

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 671 027**

51 Int. Cl.:

C12N 15/10 (2006.01)

C12Q 1/68 (2008.01)

C40B 40/06 (2006.01)

C40B 50/06 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **18.09.2009 E 15199741 (8)**

97 Fecha y número de publicación de la concesión europea: **02.05.2018 EP 3026113**

54 Título: **Síntesis en paralelo automatizada combinada de variantes de polinucleótidos**

30 Prioridad:

11.06.2009 WO PCT/US2009/047046

11.06.2009 US 483089

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

04.06.2018

73 Titular/es:

**CODEXIS, INC. (100.0%)
200 Penobscot Drive
Redwood City, CA 94063, US**

72 Inventor/es:

**COLBECK, JEFFREY;
MIJTS, BENJAMIN;
GIVER, LORRAINE JOAN;
FOX, RICHARD, J.;
MITCHELL, VESNA;
PAK, BUMSHIK, ROBERT y
GILSON, LYNNE**

74 Agente/Representante:

IZQUIERDO BLANCO, María Alicia

ES 2 671 027 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).

Síntesis en paralelo automatizada combinada de variantes de polinucleótidos**Descripción**

5 1. ÀREA TECNOLÒGICA

Esta divulgación se refiere a métodos para una síntesis, clonación, transformación y revisión eficiente de bibliotecas diversas y grandes de variaciones de polinucleótidos que tienen diferencias bien definidas de nucleótidos en relación a un polinucleótido referencial.

10

2. ANTECEDENTES

Varias técnicas de evolución dirigida que se basa in silico e in vitro de funciones proteínicas han permitido la generación de proteínas con propiedades nuevas. Por ejemplo, enzimas de citocromo P450 han sido evolucionadas para tener actividades en contra de sustratos que normalmente no son reconocidos por la enzima que ocurre naturalmente (refiérase a, por ejemplo, Landwehr et al., 2007, Chem Biol 14(3):269-78; Kubo et al., 2006, Chemistry (Química) 12(4):1216-20.). Comúnmente, para generar aquellas nuevas enzimas, un polinucleótido que codifica a un polipéptido referencial, tal como una enzima de tipo silvestre, es sujeto a mutagénesis para generar polinucleótidos que codifican a variaciones de polipéptidos con cambios en las secuencias de aminoácidos. La revisión de variaciones para detectar una propiedad deseada, tal como una mejora en una estabilidad o actividad enzimática en contra de nuevos sustratos, permite la identificación de los residuos de aminoácidos asociados con la propiedad cambiada. Sin embargo, no todas las combinaciones de las mutaciones estarán presentes en una población de variaciones examinadas. Por ejemplo, una mutación asociada con la estabilidad térmica de una enzima podría no encontrarse en asociación con una mutación asociada con un cambio en la especificidad de sustrato. Esta parcialización en la población puede surgir de varios factores, incluyendo, entre otros, la secuencia paternal de aminoácidos codificados por el polinucleótido utilizado para la mutagénesis, la selección posible en contra de la combinación durante la propagación in vivo del polinucleótido, y la parcialización de la técnica utilizada para la mutagénesis (por ejemplo, el uso de polimerasas para introducir errores).

Puesto que las mutaciones en posiciones definidas de residuos de aminoácidos de una secuencia de polipéptidos referenciales pueden suministrar una riqueza de información acerca de las actividades biológicas del polipéptido, una vez que las mutaciones han sido identificadas inicialmente, es deseable preparar varias combinaciones de las mutaciones que no fueron encontradas en el conjunto inicial de variaciones examinadas que pueden ser probadas para detectar a la propiedad deseada. Una selección que se basa in silico de mutaciones definidas o conjuntos de mutaciones suministra un marco referencial para generar un número grande de combinaciones posibles de mutaciones. Por ejemplo, mutaciones que afectan la especificidad de sustratos pueden ser combinadas con mutaciones que afectan otras propiedades enzimáticas, incluyendo, entre otras, la actividad enzimática, la estabilidad térmica y resistencia inhibitoria. Comúnmente, el método para generar a estos polipéptidos que tienen combinaciones nuevas de mutaciones es el sintetizar especies individuales (es decir, la síntesis de cada polinucleótido que codifica al gen mutante). Esto puede ser logrado por medio de síntesis química y/o enzimática de los polinucleótidos en combinación con técnicas estándar de recombinación. Aquellas técnicas de síntesis desde cero requieren una síntesis genética completa de cada variación de polinucleótidos y/o la síntesis de números grandes de cebadores de oligonucleótidos los cuales son utilizados entonces para sintetizar a variantes completas de polinucleótidos (por ejemplo, por medio de ION-PCR). Estas técnicas requieren más síntesis de oligonucleótidos y resultan en producciones inferiores de variaciones que tienen la secuencia correcta. Consecuentemente, si el conjunto de datos de mutaciones es grande, el costo y eficiencia de generar las combinaciones de mutaciones podría limitar la capacidad para examinar un número grande de combinaciones nuevas. Por lo tanto, métodos eficientes y efectivos para generar polinucleótidos que codifican a combinaciones de mutaciones definidas son deseables.

La WO 03/010183 considera métodos en los cuales varios PCRs separados con cebadores mutagénicos son realizados para generar amplicones que entonces son "mezclados, preferiblemente en cantidades equimolares" antes del ensamblaje.

La WO 02/34762 describe un método para producir bibliotecas de moléculas de ácidos nucleicos tales bibliotecas se derivan de una plantilla de ácido nucleico.

La WO 2006/105082 describe hidroxilasas modificadas, células que expresan tales hidroxilasas modificadas, y métodos para producir alcanos hidroxilados poniendo en contacto un sustrato adecuado con tales células.

Ling et al. (Anal Biochem. 15 de diciembre de 1997; 254 (2): 157 - 78) presentan una visión general de enfoques de mutagénesis de ADN.

La WO 01/12802 describe un método para mutagenizar ácidos nucleicos y proteínas con respecto a una

secuencia de ácidos nucleicos inicial mediante la inserción, deleción o sustitución de nucleótido(s) en el ácido nucleico objetivo durante la amplificación.

3. RESUMEN

Este invento proporciona un método para sintetizar una pluralidad de variantes de polinucleótidos que tienen una mezcla estocástica de diferencias definidas de nucleótidos en relación a una secuencia de polinucleótido de referencia, en donde el polinucleótido de referencia codifica un polipéptido de referencia y cada una de la pluralidad de variantes del polinucleótido codifica un polipéptido que tiene por lo menos una diferencia de secuencia de aminoácidos en relación al polipéptido de referencia, el método comprendiendo:

(a) proporcionar una pluralidad de parejas de cebadores directos e inversos, en donde cada una de la pluralidad de cebadores directos e inversos comprende una mezcla de cebadores mutagénicos y no mutagénicos, cada cebador mutagénico comprendiendo una diferencia definida en una posición seleccionada y cada cebador no mutagénico no comprendiendo diferencia en la posición seleccionada, y en donde cada una de la pluralidad de parejas de cebadores directos e inversos comprende una mezcla que tiene una proporción definida de cebadores mutagénicos a no mutagénicos, y en donde cada pareja genera un amplicón que comprende una secuencia capaz de unirse a una secuencia superpuesta adyacente de por lo menos un otro amplicón;

(b) amplificar una plantilla de polinucleótido de referencia con cada una de la pluralidad de parejas de cebadores directos e inversos, generando de este modo una pluralidad de conjuntos de amplicones, en donde los amplicones resultantes comprenden una mezcla de secuencias con o sin las diferencias definidas en las posiciones seleccionadas, y en donde cada conjunto comprende amplicones que tienen secuencias superpuestas adyacentes capaces de unirse para formar la longitud completa de la secuencia de polinucleótidos de referencia;

(c) ensamblar y replicar la pluralidad de conjuntos de los amplicones, sintetizando de este modo una pluralidad de variantes de polinucleótido que tienen una mezcla estocástica de diferencias de nucleótidos definidas.

