspecto, los oligonucleótidos utilizados en los métodos en la presente memoria presentan un uso alterado de los codones en comparación con las secuencias parentales a partir de las que se derivan los oligonucleótidos. En particular, resulta útil, por ejemplo, modificar la preferencia de los codones para optimizar la expresión en una célula en la que debe evaluarse, o de otra manera seleccionarse, un producto recombinante de un procedimiento de trasposición de oligonucleótidos. La conformación de un ácido nucleico recombinante al sesgo de codones de una célula particular en la que tiene lugar la selección típicamente resulta en la maximización de la expresión del ácido nucleico recombinante. Debido a que los oligonucleótidos utilizados en las diversas estrategias en la presente invención típicamente se construyen sintéticamente, la selección de la preferencia óptima de codones se realiza simplemente haciendo referencia a tablas de sesgo de codones bien conocidas. Los métodos sintéticos basados en codones, tal como se describe *supra*, opcionalmente se utilizan para modificar codones en protocolos de síntesis.

Además de la selección de secuencias oligonucleótidas para optimizar la expresión, también puede utilizarse la preferencia de codones para incrementar la similitud de secuencia entre ácidos nucleicos distantemente relacionados que deben recombinarse. Mediante la selección de qué codones se utilizan en posiciones particulares, resulta posible incrementar la similitud entre ácidos nucleicos que, a su vez, incrementa la frecuencia de recombinación entre los ácidos nucleicos. Se pueden encontrar detalles adicionales sobre los procedimientos de modificación de codones y su aplicación a la trasposición del ADN en Paten y Stemmer, documento USSN n° 60/102.362, titulado “Shuffling of codon altered nucleic acids”, presentado el 29 de septiembre de 1998, y la solicitud relacionada de Paten y Stemmer, número de expediente del cesionario 018097-028510, titulada “Shuffling of codon altered nucleic acids”, presentada el 29 de enero de 1999.

*Variación de la longitud mediante trasposición modular*

Muchos dominios de secuencia funcional para genes y elementos génicos están compuestos de dominios de sub-secuencia funcional. Por ejemplo, las secuencias de promotor están constituidas de varios elementos de secuencia funcionales que se unen a factores de transcripción que, a su vez, regulan la expresión génica. Los elementos intensificadores pueden combinarse con elementos promotores para incrementar la expresión de un gen dado. De manera similar, por lo menos algunos exones representan dominios modulares de una proteína codificada, y pueden multimerizarse o delecionarse exones respecto a un gen de tipo salvaje, y los ácidos nucleicos resultantes pueden recombinarse para proporcionar bibliotecas de módulos génicos (o proteínas codificadas) alterados (es decir, bibliotecas de ácidos nucleicos de módulo insertado o delecionado). El número y disposición de las secuencias modulares, así como su composición de secuencia, pueden afectar a la actividad global del promotor, exón u otro módulo genético.

El concepto de exones como módulos de genes y proteínas codificadas se encuentra bien establecido, particularmente para proteínas que se han desarrollado en eucariotas (ver, por ejemplo, Gilbert y Glynias, *Gene* 137-144, 1993; Dolittle y Bork, *Scientific American* 50-56 (octubre de 1993), y Patthy, *Current Opinions in Structural Biology* 1:351-361, 1991. La trasposición de módulos de exones se optimiza a partir de una comprensión de las reglas de trasposición de exones. Los intrones (y en consecuencia los exones) se producen en tres etapas diferentes, dependiendo del límite de procesamiento de un codón en el límite exón-intrón (ver Stemmer, *Biotechnology* 13:549-553, 1995; Patthy, *Current Opinions in Structural Biology* 4:383-392, 1994, y Patthy, *Current Opinions in Structural Biology* 1:351-361, 1991).

En la naturaleza, los límites de procesamiento de exones traspuestos deben ser “de fase compatible” con aquellos de los exones vecinos; en caso contrario, se produce un desplazamiento del marco de lectura, eliminando la información del módulo de exones. Las tres posibles fases de un intrón son las fases 1, 2 ó 0, según la posición de la base dentro del codón en el límite intrón-exón en la que se encuentra presente el intrón. La clasificación de los intrones según su localización respecto al marco de lectura es la siguiente: un intrón de fase 0 se encuentra localizado entre dos codones de exones flanqueantes; un intrón de fase 1 se encuentra localizado entre el primer y segundo nucleótidos de un codón, y un intrón de fase 2 se encuentra localizado entre el segundo y tercer nucleótidos de un codón. Los intrones de fase 1 son los más comunes en la Naturaleza.

Un aspecto de la presente exposición es la trasposición de secuencias modulares (incluyendo, por ejemplo, elementos promotores y exones) para modificar la secuencia de dichos módulos, el número de repeticiones de los módulos (entre 0 (es decir, una delección del elemento) y un número deseado de copias) y la longitud de los módulos. En particular, los métodos estándares de traslocación y/o los métodos mediados por oligonucleótidos en la presente memoria, pueden combinarse con enfoques de duplicación de elementos y variación de la longitud de los mismos, simplemente con pulsos de fragmentos apropiadamente diseñados u oligonucleótidos en una mezcla de recombinación.

Por ejemplo, un fragmento generado por PCR que contiene el elemento que debe repetirse se pulsa en una reacción de recombinación, causando la creación de multímeros en una reacción de recombinación posterior. Estos multímeros pueden incorporarse en los productos traspuestos finales mediante recombinación homóloga en los extremos de los multímeros, siendo las longitudes globales de dichos multímeros dependientes de las proporciones molares de los módulos que deben multimerizarse. Los multímeros pueden construirse separadamente, o puede ser oligos en una reacción de reensamblaje/recombinación génicas, tal como se ha comentado *supra*.

En un aspecto preferente, se seleccionan oligos para generar multímeros y/o para deleccionar módulos seleccionados, tales como exones, elementos promotores, intensificadores o similares durante la recombinación de oligonucleótidos y el ensamblaje génico, evitando de esta manera la necesidad de construir multímeros o ácidos nucleicos que comprenden delecciones de módulos separadamente. De esta manera, en un aspecto, se construye un conjunto de oligonucleótidos solapantes de trasposición génica familiar, para que comprenda oligos que permiten la delección o multimerización de elementos de módulo de secuencias. Estos oligonucleótidos de “trasposición de módulos” pueden utilizarse conjuntamente con cualquiera de los demás enfoques en la presente memoria para recombinar ácidos nucleicos homólogos. De esta manera, los elementos de módulo de secuencias son aquellas subsecuencias de un ácido nucleico dado que proporcionan una actividad o componente diferenciado de una actividad de un ácido nucleico seleccionado, mientras que los oligonucleótidos de trasposición de módulos son oligonucleótidos que permiten la inserción, delección o multimerización de módulos de secuencias. Entre los ejemplos de dichos oligonucleótidos se incluyen aquellos que presentan subsecuencias correspondientes a más de un módulo de secuencias (permitiendo la delección de secuencias intermedias y/o la inserción de un módulo en una posición seleccionada), uno o más oligonucleótidos con extremos que presentan regiones de identidad que permiten la multimerización del oligonucleótido u oligonucleótidos (y, opcionalmente, de secuencias asociadas) durante la hibridación y elongación de una mezcla de oligonucleótidos, y similares.

Las bibliotecas resultantes de estrategias de delección/inserción de módulos indicadas anteriormente varían en el número de copias y en la disposición de un módulo dado respecto a un ácido nucleico parental correspondiente u homólogo. En el caso de que los módulos sean exones, los oligonucleótidos utilizados en los métodos e recombinación típicamente se seleccionan para resultar en la unión de exones en la misma fase (es decir, que presentan el mismo marco de lectura), para incrementar la probabilidad de que cualquier miembro dado de una biblioteca sea funcionalmente activo. Esto se ilustra esquemáticamente en la figura 3. Los módulos de diferente sombreado representan exones separados, indicando la fase del exón como 1, 2 ó 0.

#### 25 *Trasposición de intermediarios cladísticos*

La presente exposición permite la trasposición de “intermediarios evolutivos”. En el contexto de la presente exposición, los intermediarios evolutivos son constructos artificiales que presentan un carácter de intermediario entre dos o más secuencias homólogas, por ejemplo en el caso de que las secuencias se agrupen en un dendrograma evolutivo.