Esta divulgación se refiere a métodos para generar eficientemente polinucleótidos que tienen combinaciones diferentes de cambios secuenciales definidos (por ejemplo, mutaciones deseadas de aminoácidos) en comparación a una secuencia referencial de polinucleótidos. Los métodos se basan en el uso de una biblioteca de fragmentos de polinucleótidos (es decir, amplicones) que tienen regiones adyacentes en superposición de tal forma que conjuntos de los fragmentos de polinucleótidos pueden ser ensamblados para generar una pluralidad de variantes de polinucleótidos donde cada uno tiene un conjunto definido de cambios secuenciales. Cebadores seleccionados directos e inversos son utilizados para introducir los cambios secuenciales al amplificar a una plantilla referencial de polinucleótidos y generar, por lo tanto, fragmentos de polinucleótidos que comprenden a los cambios secuenciales definidos. La biblioteca está diseñada para tener suficientes fragmentos de polinucleótidos para ensamblar a por lo menos dos secuencias diferentes de variantes de polinucleótidos. En algunas realizaciones, la biblioteca de los fragmentos de polinucleótidos contienen a miembros que tienen a todas las diferencias definidas en una secuencia de polinucleótidos (por ejemplo, los cambios deseados de nucleótidos) en comparación a una secuencia referencial de tal forma que todas las permutaciones secuenciales pueden ser ensambladas. En algunas realizaciones, los polinucleótidos pueden ser diseñados para codificar a polipéptidos que tienen diferencias definidas en la secuencia de aminoácidos en comparación a una secuencia referencial de aminoácidos.

Los métodos de esta divulgación son capaces de producir bibliotecas grandes de secuencias variantes de polinucleótidos que tienen diferencias definidas de nucleótidos (por ejemplo, bibliotecas de 10, 50, 100, 150, 300, 500, 700, 1000 o más variantes, donde cada una tiene 1, 2, 3, 5, 9, 12, 15, 20, 25, 30, 35, 40, 50, o más cambios deseados), con relativamente pocos oligonucleótidos (por ejemplo, en comparación a métodos de síntesis de genes completos), y oligonucleótidos relativamente cortos (por ejemplo, 35-mer o menos), y donde el porcentaje promedio de secuencias correctas es sorpresivamente alto (por ejemplo, por lo menos el 65%, 75%, 85%, 95% o más).

En algunas realizaciones, el método para formar variantes de polipéptidos que codifican a polinucleótidos puede comprender: seleccionar una pluralidad de diferencias definidas de residuos de aminoácidos en relación a una secuencia referencial de aminoácidos; definir segmentos en superposición de una secuencia de polinucleótidos que codifican al polipéptido con la secuencia diferente de aminoácidos o la secuencia referencial de aminoácidos, donde cada segmento está enlazado por un conjunto de secuencias enlazadoras de cebadores directos e inversos y donde una diferencia secuencial de polinucleótidos que codifica a cada una de las pluralidades de las diferencias de residuos de aminoácidos es cubierta en la secuencia enlazadora de cebadores; amplificar cada segmento con un conjunto de cebadores directos e inversos, donde los cebadores seleccionados directos y/o inversos contienen a las diferencias secuenciales de polinucleótidos, para generar una biblioteca de amplicones que comprende a miembros que codifican a las diferencias de residuos de aminoácidos, donde la biblioteca comprende a miembros que ensamblan a 2 o más permutaciones diferentes de secuencias de aminoácidos de las diferencias definidas de

aminoácidos; ensamblar a partir de la biblioteca a un conjunto de amplicones que tengan regiones complementarias adyacentes en superposición, donde el conjunto de amplicones codifican juntos al polipéptido con una permutación definida de secuencias de aminoácidos que tiene una o más diferencias residuales de aminoácidos; y replicar el conjunto de amplicones ensamblados para sintetizar al polinucleótido que codifica al polipéptido.

5 Además se describen este documento métodos para generar la biblioteca de fragmentos de polinucleótidos, donde el método comprende: (a) generar una pluralidad de permutaciones de secuencias de aminoácidos que difieren de una secuencia referencial de aminoácidos que se basa en una pluralidad de diferencias definidas de residuos de aminoácidos en comparación con la secuencia referencial de aminoácidos, y para cada permutación; (i) 10 determinar una secuencia de polinucleótidos que codifica a la permutación secuencial de aminoácidos que se basa en una secuencia referencial de polinucleótidos; (ii) identificar un cambio en la secuencia de polinucleótidos que codifica a una diferencia residual de aminoácidos en comparación a una secuencia referencial de aminoácidos, y determinar la proximidad de un cambio del vecino más cercano en la secuencia de polinucleótidos que codifica a otra diferencia residual de aminoácidos en la permutación secuencial de aminoácidos; (iii) seleccionar un cebador de 15 oligonucleótidos directos que tenga una secuencia que codifica a las diferencias residuales de aminoácidos, y opcionalmente incluir al cambio del vecino más cercano en la secuencia de polinucleótidos en el mismo cebador de oligonucleótidos directos si está cerca del primer cambio en la secuencia de polinucleótidos; (iv) identificar el próximo cambio en secuencias de polinucleótidos o hasta que el final del polinucleótido sea alcanzado, y seleccionar un cebador de oligonucleótidos inverso para amplificar a un fragmento de polinucleótidos con los cebadores de 20 oligonucleótidos hacia delante, donde el cebador inverso codifica opcionalmente al siguiente cambio en la diferencia residual de aminoácidos; (v) reiterar los pasos (ii) al (iv) para cada cambio en secuencias de polinucleótidos que codifican a una diferencia residual de aminoácidos de tal forma que todos los cambios en las secuencias de polinucleótidos están presentes en los cebadores de oligonucleótidos; y (b) amplificar con cada conjunto de cebadores de oligonucleótidos directos e inversos para generar la biblioteca de amplicones en superposición que 25 tengan miembros que codifican a diferencias residuales de aminoácidos.

En otro aspecto, esta divulgación suministra una biblioteca de aquellos fragmentos de polinucleótidos (es decir, amplicones) que pueden ser utilizados para ensamblar a las variantes de polinucleótidos. En algunas realizaciones, la pluralidad de fragmentos de nucleótidos comprende a fragmentos de polinucleótidos con regiones adyacentes superpuestas, donde cada fragmento de polinucleótidos está enlazado por secuencias enlazadoras de 30 cebadores directos e inversos, donde la pluralidad de fragmentos de polinucleótidos tiene miembros que codifican en las secuencias enlazadoras de cebadores a una diferencia específica de residuos de aminoácidos a partir de una pluralidad definida de diferencias residuales de aminoácidos en relación a una secuencia referencial de aminoácidos de tal forma que la pluralidad de fragmentos de polinucleótidos codifica a toda una pluralidad seleccionada de 35 diferencias residuales de aminoácidos a partir de la pluralidad definida de diferencias residuales de aminoácidos; y donde la pluralidad de fragmentos de polinucleótidos comprende a miembros para ensamblar 2 o más permutaciones diferentes de secuencias de aminoácidos de la pluralidad seleccionada de diferencias residuales de aminoácidos.

40 En otro aspecto, la siguiente divulgación suministra un método para sintetizar una pluralidad de variantes de polinucleótidos donde cada una tiene una diferencia definida de nucleótidos en relación a una secuencia referencial de polinucleótidos, donde el método comprende: (a) amplificar por separado a una plantilla referencial de polinucleótidos con cada una de una pluralidad de parejas de cebadores directos e inversos, donde la pluralidad de 45 parejas de cebadores directos e inversos comprende a la pluralidad de diferencias definidas de nucleótidos y donde cada pareja genera un amplicón que comprende a una secuencia capaz de enlazarse a una secuencia adyacente superpuesta de por lo menos un amplicón adicional; (b) ensamblar por separado a una pluralidad de conjuntos de amplicones, donde cada conjunto comprende a amplicones que tienen secuencias adyacentes superpuestas capaces de enlazarse para formar la longitud completa de la secuencia referencial de polinucleótidos; y (c) replicar la pluralidad de conjuntos de amplicones ensamblados, sintetizando, por lo tanto, a una pluralidad de variantes de 50 polinucleótidos.

En otra realización, la divulgación suministra un método para sintetizar a una pluralidad de variantes de polinucleótidos donde cada una tiene una diferencia definida de nucleótidos en relación a una secuencia referencial de polinucleótidos, donde el método comprende: (a) seleccionar una pluralidad de diferencias definidas de 55 nucleótidos en relación a la secuencia referencial de polinucleótidos; (b) definir una pluralidad de segmentos de la secuencia referencial de polinucleótidos, donde cada segmento está superpuesto con por lo menos un segmento adyacente y es enlazado por una pareja de secuencias enlazadoras de cebadores directos e inversos, donde los cebadores directos y/o inversos comprenden a por lo menos una de las pluralidades de las diferencias definidas de nucleótidos; (c) amplificar por separado a una plantilla referencial de polinucleótidos con cada pluralidad de parejas 60 de cebadores directos e inversos, donde cada pareja de cebadores comprende a por lo menos una de las pluralidades de diferencias definidas de nucleótidos, generando, por lo tanto, a una biblioteca de amplicones a la que se le pueden asignar direcciones referenciales donde cada uno corresponde a un segmento de la secuencia referencial de polinucleótidos que tiene diferencias definidas de nucleótidos; (d) ensamblar por separado a una pluralidad de conjuntos de amplicones a partir de la biblioteca de amplicones a la que se le puede asignar 65 direcciones referenciales, donde cada conjunto comprende a amplicones que corresponden a segmentos

adyacentes superpuestos que comprenden la longitud completa de la secuencia referencial de polinucleótidos; y (e) replicar la pluralidad de conjuntos de amplicones ensamblados, sintetizando, por lo tanto, a una pluralidad de variantes de polinucleótidos.

5 En algunas realizaciones de los métodos para sintetizar una pluralidad de variantes de polinucleótidos, después de que son ensamblados y replicados por separado (por ejemplo, por medio de SOE-PCR), las variantes de polinucleótidos son combinadas (es decir, son agrupadas) en vez de ser mantenidas por separado (por ejemplo, en un ensayo en el cual se pueden asignar direcciones referenciales). La agrupación es clonada entonces en un vector de expresión, transformada en células, y puesta en placas, lo cual suministra una ventaja sorpresiva al
10 facilitar la revisión de bibliotecas grandes comprendidas de cientos o miles de variantes de longitudes completas que tienen diferencias definidas de nucleótidos, seguido por una secuenciación limitada a solamente aquellos polipéptidos que codifican a variantes que tienen algún nivel deseado de actividad u otra propiedad mejorada. Por lo tanto, en algunas realizaciones, el método de sintetizar una pluralidad de variantes de polinucleótidos, donde cada una tiene a por lo menos una diferencia definida de nucleótidos en relación a una secuencia referencial de polinucleótidos, comprende: (a) amplificar por separado a una plantilla referencial de polinucleótidos con cada una de una pluralidad de parejas de cebadores directos e inversos, donde la pluralidad de parejas de cebadores directos e inversos comprende la pluralidad de diferencias definidas de nucleótidos y donde cada pareja genera un amplicón que comprende a una secuencia capaz de enlazarse a una secuencia adyacente superpuesta de por lo menos un amplicón adicional; (b) ensamblar por separado a una pluralidad de conjuntos de amplicones, donde cada conjunto comprende a amplicones que tienen secuencias adyacentes superpuestas capaces de enlazarse para formar a la longitud completa de la secuencia referencial de polinucleótidos; (c) replicar la pluralidad de conjuntos de amplicones ensamblados; y (d) combinar la pluralidad de variantes de polinucleótidos, sintetizando, por lo tanto, a una agrupación que comprenden a una pluralidad combinada de variantes de polinucleótidos, donde cada una tiene a una diferencia definida de nucleótidos. En algunas realizaciones del método el polinucleótido de referencia codifica a un polipéptido referencial y cada una de las pluralidades de variantes de polinucleótidos codifican a un polipéptido que tiene a por lo menos una diferencia secuencial de aminoácidos. En otras realizaciones, el método comprende además después del paso de combinar la pluralidad de variantes de polinucleótidos: (e) un paso de clonación de la pluralidad combinada de variantes de polinucleótidos en un vector de expresión, generando, por lo tanto, a una pluralidad combinada de vectores de expresión donde cada uno comprende a una variante de polinucleótidos; (f) transformar las células con la pluralidad combinada de vectores de expresión; (g) examinar las células transformadas para detectar la actividad de los polipéptidos codificados por las variantes de polinucleótidos; (h) secuenciar a variantes de polinucleótidos que codifican a polipéptidos que tienen la actividad; o (i) aislar a por lo menos un polipéptido codificado por las variantes de polinucleótidos.