Los ácidos nucleicos con frecuencia se clasifican en dendrogramas (o “árboles”) evolutivos que muestran los puntos de ramificación evolutiva y, opcionalmente, el grado de parentesco. Por ejemplo, el análisis cladístico es un método de clasificación en el que se ordenan y asignan rangos los organismos y caracteres (incluyendo las secuencias de ácidos nucleicos y polipeptídicas) sobre una base que refleja el origen a partir de un ancestro común postulado (una forma intermedia de los caracteres u organismos divergentes). El análisis cladístico se refiere principalmente a la ramificación de los árboles de parentesco (o “dendrogramas”), que muestran el grado de parentesco, aunque el grado de las diferencias también puede evaluarse (una distinción realizada en ocasiones entre los taxónomos evolutivos, quienes consideran los grados de las diferencias y aquellos que simplemente determinan los puntos de ramificación en un dendrograma evolutivo (análisis cladístico clásico); para los fines de la presente exposición, sin embargo, los árboles de parentesco producidos mediante cualquiera de los dos métodos pueden producir intermediarios evolutivos).

Pueden determinarse los intermediarios cladísticos u otros intermediarios evolutivos mediante la selección de ácidos nucleicos que presentan una secuencia intermedia entre dos o más ácidos nucleicos existentes. Aunque la secuencia podría no existir en la Naturaleza, representa una secuencia que es similar a una secuencia en la Naturaleza que ha sido seleccionada, es decir, una secuencia intermedia entre dos o más secuencias representa una secuencia similar al ancestro común de los dos o más ácidos nucleicos existentes. De esta manera, los intermediarios evolutivos son un sustrato de trasposición preferente, debido a que representan secuencias “pseudoseleccionadas” que es más probable que presenten actividad que secuencias seleccionadas aleatoriamente.

Un beneficio de la utilización de intermediarios evolutivos como sustratos para la trasposición (o de la utilización de oligonucleótidos que corresponden a dichas secuencias) es que puede representarse una considerable diversidad de secuencia en menos sustratos de partida (es decir, en el caso de que se parta de las cadenas parentales A y B, un único intermediario “C” presenta por lo menos una representación parcial tanto de A como de B). Esto simplifica la síntesis de oligonucleótidos para los métodos de reconstrucción/recombinación génicas, mejorando la eficiencia del procedimiento. Además, la búsqueda de las bases de datos de secuencias con intermediarios evolutivos incrementa las probabilidades de identificar ácidos nucleicos relacionados utilizando programas de búsqueda estándares, tales como BLAST.

También pueden seleccionarse secuencias intermedias entre dos o más secuencias sintéticas que no se encuentran representadas en la Naturaleza, partiendo simplemente de dos secuencias sintéticas. Estas secuencias sintéticas pueden incluir intermediarios evolutivos, secuencias génicas que se han propuesto, u otras secuencias de interés de secuencias relacionadas. Estos “intermediarios artificiales” también resultan útiles para reducir la complejidad de los métodos de reconstrucción génica y para mejorar la capacidad de búsqueda en bases de datos evolutivas.

De acuerdo con lo anteriormente expuesto, en un aspecto significativo de la exposición, en primer lugar se determinan cadenas de caracteres que representan intermediarios evolutivos o artificiales utilizando programas informáticos de alineación y de relación entre secuencias (BLAST, PILEUP, etc.) y después se sintetizan utilizando métodos de

reconstrucción de oligonucleótidos. Alternativamente, los intermediarios pueden formar la base para la selección de oligonucleótidos utilizados en los métodos de reconstrucción génica proporcionados en la presente memoria.

#### *Trasposición de dominios de proteína*

La trasposición de genes familiar es una buena manera de acceder a la diversidad funcional de las proteínas codificadas. Puede resultar ventajoso, sin embargo, trasponer únicamente una parte de una proteína codificada que proporcione una actividad de interés, particularmente en el caso de que la proteína sea multifuncional y pueda localizarse una o más actividades en una subsecuencia (un dominio) de la proteína total.

Por ejemplo, enzimas tales como las glucosiltransferasas presentan dos sustratos: el aceptor y el donante activado de sacárido. Para modificar el azúcar que debe transferirse sin alterar el aceptor, puede resultar preferible la trasposición familiar de únicamente el dominio de unión a sacárido, debido a que la trasposición familiar del dominio aceptor de sacárido puede resultar en un menor número de moléculas del aceptor deseado.

En un ejemplo existen 5 enzimas, eA-eE (cada uno de 500 aminoácidos) que transfieren sacáridos a-e a los aceptores A-E. La generación de una biblioteca de enzimas que transfieren sacáridos a-e al aceptor A puede resultar preferible a trasladar los dominios de unión de sacárido de eA-eE, combinándolos con dominios de unión a aceptor de eA.

Un reto técnico en la práctica de dicha estrategia es que puede disponerse de una cantidad insuficiente de datos para identificar dichos dominios funcionales en una proteína de interés. En este caso puede generarse un conjunto de bibliotecas mediante la trasposición familiar de partes aleatorias del enzima. Por ejemplo, según se aplica a la trasposición familiar de enzimas eA-eE, anteriormente, puede construirse una primera biblioteca codificante de los primeros 100 aminoácidos de eA-eE, en combinación con los últimos 400 aminoácidos de cualquiera de entre eA-eE mediante la selección apropiada de conjuntos de oligonucleótidos para la recombinación y elongación. Puede construirse una segunda biblioteca de trasposición familiar de los siguientes 100 aminoácidos de eA-eE, en combinación con la codificación de los primeros 100 aminoácidos de cualquiera de entre eA-Ae y los últimos 300 aminoácidos de cualquiera de entre eA-Ae, y de esta manera sucesivamente. Se criban subconjuntos pequeños de estas bibliotecas para una primera función deseada. Las bibliotecas que conservan la primera función deseada (por ejemplo actividad aceptor en el ejemplo anterior) presentan una proporción relativamente más alta de variantes en funciones seleccionables adicionales (por ejemplo la transferencia de sacáridos en el ejemplo anterior).

Este enfoque puede utilizarse para la diversificación de cualquier proteína multifuncional en la que una propiedad se encuentre deseablemente conservada. Esta estrategia resulta particularmente ventajosa en el caso de que la propiedad que deba conservarse sea compleja (por ejemplo la especificidad de sustrato para, por ejemplo, poliquétidos, péptidos no ribosómicos u otros productos naturales).

En general, la selección de oligonucleótidos para proporcionar la trasposición de dominios individuales (correspondientes a subsecuencias funcionales conocidas o a subsecuencias de función desconocida, tal como se ha indicado anteriormente) se lleva a cabo proporcionando dos tipos generales de oligonucleótido de secuencias relacionadas. El primer tipo se proporciona mediante la selección de conjuntos de oligonucleótidos solapantes de secuencias relacionadas correspondientes a regiones en las que se desea la recombinación (es decir, según las estrategias indicadas en la presente memoria), mientras que el segundo tipo proporciona uniones de recombinación entre los dominios que deben trasponerse y dominios no traspuestos, es decir, de manera similar a un oligonucleótido entrecruzado tal como se describe en la presente memoria. Los dominios no traspuestos pueden producirse mediante métodos simples de reconstrucción génica de oligonucleótidos (por ejemplo mediante reacciones de ligación o de extensión mediada por polimerasa para concatenar oligonucleótidos) o los dominios no traspuestos pueden producirse mediante corte enzimático de ácidos nucleicos más grandes.

#### *Trasposición familiar expandida que incorpora modelaje molecular y escaneo de alaninas*

La trasposición de oligos de tipo familiar implica la recombinación de ácidos nucleicos homólogos mediante la construcción de conjuntos de oligonucleótidos de trasposición familiar que se recombinan en protocolos de síntesis y recombinación génicas, tal como se ha comentado *supra*. Tal como se ha indicado, los ácidos nucleicos homólogos pueden ser homólogos naturales o no naturales (es decir, artificiales).

Una ventaja de recombinar los homólogos no naturales es que los ácidos nucleicos recombinantes acceden a espacio de secuencia diferente del natural. Esta diversidad adicional permite el desarrollo u obtención de propiedades funcionales no proporcionadas por la recombinación de ácidos nucleicos que representan la diversidad natural.

La desventaja principal de crear homólogos aleatorios para la recombinación es que muchos de los homólogos resultantes no son funcionales con respecto a una característica relevante. Para estos homólogos, una gran parte del incremento resultante del espacio de secuencia seleccionable es "ruido" no deseable que debe excluirse de la población. En contraste, la diversidad natural representa moléculas sometidas a la prueba de la evolución, representando un espacio global de secuencia potencial más focalizado en el que se produce la recombinación.

Una manera de capturar diversidad no natural sin incrementar significativamente espacio de secuencia no deseable es definir aquellas posiciones que pueden modificarse en un gen dado sin degradar significativamente la propiedad

funcional deseada de una molécula codificada (proteína, ARN, etc.). Pueden utilizarse por lo menos dos enfoques básicos.

5 En primer lugar, la mutagénesis puntual (por ejemplo el escaneo de alaninas) para definir posiciones que pueden mutarse sin una pérdida significativa de función. En principio, podrían someterse a ensayo la totalidad de los 20 aminoácidos en cada posición, definiendo un gran espectro de mutaciones puntuales que son esencialmente neutras con respecto a la función. A continuación, se construyen conjuntos de oligos de trasposición que capturan dichos homólogos no naturales (aunque todavía activos). Para muchas proteínas comercialmente importantes, la información del escaneo de alaninas ya se encuentra disponible. Por ejemplo, Young *et al.*, Protein Science 6:1228-1236, 1997, describe el escaneo de alaninas del factor de estimulación de colonias de granulocitos (G-CSF).

15 En segundo lugar, en donde se disponga de información estructural para una proteína (y, por ejemplo, de cómo interacciona la proteína con un ligando), pueden definirse regiones que se predice que son mutables con poco o ningún cambio de función. A continuación, se construyen conjuntos de oligos de trasposición familiar que capturan estos homólogos no naturales (aunque predeciblemente todavía activos). Se encuentra disponible una diversidad de estructuras cristalinas de proteínas (entre ellas, por ejemplo, la estructura cristalina de G-CSF; Hill *et al.*, PNAS 90:5167, 1993).

20 De manera similar, aunque no se disponga de información estructural, puede llevarse a cabo modelaje molecular para proporcionar una estructura predicha, que también puede utilizarse para predecir qué residuos pueden modificarse sin alterar la función. Se encuentra disponible una diversidad de programas de modelaje de la estructura de proteínas para predecir la estructura de proteínas. Además, las tendencias relativas de los aminoácidos para formar regiones de superestructura (hélices, láminas  $\beta$ , etc.) se encuentran bien establecidas. Por ejemplo, O'Neil y DeGrado, Science volumen 250, proporcionan un comentario sobre las tendencias de formación de hélice de los aminoácidos comunes. Pueden utilizarse las tablas de actividad relativa de formación de estructura de los aminoácidos a modo de tablas de sustitución para predecir qué residuos pueden sustituirse funcionalmente en una parte dada. A continuación, se construyen conjuntos de oligos de trasposición familiar que capturan estos homólogos no naturales (aunque predeciblemente activos).

30 Por ejemplo, la automatización del diseño de proteínas (PDA) es un sistema controlado por ordenador para el diseño y optimización de proteínas y péptidos, así como para el diseño de proteínas y péptidos. Típicamente, la PDA se inicia con una estructura de esqueleto de una proteína y diseña la secuencia de aminoácidos para modificar las propiedades de la proteína, manteniendo simultáneamente sus propiedades de plegamiento tridimensional. Puede manipularse un gran número de secuencias utilizando PDA, permitiendo el diseño de estructuras de proteína (secuencias, subsecuencias, etc.). La PDA se describe en varias publicaciones, incluyendo, por ejemplo, Malakauskas y Mayo, "Design, Structure and Stability of a Hyperthermophilic Protein Variant", Nature Struct. Biol. 5:470, 1998; Dahiyat y Mayo, "De Novo Protein Design: Fully Automated Sequence Selection", Science 278:82-87, 1997; DeGrado, "Proteins from Scratch", Science 278:80-81, 1997; Dahiyat, Sarisky y Mayo, "De Novo Protein Design: Towards Fully Automated Sequence Selection", J. Mol. Biol. 273:789-796, 1997; Dahiyat y Mayo, "Probing the Role of Packing Specificity in Protein Design", Proc. Natl. Acad. Sci. USA 94:10172-10177, 1997; Hellinga, "Rational Protein Design Combining Theory and Experiment", Proc. Natl. Acad. Sci. USA 94:10015-10017, 1997; Su y Mayo, "Coupling Backbone Flexibility and Amino Acid Sequence Selection in Protein Design", Prot. Sci. 6:1701-1707, 1997; Dahiyat, Gordon y Mayo, "Automated Design of the Surface Positions of Protein Helices", Prot. Sci. 6:1333-1337, 1997; Dahiyat y Mayo, "Protein Design Automation", Prot. Sci. 5:895-903, 1996. Se encuentran disponibles detalles adicionales sobre la PDA, por ejemplo en <http://www.xencor.com/>. Puede utilizarse la PDA para identificar variantes de una secuencia que es probable que conserven actividad, proporcionando un conjunto de ácidos nucleicos homólogos que pueden utilizarse como base para la recombinación mediada por oligonucleótidos.

#### Técnica de cribado posterior a la recombinación

50 El método exacto de recombinación que se utilice en los diversos procedimientos de trasposición en la presente memoria no es un aspecto crítico de la exposición. En general, el experto en la materia podrá poner en práctica métodos apropiados de cribado (es decir, de selección), en referencia a la actividad que se selecciona.

55 En cualquier caso, tras uno o más ciclos de recombinación habitualmente sigue uno o más ciclos de cribado o selección para moléculas o células transformadas u organismo que presenten una propiedad, carácter o característica deseado. Si se lleva a cabo un ciclo de recombinación *in vitro*, los productos de recombinación, es decir, los segmentos recombinantes, en ocasiones se introducen en las células antes de la etapa de cribado. Los segmentos recombinantes también pueden unirse a un vector apropiado o a otras secuencias reguladoras antes del cribado. Alternativamente, en ocasiones se empaquetan en virus (por ejemplo bacteriófagos) los productos de recombinación generados *in vitro* antes del cribado. Si se lleva a cabo la recombinación *in vivo*, los productos de recombinación en ocasiones pueden cribarse en las células en las que se produjo la recombinación. En otras aplicaciones, se extraen los segmentos recombinantes de las células y opcionalmente se empaquetan como virus, antes del cribado.

65 La naturaleza del cribado o selección depende de la propiedad o característica que debe adquirirse, o de la propiedad o característica para la que se desea una mejora, y se comentan posteriormente muchos ejemplos. Habitualmente no resulta necesario comprender la base molecular por la que productos de recombinación particulares (segmentos recombinantes) han adquirido propiedades o características nuevas o mejoradas respecto a los sustratos de partida. Por

ejemplo, un gen puede presentar muchas secuencias-componente, presentando cada una, un papel pretendido diferente (por ejemplo secuencia codificante, secuencias reguladoras, secuencias de reconocimiento, secuencias que proporcionan estabilidad, secuencias de subunidad y secuencias que afectan a la integración). Cada una de estas secuencias-componente puede modificarse y recombinarse simultáneamente. A continuación, puede llevarse a cabo un cribado/selección, por ejemplo, para segmentos recombinantes que presentan una capacidad incrementada de proporcionar actividad a una célula sin la necesidad de atribuir dicha mejora a ninguna de las secuencias-componente individuales del vector.

Dependiendo del protocolo de cribado particular que se utilice para una propiedad deseada, en ocasiones pueden llevarse a cabo una o más rondas iniciales de cribado utilizando células bacterianas, debido a sus elevadas eficiencias de transfección y facilidad de cultivo. Sin embargo, la expresión bacteriana con frecuencia no resulta práctica o deseable, por lo que también se utilizan sistemas de levadura, fúngicos u otros sistemas eucarióticos para la expresión y cribado de bibliotecas. De manera similar, se llevan a cabo otros tipos de cribado que no pueden realizarse en bibliotecas de células bacterianas o eucarióticas simples, en células seleccionadas para la utilización en un ambiente similar al de su uso pretendido. Pueden llevarse a cabo rondas finales de cribado en el tipo celular exacto del uso pretendido.