35 En algunas realizaciones, la divulgación suministra métodos de revisión de una biblioteca, de variantes de polipéptidos que codifican a variantes de polinucleótidos, a la que se le puede asignar direcciones referenciales, donde el método comprende: (a) combinar a miembros de la biblioteca de variantes de polinucleótidos a la que se le pueden asignar direcciones referenciales en un grupo; (b) clonar al grupo de variantes de polinucleótidos en un vector de expresión; (c) transformar a células con el vector de expresión; (d) colocar en placas a las células transformadas para generar una pluralidad de clones separados que comprenden a la biblioteca de las variantes de polinucleótidos; y (e) examinar a los clones para detectar una propiedad mejorada en relación a un polipéptido referencial. En algunas realizaciones del método de revisión, el método es ejecutado donde la biblioteca a la que se le puede asignar direcciones referenciales de variantes de polinucleótidos comprende a por lo menos 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, o más variantes diferentes de polinucleótidos. En algunas realizaciones del método de revisión, el método es ejecutado en casos en los cuales una biblioteca a la que se le puede asignar direcciones referenciales de variantes de polinucleótidos comprende a por lo menos 100, 200, 400, 800, 1000, o más variantes diferentes de polinucleótidos donde cada una comprende a una diferencia definida diferente de nucleótidos en una de por lo menos 5, 10, 20, 40, 50, o más posiciones seleccionadas diferentes, respectivamente. En algunas realizaciones del método de revisión, el método es ejecutado donde la biblioteca de variantes de polinucleótidos a la que se le puede asignar direcciones referenciales comprende a por lo menos 460 variantes diferentes de polinucleótidos donde cada una comprende a uno de 23 codones diferentes en una de 20 posiciones seleccionadas diferentes. En algunas realizaciones de los métodos mencionados anteriormente de revisión, por lo menos el 75%, por lo menos el 80%, por lo menos el 85%, por lo menos el 90%, por lo menos el 95%, o más de la biblioteca de variantes de polinucleótidos a la que se le puede asignar direcciones referenciales comprende a la secuencia correcta.

60 En otras realizaciones, esta divulgación suministra un método para sintetizar una pluralidad de variantes de polinucleótidos que tienen una mezcla estocástica de diferencias definidas de nucleótidos en relación a una secuencia referencial de polinucleótidos. Aquella sección suministra una capacidad fácil para crear bibliotecas combinatorias de variantes de polinucleótidos que tienen combinaciones estocásticas de diferencias definidas de nucleótidos. El método comprende: (a) suministrar una pluralidad de parejas de cebadores directos e inversos, donde la pluralidad de parejas de cebadores directos e inversos comprende a una mezcla de cebadores mutagénicos y no mutagénicos, donde los cebadores mutagénicos comprenden a una pluralidad de diferencias definidas de nucleótidos, y donde cada pareja genera a un amplicón que comprende a una secuencia capaz de enlazarse a una secuencia adyacente superpuesta de por lo menos un amplicón adicional; (b) amplificar una plantilla
65

referencial de polinucleótidos con cada una de las pluralidades de parejas de cebadores directos e inversos, generando, por lo tanto, a una pluralidad de conjuntos de amplicones, donde cada conjunto comprende a amplicones que tienen secuencias adyacentes superpuestas capaces de enlazarse para formar a la longitud completa de la secuencia referencial de polinucleótidos; y (c) ensamblar y replicar la pluralidad de conjuntos de los amplicones, sintetizando, por lo tanto, a una pluralidad de variantes de polinucleótidos que tienen una mezcla estocástica de diferencias definidas de nucleótidos. En otra realización, este método puede ser ejecutado donde la pluralidad de parejas de cebadores directos e inversos es combinada antes de amplificar la plantilla referencial de polinucleótidos, generando, por lo tanto, a una pluralidad combinada de conjuntos de amplicones. En otra realización, este método puede ser ejecutado donde la amplificación de la plantilla referencial de polinucleótidos es ejecutada por separado con cada una de las pluralidades de parejas de cebadores directos e inversos, y la pluralidad de conjuntos de amplicones es combinada antes del ensamblaje y la replicación.

En algunas realizaciones de los métodos aquí descritos para sintetizar a una pluralidad de variantes de polinucleótidos, los métodos pueden ser ejecutados de tal forma que el polinucleótido referencial codifica a un dipéptido referencial y cada una de las pluralidades de variantes de polinucleótidos codifica a un polipéptido que tiene a por lo menos una diferencia secuencial de aminoácidos. En realizaciones adicionales, los métodos pueden ser ejecutados donde comprenden adicionalmente a los pasos de: (i) clonar a cada una de las pluralidades de variantes de polinucleótidos en un vector de expresión; (ii) transformar las células con los vectores de expresión; (iii) examinar a las células transformadas para detectar la actividad de los polipéptidos codificados por las variantes de polinucleótidos; o (iv) aislar a por lo menos un polipéptido codificado por las variantes de polinucleótidos. Adicionalmente, los métodos pueden ser ejecutados de tal forma que cada variante de polinucleótidos es ensamblada en una posición conocida en una formación.

En otras realizaciones de los métodos para sintetizar una pluralidad de variantes de polinucleótidos, los métodos pueden ser ejecutados de tal forma que una pluralidad de cambios diferentes de aminoácidos es codificada en una o más de las posiciones objetivo en un polipéptido. En una de esas realizaciones, la pluralidad de parejas de cebadores directos e inversos comprenden a cebadores degenerados (por ejemplo, un conjunto de cebadores que tienen a un codón degenerado en la misma posición en cada una de las secuencias cebadoras) que codifican a una pluralidad de diferencias de aminoácidos en una sola posición en el polipéptido. Por ejemplo, en algunas realizaciones los cebadores degenerados pueden comprender un conjunto de cebadores que tienen codones degenerados seleccionados de un grupo y consiste de NHT, NNB, NNG, NNK, NNN, NNS, NNT, NDT, RMG, RNG, RRS, SNT, VNS, VNT, y VWG. En algunas realizaciones, los métodos pueden ser ejecutados de tal forma que los diferentes codones consisten de los codones degenerados NNT y VWG, y el codón TGG. Estos codones codifican a todos los 20 aminoácidos naturales con solamente 23 codones. En algunas realizaciones, los diferentes codones degenerados pueden ser utilizados para diferencias de aminoácidos en posiciones diferentes del polinucleótido. Por ejemplo, en una posición, los cebadores degenerados que codifican a todos los 20 aminoácidos pueden ser utilizados, mientras que, en otra posición, pueden utilizarse cebadores degenerados que codifican a solamente 16 o menos aminoácidos.

En algunas realizaciones de los métodos aquí descritos para sintetizar una pluralidad de variantes de polinucleótidos, los métodos pueden ser ejecutados de tal forma que la pluralidad de secuencias de cebadores directos e inversos son generadas por medio de los pasos de: (i) identificar una primera diferencia definida en la secuencia variante de polinucleótidos en comparación a la secuencia referencial, y determinar la proximidad de la diferencia definida del vecino más cercano en la secuencia de polinucleótidos; (ii) seleccionar un cebador directo que tenga una secuencia que comprenda a la primera diferencia definida de nucleótidos, y opcionalmente que incluya a la primera diferencia definida de nucleótidos; (iii) identificar una siguiente diferencia definida en la secuencia variante de polinucleótidos en comparación a la secuencia referencial, y determinar la proximidad de una diferencia definida del vecino más cercano en la secuencia de polinucleótidos, o identificar que el final de la variante de polinucleótidos ha sido alcanzada; (iv) seleccionar a un cebador inverso que tenga una secuencia que comprenda a la siguiente diferencia definida de nucleótidos, y opcionalmente incluir a cualquier diferencia definida del vecino más cercano en el mismo cebador hacia delante si estuviere próximo a la siguiente diferencia definida de nucleótidos; y (v) repetir los pasos desde el (iii) hasta el (iv) para cada diferencia definida en la secuencia variante de polinucleótidos de tal forma que todas las diferencias definidas estén presentes en los cebadores.