Un enfoque para el cribado de bibliotecas diversas es utilizar un procedimiento de fase sólida masivamente en paralelo para cribar productos de ácidos nucleicos traspuestos, por ejemplo enzimas codificados, para actividad incrementada. Se encuentra disponible un aparato de cribado en fase sólida masivamente en paralelo utilizando absorción, fluorescencia o FRET (ver, por ejemplo, la patente US nº 5.914.245, de Bylina *et al.*, 1999; ver también <http://www.kairos-scientific.com/>; Youvan *et al.*, "Fluorescence Imaging Micro-Spectrophotometer (FIMS)", *Biotechnology et alia* <[www.et-al.com](http://www.et-al.com)> 1:1-16; Yang *et al.*, "High Resolution Imaging Microscope (HIRIM)", *Biotechnology et alia*, <[www.et-al.com](http://www.et-al.com)>4:1-20, y Youvan *et al.*, "Calibration of Fluorescence Resonance Energy Transfer in Microscopy Using Genetically Engineered GFP Derivatives on Nickel Chelating Beads", enviado a [www.kairos-scientific.com](http://www.kairos-scientific.com), 1999). Tras el cribado mediante estas técnicas, típicamente se aíslan secuencias de interés, opcionalmente se secuencian y las secuencias utilizadas se proporcionan en la presente memoria para diseñar nuevos métodos de trasposición de oligonucleótidos.

Si se desea una mejora adicional en una propiedad, por lo menos un segmento, y habitualmente una colección de segmentos recombinantes que sobreviven una primera ronda de cribado/selección se someten a una ronda posterior de recombinación. Estos segmentos recombinantes pueden recombinarse entre sí o con segmentos exógenos que representan los sustratos originales o variantes adicionales de los mismos. Nuevamente, la recombinación puede producirse *in vitro* o *in vivo*. En el caso de que la etapa de cribado anterior identifique segmentos recombinantes deseados como componentes de células, los componentes pueden someterse a recombinación adicional *in vivo*, o pueden someterse a recombinación adicional *in vitro*, o pueden aislarse antes de llevar a cabo una ronda de recombinación *in vitro*. A la inversa, en el caso de que la etapa de cribado anterior identifique segmentos recombinantes deseados en forma desnuda o como componentes de virus, estos segmentos pueden introducirse en células para llevar a cabo una ronda de recombinación *in vivo*. La segunda ronda de recombinación, con independencia de cómo se lleva a cabo, genera segmentos recombinantes adicionales que comprenden diversidad adicional a la presente en segmentos recombinantes resultantes de rondas anteriores.

A la segunda ronda de recombinación puede seguir una ronda adicional de cribado/selección según los principios comentados anteriormente para la primera ronda. La astringencia de cribado/selección puede incrementarse entre rondas. Además, la naturaleza del cribado y la propiedad para la que se criba pueden variar entre rondas si se desea la mejora de más de una propiedad o si se desea adquirir más de una nueva propiedad. En este caso pueden llevarse a cabo rondas adicionales de recombinación y cribado hasta que los segmentos recombinantes hayan evolucionado suficientemente para adquirir la nueva o mejorada propiedad o función deseada.

#### *Procedimientos posteriores a la trasposición*

Los ácidos nucleicos producidos mediante los métodos de la invención opcionalmente se clonan en células para el cribado de actividad (o se utiliza en reacciones de transcripción *in vitro* para producir productos que se criban). Además, los ácidos nucleicos pueden secuenciarse, expresarse, amplificarse *in vitro* o tratarse en cualquier otro método recombinante común.

Entre los textos generales que describen técnicas de biología molecular que resultan útiles en la presente invención, incluyendo la clonación, la mutagénesis, la construcción de bibliotecas, los ensayos de cribado, el cultivo celular y similares, se incluyen Berger y Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology*, vol. 152, Academic Press, Inc., San Diego, CA (Berger); Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual* (2a edición), vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989 ("Sambrook") y *Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, editores, Current Protocols, un proyecto conjunto de Greene Publishing Associates, Inc. y John Wiley & Sons, Inc. (suplementado hasta 1998) ("Ausubel"). Los métodos de transducción de células, incluyendo las células vegetales y animales, con ácidos nucleicos se encuentran generalmente disponibles, al igual que los métodos de expresión de proteínas codificadas por dichos ácidos nucleicos. Además de Berger, Ausubel y Sambrook, entre las referencias generales útiles para el cultivo de células animales se incluyen Freshney (*Culture of Animal Cells, a Manual of Basic Technique*, tercera edición, Wiley-Liss, New York, 1994) y las referencias citadas en la misma; Humason (*Animal Tissue Techniques*, cuarta edición, W.H. Freeman and Company, 1979), y Ricciardelli *et al.*, *In vitro Cell Dev. Biol.* 25:1016-1024, 1989). Entre las referencias para

la clonación, cultivo y regeneración de células vegetales se incluyen Payne *et al.*, Plant Cell and Tissue Culture in Liquid Systems, John Wiley & Sons, Inc., New York, NY, 1992 (Payne), y Gamborg y Phillips (editores), Plant Cell, Tissue and Organ Culture, Fundamental Methods, Springer Lab. Manual, Springer-Verlag (Berlin, Heidelberg, New York), 1995, (Gamborg). Se describe una diversidad de medios de cultivo celular en Atlas y Parks (editores), The Handbook of Microbiological Media, CRC Press, Boca Raton, FL, 1993 (Atlas). Se puede encontrar información adicional para el cultivo de células vegetales en la literatura comercial disponible, tal como el catálogo Life Science Research Cell Culture Catalogue (1998), de Sigma-Aldrich, Inc. (St. Louis, MO) (Sigma-LSRCCC) y, por ejemplo, el catálogo Plant Culture Catalogue y suplemento, 1997, también de Sigma-Aldrich, Inc. (St. Louis, MO) (Sigma-PCCS).

Entre los ejemplos de técnicas suficientes para dirigir a expertos en la materia a través de los métodos de amplificación *in vitro*, que resultan útiles, por ejemplo, para amplificar ácidos nucleicos de trasposición de oligonucleótidos, se incluyen la reacción en cadena de la polimerasa (PCR), la reacción en cadena de ligasa (LCR), la amplificación con replicasa Q $\beta$  y otras técnicas mediadas por ARN polimerasa (por ejemplo NASBA). Estas técnicas pueden encontrarse en Berger, Sambrook y Ausubel, *id.*, así como en Mullis *et al.*, patente US n° 4.683.202, 1987; PCR Protocols: A Guide to Methods and Applications (Innis *et al.*, editores), Academic Press, Inc., San Diego, CA, 1990 (Innis); Arnheim y Levinson (1 de octubre de 1990) C&EN 36-47; The Journal Of NIH Research 3:81-94, 1991; Kwoh *et al.*, Proc. Natl. Acad. Sci. USA 86:1173, 1989; Guatelli *et al.*, Proc. Natl. Acad. Sci. USA 87:1874, 1990; Lomell *et al.*, J. Clin. Chem. 35:1826, 1989; Landegren *et al.*, Science 241:1077-1080, 1988; Van Brunt, Biotechnology 8:291-294, 1990; Wu y Wallace, Gene 4:560, 1989; Barringer *et al.*, Gene 89:117, 1990, y Sooknanan y Malei, Biotechnology 13:563-564, 1995. Se describen métodos mejorados de clonación *in vitro* de ácidos nucleicos amplificados en Wallace *et al.*, patente US n° 5.426.039. Los métodos mejorados de amplificación de grandes ácidos nucleicos mediante PCR se resumen en Cheng *et al.*, Nature 369:684-685, 1994, y referencias citadas en la misma, en la que se generan amplicones por PCR de hasta 40 kb. El experto en la materia apreciará que esencialmente cualquier ARN puede convertirse en un ADN de doble cadena adecuado para la digestión de restricción, la expansión por PCR y la secuenciación utilizando transcriptasa inversa y una polimerasa (ver Ausubel, Sambrook y Berger, todos *supra*). En un método preferente, se comprueban las secuencias reensambladas para la incorporación de oligonucleótidos de trasposición génica familiar. Esto puede realizarse mediante clonación y secuenciación de los ácidos nucleicos y/o mediante digestión de restricción, por ejemplo tal como esencialmente se enseña en Sambrook, Berger y Ausubel, anteriormente. Además, las secuencias pueden amplificarse por PCR y secuenciarse directamente. De esta manera, además de, por ejemplo, Sambrook, Berger, Ausubel e Innis (*id.* y *anteriormente*), también resultan particularmente útiles metodologías de secuenciación por PCR adicionales. Por ejemplo, se ha llevado a cabo la secuenciación directa de amplicones generados mediante PCR mediante la incorporación selectiva de nucleótidos boronados resistentes a nucleasas en los amplicones durante la PCR, y la digestión de los amplicones con una nucleasa para producir fragmentos de molde de diferentes tamaños (Porter *et al.*, Nucleic Acids Research 25(8):1611-1617, 1997). En los métodos, se llevan a cabo 4 reacciones de pPCR en un molde, en cada una de las cuales uno de los nucleótidos trifosfato en la mezcla de reacción de PCR se sustituye parcialmente con un 2'-desoxinucleósido 5'-[P-borano]-trifosfato. El nucleótido boronado se incorpora aleatoriamente en productos de PCR en diversas posiciones a lo largo del amplicón de PCR en un conjunto anidado de fragmentos de PCR del molde. Se utiliza una exonucleasa que se encuentra bloqueada por nucleótidos boronados incorporados para cortar los amplicones de PCR. Los amplicones cortados seguidamente se separan según tamaño utilizando electroforesis en gel de poliacrilamida, proporcionando la secuencia de los amplicones. Una ventaja de este método es que utiliza menos manipulaciones bioquímicas que la secuenciación estándar de estilo Sanger de los amplicones de PCR.

#### 45 *Trasposición in silico*

La trasposición "*in silico*" utiliza algoritmos informáticos para llevar a cabo una trasposición "virtual" utilizando operadores genéticos en un ordenador. Según se aplica a la presente exposición, se recombinan cadenas de secuencia génica en un sistema informático y se construyen productos deseables, por ejemplo mediante PCR de reensamblaje de oligonucleótidos sintéticos, tal como se describe en la presente memoria.

Brevemente, se utilizan operadores genéticos (algoritmos que representan sucesos genéticos dados, tales como mutaciones puntuales, la recombinación de dos cadenas de ácidos nucleicos homólogos, etc.) para modelar sucesos de recombinación o mutación que pueden producirse en uno o más ácidos nucleicos, por ejemplo mediante alineación de cadenas de secuencia de ácidos nucleicos (utilizando programas informáticos estándares de alineación, o mediante inspección manual y alineación), tales como aquéllas que representan ácidos nucleicos homólogos y que predicen resultados de recombinación. Los resultados de recombinación predichos se utilizan para producir moléculas correspondientes, por ejemplo mediante síntesis de oligonucleótidos y PCR de reensamblaje.

#### 60 *Ensayos integrados y elementos de sistema integrado*

Tal como se indica en la totalidad de la presente memoria, una característica de la presente invención es la alineación de ácidos nucleicos utilizando un ordenador y programas informáticos de alineación de secuencias. De manera similar, pueden utilizarse ordenadores que presenten programas informáticos apropiados para llevar a cabo la trasposición "*in silico*" previamente a la síntesis física de los oligonucleótidos. Además, otros componentes importantes del sistema integrado pueden proporcionar ensayos de cribado de alto rendimiento, así como el acoplamiento de dichos ensayos a la selección, síntesis y recombinación de oligonucleótidos.

Evidentemente el ensayo relevante dependerá de la aplicación. Se conocen muchos ensayos para proteínas, receptores, ligandos y similares. Entre los formatos se incluyen la unión a componentes inmovilizados, la viabilidad de una célula u organismo, la producción de composiciones informadoras, y similares.

5 En los ensayos de alto rendimiento, resulta posible cribar hasta varios miles de variantes de trasposición diferentes en un solo día. En particular, puede utilizarse cada pocillo de una placa de microtitulación para llevar a cabo un ensayo separado o, si deben observarse efectos de la concentración o del periodo de incubación, cada 5 a 10 pocillos pueden someter a ensayo una única variante. De esta manera, un sola placa estándar de microtitulación puede someter a ensayo aproximadamente 100 (por ejemplo 96) reacciones. Si se utilizan placas de 1.536 pocillos, una sola placa fácilmente puede someter a ensayo entre aproximadamente 100 y aproximadamente 1.500 reacciones diferentes. Resulta posible someter a ensayo varias placas diferentes en un día; los cribados de ensayo de entre aproximadamente 6.000 y 20.000 ensayos diferentes (es decir, que implican diferentes ácidos nucleicos, proteínas codificadas, concentraciones, etc.) resultan posibles utilizando los sistemas integrados de la invención. Más recientemente, se han desarrollado enfoques de microfluidos para la manipulación de reactivos, por ejemplo por Caliper Technologies (Mountain View, CA).

15 En un aspecto, se separan miembros de una biblioteca, por ejemplo células, placas víricas, esporas o similares, sobre medios sólidos para producir colonias (o placas) individuales. Mediante la utilización de un recolector automático de colonias (por ejemplo el Q-bot, Genetix, Reino Unido), se identifican colonias o placas, se recolectan y se inoculan hasta 10.000 mutantes diferentes en placas de microtitulación de 96 pocillos que contienen dos bolas de vidrio de 3 mm/pocillo. El Q-bot no recolecta una colonia entera sino que inserta una aguja a través del centro de la colonia y sale con una muestra pequeña de células (o micelios) y esporas (o virus en aplicaciones de placas). El tiempo durante el que la aguja se encuentra dentro de la colonia, el número de inserciones para inocular el medio de cultivo y el tiempo durante el que la aguja se encuentra dentro de dicho medio, afectan al tamaño del inóculo, y cada uno de ellos puede controlarse y optimizarse. El procedimiento uniforme del Q-bot reduce el error debido a la manipulación humana e incrementa la tasa de cultivos establecidos (aproximadamente 10.000/4 horas). A continuación, estos cultivos se agitan en un incubador de temperatura y humedad controladas. Las bolas de vidrio en las placas de microtitulación actúan estimulando la aireación uniforme de las células y la dispersión de los fragmentos miceliares de manera similar a las palas de un fermentador. Los clones de los cultivos de interés pueden clonarse mediante dilución limitante. Tal como también se ha descrito *supra*, las placas o células que constituyen bibliotecas también pueden cribarse directamente para la producción de proteínas, mediante la detección de hibridación, actividad de una proteína, unión de una proteína a anticuerpos, o similar.

También se han desarrollado varios sistemas robóticos bien conocidos para químicas de fase solución que resultan útiles en sistemas de ensayo. Entre estos sistemas se incluyen estaciones de trabajo automáticas como el aparato automático de síntesis desarrollado por Takeda Chemical Industries, Ltd. (Osaka, Japón) y muchos sistemas robóticas que utilizan brazos robóticos (Zymate II, Zymark Corporation, Hopkinton, Mass.; Orca, Beckman Coulter, Inc. (Fullerton, CA)) que imitan las operaciones sintéticas manuales realizadas por un científico. Cualquiera de los dispositivos anteriormente indicados resulta adecuado para la utilización con la presente invención, por ejemplo para el cribado de alto rendimiento de moléculas ensambladas a partir de los diversos conjuntos de oligonucleótidos indicados en la presente memoria. La naturaleza e implementación de modificaciones en estos dispositivos (si se lleva a cabo alguna) de manera que puedan funcionar tal como se ha comentado en la presente memoria haciendo referencia al sistema integrado, resultarán evidentes para el experto en la materia relevante.