En algunas realizaciones de los métodos aquí descritos para sintetizar una pluralidad de variantes de polinucleótidos, la pluralidad de variantes de polinucleótidos comprende a por lo menos 10, 25, 35, 50, 75, 90, 120, 150, 180, 300, 500, 700, 900, o aún más variantes diferentes de polinucleótidos.

En algunas realizaciones de los métodos aquí descritos para sintetizar una pluralidad de variantes de polinucleótidos, por lo menos una de las pluralidades de las variantes de polinucleótidos comprende a por lo menos 2, 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, o aún más diferencias definidas de nucleótidos en relación a la secuencia referencial de polinucleótidos. En algunas realizaciones, 2 o más, o en algunas realizaciones cada una de las pluralidades de variantes de polinucleótidos comprenden a por lo menos 1, 2, 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, o aún más diferencias definidas de nucleótidos en relación a la secuencia referencial de polinucleótidos.

En algunas realizaciones de los métodos, la pluralidad de variantes de nucleótidos sintetizada utilizando los métodos puede comprender a por lo menos 10, 20, 30, 40, 50, 100, 200, o aún más variantes diferentes de polinucleótidos, donde cada variante comprende a una diferencia diferente definida de nucleótidos en una de 10, 20, 30, 40, 50, 100, 200, o más posiciones seleccionadas diferentes (es decir, 10, 20, 30, 40, 50, 100, 200, o más mutaciones distintas en un solo lugar). También es accesible una diversidad a mucho mayor escala, por ejemplo, cuando se usan cebadores degenerados, el método puede suministrar bibliotecas de "mutagénesis de saturación" donde cada miembro variante de polinucleótidos tiene una secuencia diferente que comprende a uno de 23 codones diferentes a una posición seleccionada diferente de la secuencia referencial de polinucleótidos. Asimismo, los métodos aquí presentados pueden ser utilizados para sintetizar una pluralidad de variantes de polinucleótidos que comprende a por lo menos 23, 46, 69, 92, 115, 230, 460, 920 o más variantes diferentes de polinucleótidos donde cada una comprende a uno de 23 codones diferentes en una de 1, 2, 3, 4, 5, 10, 20, 40, o más posiciones seleccionadas diferentes. En algunas realizaciones, los métodos pueden ser ejecutados de tal forma que diferentes conjuntos de codones específicos y/o degenerados son utilizados en diferentes posiciones seleccionadas para diferencias definidas de nucleótidos en la pluralidad de cebadores directos e inversos, suministrando, por lo tanto, bibliotecas diversas de variantes de polinucleótidos que tienen mutaciones indefinidas.

En algunas realizaciones de los métodos aquí descritos para sintetizar una pluralidad de variantes de polinucleótidos, por lo menos una de las pluralidades de conjuntos de amplicones comprende a por lo menos 3, a por lo menos 5, a por lo menos 7, a por lo menos 10, o más amplicones diferentes. En algunas realizaciones, 2 o más, o en algunas realizaciones cada una de las pluralidades de los conjuntos de amplicones comprende a por lo menos 3, a por lo menos 5, a por lo menos 7, a por lo menos 10, o más amplicones diferentes. En realizaciones en las cuales se usan a cebadores degenerados, los amplicones diferentes son degenerados. Asimismo, en algunas realizaciones por lo menos una de las pluralidades de conjuntos de amplicones comprende a por lo menos 3, a por lo menos 5, a por lo menos 7, a por lo menos 10 o más amplicones degenerados diferentes.

En algunas realizaciones de los métodos aquí descritos para sintetizar una pluralidad de variantes de polinucleótidos, la longitud de la secuencia referencial de polinucleótidos es de por lo menos 500 bp, 750 bp, 1000 bp, 1250 bp, 1500 bp, o incluso más larga.

En algunas realizaciones de los métodos aquí descritos para sintetizar a una pluralidad de variantes de polinucleótidos, la pluralidad de parejas de cebadores directos e inversos comprende a 400 o menos, 300 o menos, 200 o menos, 100 o menos, 50 o menos, o incluso 25 o menos. En algunas realizaciones, la pluralidad de parejas de cebadores directos e inversos comprende a desde 6 a alrededor de 200, desde 6 a alrededor de 150, desde 6 a alrededor de 100, desde 6 a alrededor de 50, desde 6 a alrededor de 40, desde 6 a alrededor de 30, desde 6 a alrededor de 25, desde 6 a alrededor de 20, desde 6 a alrededor de 15, o incluso menos oligonucleótidos diferentes, y donde las longitudes de los oligonucleótidos son desde alrededor de 20 a alrededor de 50 nucleótidos, desde alrededor de 20 a alrededor de 40 nucleótidos, o desde alrededor de 25 hasta alrededor de 35 nucleótidos.

En algunas realizaciones de los métodos aquí descritos para sintetizar una pluralidad de variantes de polinucleótidos, la pluralidad de parejas de cebadores directos e inversos comprende a 400 o menos, a 300 o menos, 200 o menos, a 100 o menos, a 50 o menos, o incluso a 25 o menos. En algunas realizaciones, la pluralidad de parejas de cebadores directos e inversos comprende desde 6 a alrededor de 200, desde 6 a alrededor de 150, desde 6 a alrededor de 100, desde 6 a alrededor de 50, desde 6 a alrededor de 40, desde 6 a alrededor de 30, desde 6 a alrededor de 25, desde 6 a alrededor de 20, desde 6 a alrededor de 15, o incluso menos oligonucleótidos diferentes, y donde la longitud de los oligonucleótidos es desde alrededor de 20 a alrededor de 50 nucleótidos, desde alrededor de 20 a alrededor de 40 nucleótidos, o desde alrededor de 25 a alrededor de 35 nucleótidos.

En algunas realizaciones de los métodos aquí descritos para sintetizar una pluralidad de variantes de polinucleótidos, el porcentaje promedio de la pluralidad sintetizada de variantes de polinucleótidos comprende a la secuencia correcta que es por lo menos alrededor del 65%, 70%, 75%, 80%, 85%, 90% o 95% o más. Asimismo, en cualquiera de las realizaciones el método aquí descrito es capaz de sintetizar a bibliotecas de variantes de polinucleótidos que tienen altos niveles de diversidad (diferencias distintas de nucleótidos en posiciones seleccionadas diferentes) con una precisión bien definida de por lo menos alrededor del 65%, 75%, 85% o 95%, o más secuencias.