45 Los sistemas de cribado de alto rendimiento se encuentran comercialmente disponibles (ver, por ejemplo, Zymark Corp., Hopkinton, MA; Air Technical Industries, Mentor, OH; Beckman Instruments, Inc., Fullerton, CA; Precision Systems, Inc., Natick, MA, etc.). Estos sistemas típicamente automatizan procedimientos enteros, incluyendo la totalidad del pipeteado de muestras y reactivos, la dispensación de líquidos, las incubaciones temporizadas y las lecturas finales de la microplaca en uno o más detectores apropiados para el ensayo. Estos sistemas configurables proporcionan un alto rendimiento y rápido inicio, así como un elevado grado de flexibilidad y adaptación al usuario. Los fabricantes de dichos sistemas proporcionan protocolos detallados. De esta manera, Zymark Corp. proporciona boletines técnicos que describen sistemas de cribado para detectar la modulación de la transcripción génica, la unión de ligandos y similares.

55 Las imágenes ópticas visionadas (y, opcionalmente, registradas) por una cámara u otro dispositivo de registro (por ejemplo un fotodiodo y un dispositivo de almacenamiento de datos) opcionalmente se procesan adicionalmente en cualquiera de las realizaciones en la presente memoria, por ejemplo mediante digitalización de la imagen y/o almacenamiento y análisis de la imagen en un ordenador. Se encuentra disponible una diversidad de equipos periféricos comercialmente disponibles y programas informáticos para digitalizar, almacenar y analizar un vídeo digitalizado o una imagen óptica digitalizada, por ejemplo utilizando ordenadores PC (máquinas de chip Intel x86 o Pentium compatible con DOS<sup>TM</sup>, OS2<sup>TM</sup>, WINDOWS<sup>TM</sup>, WINDOWS NT<sup>TM</sup> o WIDOWS95<sup>TM</sup>), MACINTOSH<sup>TM</sup> o basados en UNIX (por ejemplo la estación de trabajo SUN<sup>TM</sup>). Un sistema convencional lleva luz del dispositivo de ensayo a una cámara con dispositivo de carga fría acoplada (CCD), de uso común en la técnica. Una cámara CCD incluye una matriz de elementos gráficos (píxels). La luz del espécimen forma una imagen sobre el CCD. Se muestren píxels particulares correspondientes a regiones del espécimen (por ejemplo sitios individuales de hibridación en una matriz de polímeros biológicos) para obtener lecturas de intensidad lumínica para cada posición. Se procesan múltiples píxels en paralelo para incrementar la velocidad. El aparato y los métodos de la invención se utilizan fácilmente para visionar cualquier muestra, por ejemplo mediante técnicas de microscopía fluorescente o de campo oscuro.

Entre los sistemas integrados para el análisis de ensayo en la presente invención típicamente se incluyen un ordenador digital con programas informáticos de alineación de secuencias y uno o más de entre: programas informáticos de control de líquidos de alto rendimiento, programas informáticos de análisis de imágenes, programas informáticos de interpretación de datos, y similares.

5 Una bobina de armadura de control robótico de líquidos para transferir soluciones de una fuente a un destino puede unirse operablemente al ordenador digital y puede utilizarse un dispositivo de entrada (por ejemplo un teclado de ordenador) para introducir datos en el ordenador digital para controlar la transferencia de alto rendimiento de líquidos, la síntesis de oligonucleótidos y similares, por ejemplo por la bobina de armadura de control robótico de líquidos. Puede utilizarse un escáner de imágenes para digitalizar las señales de marcaje procedentes del componente de ensayo marcado. El escáner de imágenes se conecta con el programa informático de análisis de imágenes para proporcionar una medición de la intensidad de marcaje de la sonda.

15 Evidentemente estos sistemas de ensayo también pueden incluir sistemas integrados que incorporen elementos de selección de oligonucleótidos, tales como un ordenador, una base de datos con secuencias de ácidos nucleicos de interés, programas informáticos de alineación de secuencias y programas informáticos de selección de oligonucleótidos. Además, estos programas informáticos pueden incluir componentes para ordenar los oligonucleótidos seleccionados y/o para dirigir la síntesis de oligonucleótidos realizada por un aparato sintetizador de oligonucleótidos conectado operablemente. De esta manera, los elementos del sistema integrado tal como se da a conocer en la presente memoria opcionalmente incluyen cualquiera de los componentes anteriormente indicados para facilitar la recombinación y selección de alto rendimiento. Se apreciará que estos elementos de recombinación de alto rendimiento pueden encontrarse en sistemas que se encuentren separados de aquellos que realizan los ensayos de selección, o los dos pueden encontrarse integrados.

25 En un aspecto, la presente exposición comprende un ordenador o un medio legible por ordenador con un juego de instrucciones para seleccionar un conjunto de oligonucleótidos utilizando los métodos descritos en la presente memoria. El juego de instrucciones alinea ácidos nucleicos homólogos con el fin de identificar las regiones de similitud y las regiones de diversidad (por ejemplo tal como en los programas informáticos de alineación típicos, tales como BLAST) y después selecciona un conjunto de oligonucleótidos solapantes que comprende las regiones de similitud y de diversidad, opcionalmente utilizando cualquiera de los factores de ponderación indicados en la presente memoria (por ejemplo la selección predominante de oligonucleótidos correspondientes a uno o más ácidos nucleicos que deben recombinarse, tal como en los métodos de mezcla génica en la presente memoria). El ordenador o medio legible por ordenador opcionalmente comprende características que facilitan la utilización por parte de un usuario, por ejemplo un campo de entrada para que el usuario introduzca selecciones de oligonucleótidos, un sistema de visualización de los resultados para controlar los resultados visionables por el usuario (por ejemplo un GUI), un archivo de resultados que dirige la síntesis de los oligonucleótidos, por ejemplo en un sintetizador automático, y similares.

#### Ejemplo

##### *Trasposición de oligos de diferentes bibliotecas*

Una ventaja de los métodos de trasposición mediada por oligonucleótidos es la capacidad de recombinar ácidos nucleicos entre bibliotecas de oligos generadas para varios sitios diferentes en un gen de interés. La generación de bibliotecas con combinaciones complejas de aleatorizaciones en diferentes regiones de un gen diana se ve facilitada por los enfoques de trasposición mediada por oligonucleótidos.

50 Por ejemplo, el sitio de unión a antígeno de un anticuerpo o de un fragmento de anticuerpo, tal como un Fv de una cadena (scFv) o un Fab comprende principalmente 6 regiones determinantes de complementariedad (CDRs). Estas CDRs se encuentran en la misma cara de la molécula correctamente plegada, pero se encuentran separadas en la secuencia génica lineal. Los oligonucleótidos sintéticos, o aquellos generados mediante PCR, de uno o más genes de anticuerpo pueden utilizarse para generar diversidad de secuencia en CDRs individuales. Este procedimiento puede repetirse con una segunda CDR, después una tercera, hasta formar una biblioteca de anticuerpos diversos. Los formatos de trasposición de ADN presentan una ventaja clara en que permiten generar simultáneamente bibliotecas de cada CDR, produciendo con elevada frecuencia sucesos de recombinación entre CDRs que potencialmente generan todas las posibles combinaciones de diferentes CDRs. La trasposición y cribado recursivos de ADN para un carácter o propiedad mejorado pueden utilizarse para optimizar la función de la proteína.

60 De manera similar, las estructuras tridimensionales de muchas citoquinas comparten una estructura de agrupación de 4 hélices con largos bucles conectores. Se han determinado los sitios de unión a receptor de algunas de estas proteínas y se localizan en 2 ó más regiones de la proteína que se encuentran separados en la secuencia génica lineal. El modelaje de proteínas relacionadas podría utilizarse para predecir regiones funcionales de proteínas desconocidas para la selección de bibliotecas. Las bibliotecas en cada una de estas regiones pueden generarse utilizando oligos sintéticos, oligos de trasposición familiar, fragmentos de genes homólogos, o combinaciones de los mismos, tal como en la presente memoria. La trasposición mediada por oligonucleótidos permite generar bibliotecas en cada una de estas regiones simultáneamente y generar recombinantes entre cada biblioteca. De esta manera, pueden cribarse para una función mejorada las combinaciones entre miembros de cada familia. Estos aislados con función mejorada seguidamente pueden someterse a rondas sucesivas de trasposición del ADN. De esta manera, pueden seleccionarse de cada

## ES 2 341 217 T3

biblioteca los aislados que presentan la actividad más alta y seleccionarse sinergias potenciales entre los miembros de diferentes bibliotecas. Otros métodos que optimizan cada biblioteca independientemente podrían no conseguir aislar estas interacciones sinérgicas.

5 Otro ejemplo es la trasposición de enzimas en los que el sitio activo y el sitio o sitios de unión al sustrato comprenden residuos próximos entre sí, aunque separados en la secuencia lineal del gen. La trasposición del ADN puede generar simultáneamente bibliotecas de cada región que interacciona con el sustrato. La trasposición del ADN también permite generar todas las posibles combinaciones de cambios entre cada biblioteca y evaluarlas para la presencia de un carácter o propiedad mejorado.

10 Los métodos dados a conocer en la presente memoria pueden utilizarse de diferentes maneras, entre ellas las proporcionadas a continuación.

15 Un sistema integrado para seleccionar oligonucleótidos de trasposición familiar (por ejemplo mediante un procedimiento que incluye la alineación de secuencias de ácidos nucleicos parentales) y para el análisis de ácidos nucleicos traspuestos para la identificación de actividades, también en un procedimiento iterativo.

20 Un ensayo, kit o sistema que aplica una utilización de cualquiera de entre: estrategias de selección, materiales, componentes, métodos o sustratos indicados anteriormente en la presente memoria. Los kits opcionalmente comprenden además instrucciones para llevar a cabo métodos o ensayos, materiales de empaquetamiento, uno o más recipientes que contienen componentes de ensayo, un dispositivo o sistema, o similares.

25 En un aspecto adicional, se dan a conocer en la presente memoria kits a modo de realizaciones de los métodos y aparato de la presente invención. Los kits de la exposición opcionalmente comprenden uno o más de los siguientes: (1) un componente de recombinación tal como se describe en la presente memoria, (2) instrucciones para la puesta en práctica de los métodos descritos en la presente memoria, y/o para llevar a cabo la síntesis de oligonucleótidos o los procedimientos de selección de genes ensamblados que se proporcionan en la presente memoria, (3) uno o más componentes de ensayo, (4) un recipiente para contener ácidos nucleicos o enzimas, u otros ácidos nucleicos, plantas, animales o células transgénicos, o similares, (5) materiales de empaquetamiento, y (6) un ordenador o un medio legible por ordenador que presenta juegos de instrucciones para alinear ácidos nucleicos diana y para seleccionar oligonucleótidos que, tras la hibridación y elongación, resultan en formas traspuestas de los ácidos nucleicos diana.

35

40

45

50

55

60

65

REIVINDICACIONES

1. Método de recombinación de ácidos nucleicos homólogos, comprendiendo el método:

- 5
- (i) alinear secuencias homólogas de ácidos nucleicos diana utilizando un programa informático para seleccionar las regiones conservadas de identidad de secuencia y las regiones de diversidad de secuencia,
  - 10 (ii) sintetizar un conjunto de oligonucleótidos solapantes de trasposición génica familiar, en donde el conjunto de oligonucleótidos corresponde a por lo menos una región de diversidad de secuencia en los ácidos nucleicos diana, y en donde el conjunto de oligonucleótidos colectivamente corresponde a 50% o más de la longitud completa de cada uno de los ácidos nucleicos diana homólogos,
  - 15 (iii) hibridar el conjunto de oligonucleótidos de trasposición génica familiar,
  - (iv) elongar el conjunto de oligonucleótidos de trasposición génica familiar utilizando una polimerasa, proporcionando de esta manera una población de ácidos nucleicos recombinados,
  - 20 (v) desnaturalizar la población de ácidos nucleicos recombinados, proporcionando de esta manera ácidos nucleicos recombinados desnaturalizados,
  - (vi) hibridar nuevamente los ácidos nucleicos desnaturalizados,
  - 25 (vii) extender los ácidos nucleicos recombinados y nuevamente hibridados que resultan utilizando una polimerasa,
  - (viii) repetir las etapas (v) a (vii) por lo menos una vez, y
  - 30 (iv) seleccionar uno o más de los ácidos nucleicos recombinados y nuevamente hibridados que resultan para un carácter o propiedad deseado, en donde se utilizan proporciones no equimolares de oligonucleótidos de trasposición familiar para sesgar la recombinación.

2. Método según la reivindicación 1, en el que el conjunto de oligonucleótidos de trasposición génica familiar corresponde colectivamente a 60% o más de la longitud completa de los ácidos nucleicos diana homólogos.

35 3. Método según la reivindicación 1, en el que el conjunto de oligonucleótidos de trasposición génica familiar corresponde colectivamente a 70% o más de la longitud completa de cada uno de los ácidos nucleicos diana homólogos.

40 4. Método según la reivindicación 1, en el que el conjunto de oligonucleótidos de trasposición génica familiar corresponde colectivamente a 80% o más de la longitud completa de cada uno de los ácidos nucleicos diana homólogos.

5. Método según la reivindicación 1, que es un método para producir ácidos nucleicos de trasposición familiar sin aislamiento ni clonación de los ácidos nucleicos diana homólogos de longitud completa.

45 6. Método según la reivindicación 1, en el que el conjunto de oligonucleótidos de trasposición génica familiar se sintetiza químicamente según el método de fase sólida de fosforamidita triéster.

7. Método según la reivindicación 1, en el que el conjunto de oligonucleótidos de trasposición familiar comprende una pluralidad de oligonucleótidos de codones modificados.