En algunas realizaciones, la divulgación suministra métodos para sintetizar una pluralidad de variantes de polinucleótidos donde cualquiera de los parámetros descritos anteriormente (por ejemplo, el número de variantes, el número de diferencias definidas de nucleótidos, la longitud de la secuencia referencial de polinucleótidos, el número de parejas de cebadores directos e inversos, la longitud de los oligonucleótidos de cebadores, y/o los porcentajes de las secuencias perfectas de longitudes completas) son combinados.

Adicionalmente a los métodos ya mencionados, la divulgación también suministra una biblioteca de variantes de polinucleótidos a la que se le pueden asignar direcciones referenciales que comprenden a una pluralidad de variantes de polinucleótidos o amplicones sintetizados de acuerdo a cualquiera de los métodos ya descritos. Asimismo, en algunas realizaciones la divulgación suministra una biblioteca de amplicones a la cual se

pueden asignar direcciones referenciales, donde cada miembro de la biblioteca de amplicones comprende a por lo menos una diferencia definida de nucleótidos en relación a una secuencia referencial de polinucleótidos y una región adyacente superpuesta capaz de enlazarse a la región adyacente superpuesta de por lo menos un amplicón adicional en la biblioteca, y donde la pluralidad de amplicones comprende a por lo menos un conjunto de amplicones capaces de enlazarse para formar a la longitud completa de la secuencia referencial de polinucleótidos. En algunas realizaciones, la dirección de amplicones a la cual se le pueden asignar direcciones referenciales comprende a miembros para ensamblar 2 o más variantes diferentes de polinucleótidos que comprende a diferencias definidas de nucleótidos en relación a la secuencia referencial de polinucleótidos. En algunas realizaciones de la biblioteca de amplicones a la cual se le pueden asignar direcciones referenciales, la secuencia referencial de polinucleótidos que codifican a un polipéptido referencial y la pluralidad de amplicones comprenden a miembros suficientes para ensamblar a todas las diferencias posibles de nucleótidos que codifican a una pluralidad seleccionada de diferencias residuales de aminoácidos.

En otras realizaciones, la divulgación suministra composiciones que comprenden a una pluralidad de variantes de polinucleótidos que tienen diferencias definidas de nucleótidos, donde las variantes de polinucleótidos son sintetizadas de acuerdo a los métodos aquí descritos. Asimismo, en algunas realizaciones, la divulgación suministra una composición que comprende a una pluralidad de vectores de expresión que comprenden a variantes de polinucleótidos sintetizadas o una pluralidad de células transformadas que comprenden a variantes de polinucleótidos sintetizadas de acuerdo a los métodos ya descritos.

En otra realización la divulgación suministra una biblioteca de variantes de polipéptidos aislada de acuerdo a los métodos de síntesis de polipéptidos que codifican a variantes de polinucleótidos, seguido por la clonación de variantes a vectores de expresión, a células transformantes y a la expresión de polipéptidos en los transformantes.

En este documento también se suministran métodos implementados por computadora para ejecutar varios pasos de los métodos aquí descritos.

4. DESCRIPCIÓN BREVE DE LAS FIGURAS

La figura 1 muestra una técnica estándar para generar polinucleótidos que codifican a variantes definidas de polipéptidos (a la izquierda) en comparación al método aquí descrito para utilizar bibliotecas de fragmentos superpuestos de polinucleótidos (a la derecha).

La figura 2 suministra un esquema de flujo de trabajo de muestra para generar bibliotecas de fragmentos superpuestos de polinucleótidos basado en la generación de cebadores de oligonucleótidos para segmentos superpuestos de un polinucleótido y el uso de oligonucleótidos en reacciones PCR para generar bibliotecas de fragmentos superpuestos de polinucleótidos.

La figura 3 muestra a los geles de agarosa de 96 muestras (y 8 controles) resultantes de ensamblar a los fragmentos superpuestos de polinucleótidos y replicar los fragmentos ensamblados de polinucleótidos para sintetizar a las variantes de polinucleótidos que codifican a la variante deseada de polipéptidos. Casi todos los geles mostraron una sola banda fuerte indicando que la secuencia de longitud esperada estaba presente.

La figura 4 muestra un flujograma para generar a variantes que codifican a polinucleótidos mediante el uso de una biblioteca de fragmentos superpuestos de polinucleótidos.

La figura 5 muestra un flujograma para generar una biblioteca de cebadores de oligonucleótidos para generar una biblioteca de amplicones superpuestos de polinucleótidos para cada permutación secuencial de aminoácidos.

La figura 6 muestra un flujograma de instrucciones para la creación y selección automatizadas de cebadores de oligonucleótidos y fragmentos superpuestos de oligonucleótidos.

5. DESCRIPCIÓN DETALLADA

Tal como se utilizan esta especificación en las reivindicaciones adjuntas, las formas singulares “un”, “uno” y “el” incluyen a referencias plurales a menos que el contexto lo indique claramente de otra forma. Por lo tanto, por ejemplo, una referencia a “una proteína” incluye a más de una proteína, y referencias a “un compuesto” se refiere a más de un compuesto.

Además, el uso de “o” significa “y/o” a menos que se indique de otra forma. Asimismo, “comprende”, “comprendiendo”, “comprender”, “incluir”, “que incluye”, e “incluye” son intercambiables y no tiene la intención de ser limitantes.

Además se entiende que en los casos en los que las descripciones de varias realizaciones utilizan el

término “comprende”, aquellas personas con conocimiento en la industria entenderán que, en algunas instancias específicas, una realización puede ser descrita alternamente utilizando un lenguaje que “consiste esencialmente de” o “consistiendo de”.

5 Los títulos de las realizaciones son utilizados aquí para propósitos de organización y no se considera que sean limitantes del asunto descrito. Tal como se utiliza aquí, los siguientes términos tienen la intención de tener los siguientes significados.

10 5.1 Definiciones

Los términos “amplificar” y “amplificación” tal como se utilizan en este documento incorporan su uso común y se refieren al uso de cualquier metodología adecuada de amplificación para generar o detectar a cualquier polinucleótido, recombinante o expresado naturalmente, que es fácil de ser amplificado in vivo o in vitro, tal como una reacción en cadena de polimerasa (PCR).

15 El término “Amplicón” se refiere al producto de la reacción de amplificación generada a través de la extensión de cualquiera de, o de ambos de, una pareja de cebadores de amplificación. Un amplicón podría contener a ácidos nucleicos exponencialmente amplificados si ambos cebadores son utilizados para hibridarse a una secuencia objetiva. Alternamente, los amplicones podrían ser generados por medio de una amplificación lineal si uno de los cebadores utilizados no se hibrida a la secuencia objetivo. Por lo tanto, este término es utilizado genéricamente en este documento y no implica la presencia de ácidos nucleicos amplificados exponencialmente.

20 El término “recocerse” o “hibridación” se refiere a las interacciones de emparejamiento de las bases en un polímero de bases nitrogenadas con otra que resulta en la formación de una estructura de doble hebra, una estructura triple o una estructura cuaternaria. Un recosido o una hibridación puede ocurrir por medio de interacciones de emparejamientos de bases de Watson-Crick, pero podrían ser reguladas por otras interacciones de enlaces de hidrógeno, tales como el emparejamiento de bases de Hoogsteen.

30 El término “Ensamblar” se refiere a juntar una pluralidad de fragmentos de polinucleótidos (por ejemplo, amplicones) bajo condiciones en las cuales las regiones complementarias entre los polinucleótidos pueden recoserse para formar a un complejo de hibridación, por ejemplo, que tenga una región hibridada de doble hebra con proyecciones para las regiones no complementarias. Una pluralidad de polinucleótidos puede ensamblarse para formar a un polinucleótido más grande que codifica a un polipéptido de interés.

35 El término “Polinucleótido de conexión” se refiere a un polinucleótido que tiene regiones complementarias en las regiones terminales de tal forma que el polinucleótido puede recoserse a una región terminal y otro polinucleótido puede recoserse a la otra región terminal del polinucleótido de conexión.

40 El término “Secuencia codificadora” se refiere a la porción de un ácido nucleico (por ejemplo, un gen) que codifica a una secuencia de aminoácidos de una proteína.

45 El término “codón optimizado” se refiere a cambios en los codones del polinucleótido que codifica a una proteína para aquellos utilizados preferencialmente en un organismo particular de tal forma que la proteína codificada es expresada eficientemente en el organismo de interés. Aunque el código genético es degenerado en que la mayoría de aminoácidos son representados por algunos codones, denominados “codones sinónimos”, es bien conocido que el uso de codones por organismos particulares no es algo aleatorio y tiene una tendencia hacia tripletes específicos de codones. Esta tendencia de uso de codones podría ser más alta en referencia a un gen específico, genes de funciones comunes o de orígenes ancestrales, proteínas altamente expresadas versus proteínas de pocos números de copias, y las regiones de codificación de proteínas acumuladas del genoma de un organismo.

50 El término “Complementarios” se refiere a la hibridación o emparejamiento de bases entre nucleótidos o ácidos nucleicos, tal como, por ejemplo, entre las 2 hebras de una molécula de ADN de doble hebra o entre un cebador de oligonucleótidos y un lugar de enlaces de cebadores en un polinucleótido de una sola hebra que va a ser secuenciado o amplificado. Los nucleótidos complementarios son, generalmente, A y T (o A y U), o C y G. 2 moléculas de ARN o de ADN de una sola hebra son conocidas por ser sustancialmente complementarias cuando una hebra de polinucleótidos (ARN o ADN) se hibrida bajo condiciones selectivas de hibridación a su complemento. Comúnmente, la hibridación selectiva ocurrirá cuando existe por lo menos alrededor de un 65% de complementariedad sobre una franja de por lo menos 14 a 25 nucleótidos, preferiblemente con por lo menos alrededor del 75%, más preferiblemente con por lo menos alrededor del 90% de complementariedad. Refiérase a, M. Kanehisa, 1984, Nucleic Acids (Ácidos Nucléicos) Res. 12:203. El término “complementario a” se utilizan en este documento con el propósito de indicar que una secuencia complementaria es sustancialmente idéntica o idéntica al complemento reverso de todo o de una porción de una secuencia referencial de polinucleótidos o que cada nucleótido en una hebra es capaz de formar a una pareja de bases con un nucleótido, o su análogo en la hebra opuesta.

El término “Sustituciones conservadoras de aminoácidos” se refiere a la intercambiabilidad de residuos que tienen cadenas laterales similares, y, por lo tanto, involucra, comúnmente, la sustitución del aminoácido en el polipéptido con aminoácidos dentro de la misma clase o dentro de una clase similar de aminoácidos. En forma de ejemplo, y sin el propósito de limitar, un aminoácido con una cadena lateral alifática puede ser sustituido con otro aminoácido alifático, por ejemplo, alanina, valina, leucina, isoleucina, y metionina; un aminoácido con una cadena lateral de hidroxilo es sustituido con otro aminoácido con una cadena lateral de hidroxilo, por ejemplo, serina y treonina; un aminoácido que tiene cadenas laterales aromáticas es sustituido con otro aminoácido que tiene cadenas laterales aromáticas, por ejemplo, fenilalanina, tirosina, triptófano, e histidina; un aminoácido con una cadena lateral básica es sustituido con otro aminoácido con una cadena lateral básica, por ejemplo, lisina, arginina, e histidina; un aminoácido con una cadena lateral ácida es sustituido con otro aminoácido con una cadena lateral ácida, por ejemplo, ácido aspártico o ácido glutámico; y un aminoácido hidrófobo o hidrofílico es reemplazado con otro aminoácido hidrófobo o hidrofílico, respectivamente.

El término “Secuencia de control” se refiere a una secuencia de polinucleótidos utilizada para efectuar la expresión de secuencias codificadoras y no calificadoras a las cuales están asociadas. La naturaleza de aquellas secuencias de control difiere dependiendo del organismo anfitrión. Las secuencias de control generalmente incluyen al promotor, al lugar de enlaces de ribosomas y a la secuencia de terminación de transcripción. El término “secuencias de control” tiene el propósito de incluir a componentes cuya presencia puede influenciar la expresión, y también puede incluir a componentes adicionales cuya presencia es ventajosa, por ejemplo, secuencias líderes y secuencias de compañeros de fusión.

El término “Diferencia definida” tal como se utiliza en este documento en el contexto de una mutación de un polinucleótido o de una secuencia de polipéptidos se refiere a un cambio previamente especificado, seleccionado y/o deseado a la secuencia (por ejemplo, un cambio de nucleótidos de c a g en una posición seleccionada o de una secuencia de polinucleótidos que resulta en un aminoácido diferente en una posición deseada del polipéptido codificado).

El término “Codón degenerado” tal como se utiliza en este documento se refiere a un codón utilizado para representar un conjunto de diferentes codones (también referido como una “codón ambiguo”). Por ejemplo, el codón degenerado “NNT” representa a un conjunto de 16 codones que tienen la secuencia triple de bases (A, C, T o G)/(A, C, T o G)/T, que codifica a un conjunto de 15 aminoácidos diferentes: F, S, Y, C, L, P, H, R, I, T, N, V, A, D, y G. Codones degenerados de ejemplo que son bien conocidos en la industria y que son útiles en los métodos de este invento incluyen a: NHT, NNB, NNG, NNK, NNN, NNS, NNT, NDT, RMG, RNG, RRS, SNT, VNS, VNT, y VWG.

El término “Eliminación” en referencia a un polipéptido o a un polinucleótido se refiere a la remoción de uno o más aminoácidos o nucleótidos del polipéptido o polinucleótido referencial, respectivamente. Las eliminaciones pueden comprender la remoción de uno o más aminoácidos o nucleótidos, 2 o más aminoácidos o nucleótidos, 3 o más aminoácidos o nucleótidos, 5 o más aminoácidos, 6 o más aminoácidos o nucleótidos, 10 o más aminoácidos o nucleótidos, 15 o más aminoácidos o nucleótidos, o 20 o más aminoácidos o nucleótidos, hasta el 10% del número total de aminoácidos o nucleótidos, o hasta el 20% del número total de aminoácidos o nucleótidos que componen al polipéptido o al polinucleótido referencial. Las eliminaciones pueden ser dirigidas a las porciones internas y/o a las porciones terminales del polipéptido o del polinucleótido. En varias realizaciones, la eliminación puede comprender a un seguimiento continuo o puede ser discontinuo.