50

55

60

65

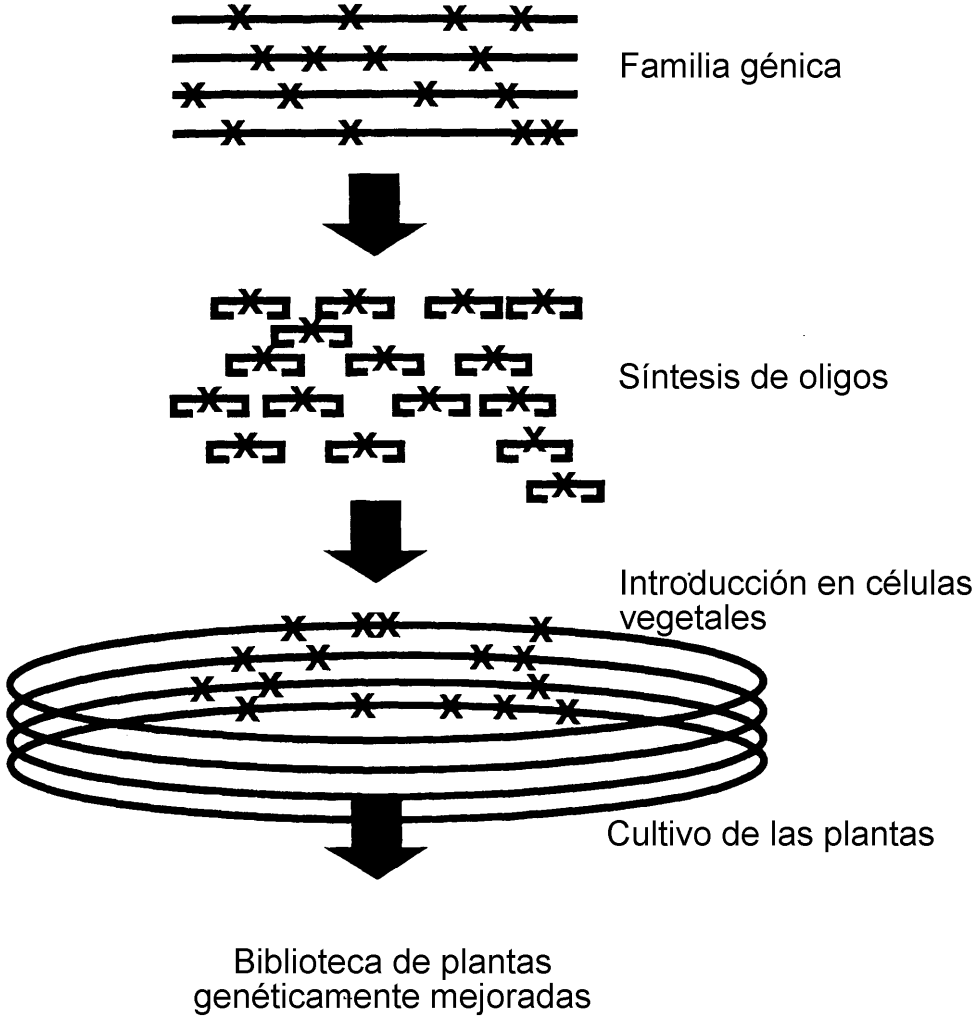


Fig. 1

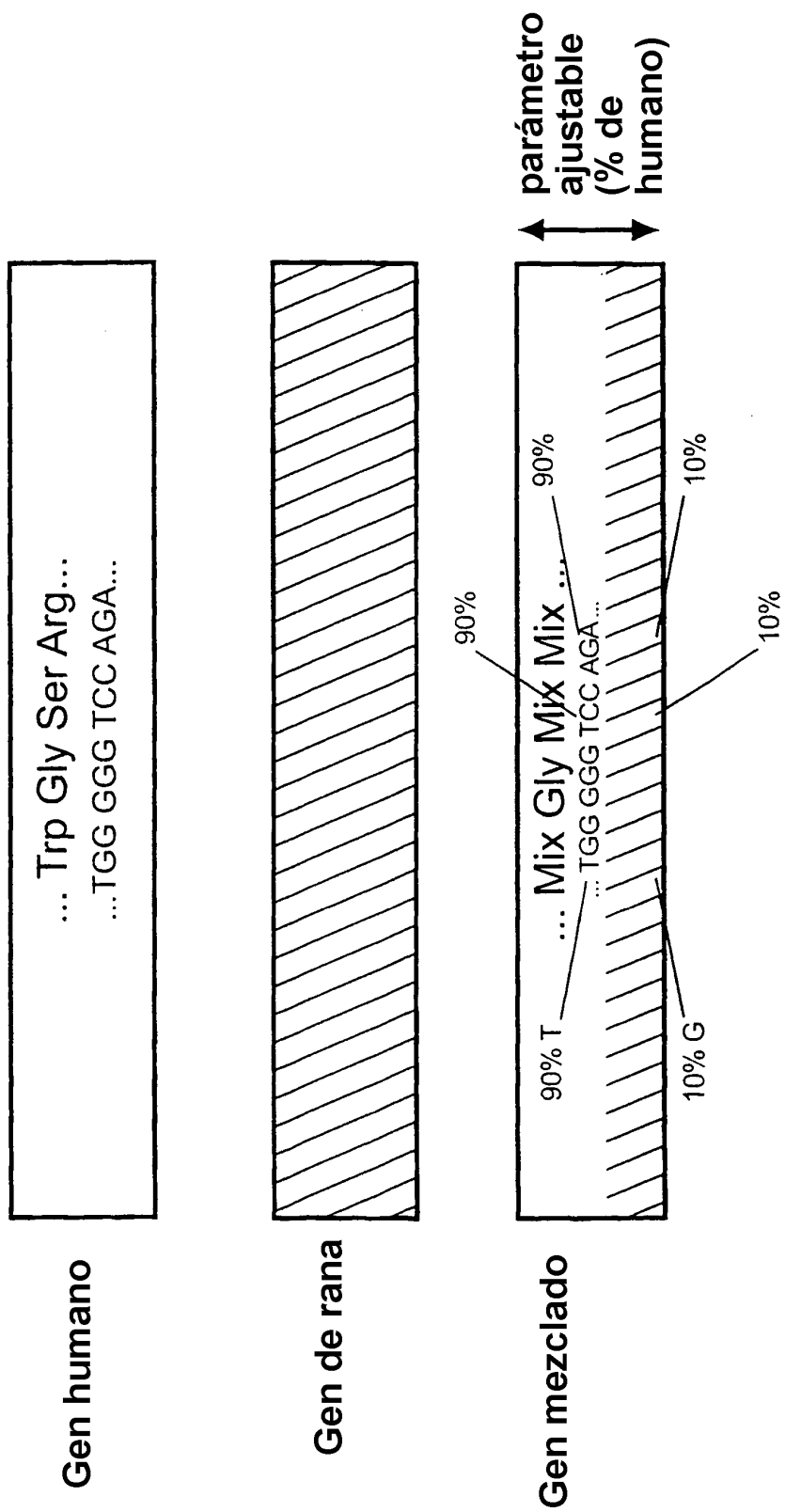


Fig. 2

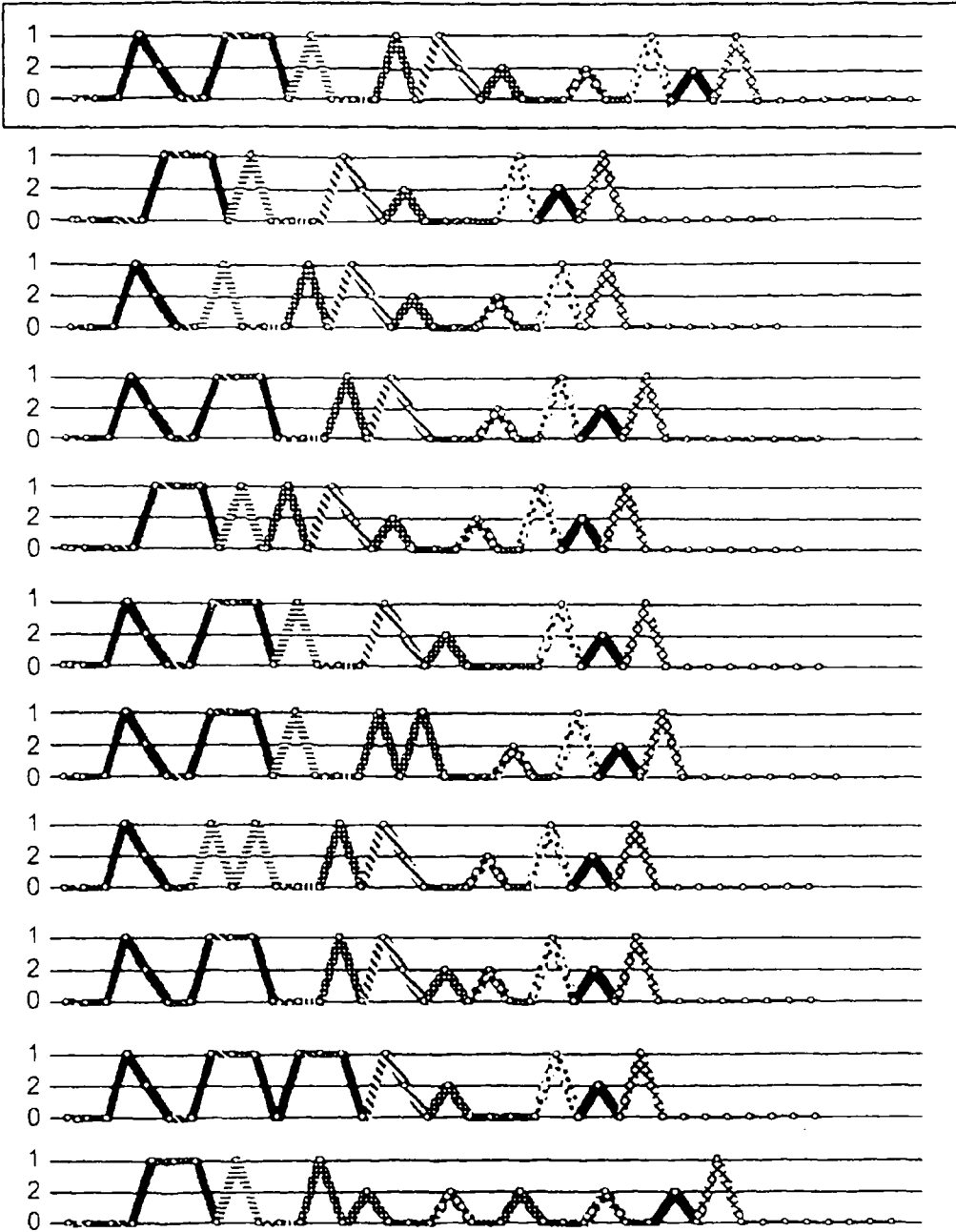


Fig. 3