El término “Heterólogo” cuando se lo utiliza con referencia a un ácido nucleico o a un polipéptido, indica que la secuencia que comprende a 2 o más sub - secuencias no se encuentran en la misma relación que en la que son encontrados normalmente en la naturaleza, o son diseñadas recombinantemente para que su nivel de expresión, o relación física en relación a otros ácidos nucleicos u otras moléculas en una célula, o estructura, no se encuentra normalmente en la naturaleza. Por ejemplo, un ácido nucleico heterólogo es comúnmente producido en una forma recombinante, teniendo 2 o más secuencias de genes no relacionados configuradas en una forma que no se encuentra en la naturaleza, por ejemplo, un marco referencial abierto de lectura (ORF - open reading frame) de ácidos nucleicos del invento ligado operativamente a una secuencia promotora insertada en una cinta de expresión, por ejemplo, un vector.

El término “Inserción” o “Adición” se refiere a un cambio en una secuencia de nucleótidos o de aminoácidos por medio de la adición de uno o más residuos de nucleótidos o de aminoácidos, respectivamente, en comparación a una secuencia referencial, tal como, por ejemplo, una secuencia de tipo silvestre.

El término “Biblioteca” se refiere a un conjunto (por ejemplo, una pluralidad) de polipéptidos o ácidos nucleicos heterogéneos. Una biblioteca está compuesta de miembros, cada uno de los cuales tiene una sola secuencia de polipéptidos o de ácidos nucleicos. En tal magnitud, la “biblioteca” es sinónimo a de un “repertorio”. Diferencias secuenciales entre miembros de bibliotecas son causantes de la diversidad presente en la biblioteca. La biblioteca podría tomar la forma de una simple mezcla de polinucleótidos o de ácidos nucleicos, o podría estar en la forma de organismos o células, por ejemplo, bacterias, virus, células animales o vegetales y similares, transformadas con una biblioteca de ácidos nucleicos.

El término "Sustitución no conservadora" se refiere a la sustitución de un aminoácido en el polipéptido con un aminoácido con propiedades significativamente diferenciadoras de cadenas laterales. Sustituciones no conservadoras podrían utilizar aminoácidos entre los grupos definidos en vez de dentro de ellos y afectan a (a) la esencia de la estructura del péptido en el área de la sustitución (por ejemplo, prolina en vez de glicina) (b) la carga o la hidrofobia, o (c) el aglutinamiento de la cadena lateral. En forma de ejemplo, pero sin limitarse, una sustitución no conservadora de ejemplo puede ser un aminoácido sustituido con un aminoácido básico o alifático; un aminoácido aromático sustituido con un aminoácido pequeño; y un aminoácido hidrofílico sustituido con un aminoácido hidrófobo.

El término "Que ocurre naturalmente" o "De tipo silvestre" se refiere a la forma encontrada en la naturaleza. Por ejemplo, una secuencia de polipéptidos o de polinucleótidos que ocurre naturalmente o de tipo silvestre es una secuencia presente en un organismo que puede ser aislada de una fuente en la naturaleza y que no ha sido modificada intencionalmente por medio de manipulación humana.

El término "Bases nitrogenadas" o "Base" se refiere a aquellas partículas que ocurren naturalmente y aquellas partículas heterocíclicas sintéticas conocidas comúnmente por aquellas personas que utilizan la tecnología de ácidos nucleicos y de polinucleótidos o utilizan la tecnología de ácidos nucleicos de poliamidas o de péptidos para generar, por lo tanto, a polímeros que pueden hibridarse a polinucleótidos en una forma específica a secuencias. Ejemplos no limitantes de bases nitrogenadas adecuadas incluyen a: la adenina, la citosina, la guanina, la timina, el uracilo, el 5-propinil-uracilo, el 2-tio-5-propinil-uracilo, la 5-metilcitosina, la pseudoisocitosina, el 2-tiouracilo y la 2-tiotimina, la 2-aminopurina, la N9-(7-deaza-guanina) y la N8-(7-deaza-8-aza-adenina). Otros ejemplos no limitantes de bases nitrogenadas adecuadas incluyen a aquellas bases nitrogenadas ilustradas en las figuras 2 (A) y 2 (B) o Buchardt et al. (WO 92/20702 o WO 92/20703).

El término "Polímero de bases nitrogenadas" u "Oligómero" se refiere a 2 o más bases nitrogenadas que están conectadas por enlaces que le permiten al polímero u oligómero resultante de bases nitrogenadas hibridarse a un polinucleótido que tiene una secuencia complementaria de bases nitrogenadas. Polímeros u oligómeros de bases nitrogenadas incluyen, pero no se limitan a, poli- y oligonucleótidos (por ejemplo, polímeros y oligómeros de ADN y de ARN), análogos de poli- y oligonucleótidos y miméticos de poli- y oligonucleótidos, tales como ácidos nucleicos de poliamidas o de péptidos. Polímeros u oligómeros de bases nitrogenadas pueden variar de tamaño desde pocas bases nitrogenadas, desde 2 a 40 bases nitrogenadas, a algunos cientos de bases nitrogenadas, a algunos miles de bases nitrogenadas, o más.

El término "Enlazados operativamente" se refiere a una relación funcional entre 2 o más segmentos de ácidos nucleicos (por ejemplo, ADN). En algunas realizaciones, se refiere a la relación funcional de una secuencia regulatoria de transcripciones a una secuencia transcrita. Por ejemplo, un promotor (que se define más adelante) está enlazado operativamente a una secuencia codificadora, tal como un ácido nucleico del invento, si estimula o modula la transcripción de la secuencia codificadora en una célula anfitriona apropiada u otro sistema de expresión. Generalmente, las secuencias promotoras regulatorias de transcripciones que son enlazadas operativamente a una secuencia transcrita están físicamente contiguas a la secuencia transcrita, es decir, están cis-actuando. Sin embargo, algunas secuencias regulatorias de transcripciones, tales como las mejoradoras, no necesitan estar físicamente contiguas o estar ubicadas en una proximidad cercana a las secuencias codificadoras cuya transcripción estas mejoran. En algunas realizaciones, la secuencia regulatoria es una secuencia regulatoria de interpretación enlazada a una secuencia codificadora.

El término "Región de superposición" se refiere a una región de un primer polinucleótido que es complementaria a un 2º polinucleótido, donde las regiones de superposición son capaces de recoserse entre sí para formar a un complejo de hibridación. Generalmente, el primer polinucleótido y el 2º polinucleótido estarán parcialmente superpuestos de tal forma que los polinucleótidos tendrán regiones no complementarias que no se recosen entre los 2 polinucleótidos.

El término "Permutaciones" se refiere a la configuración de elementos (por ejemplo, mutaciones de sustitución) provenientes de un conjunto finito específico. En el contexto de las descripciones aquí indicadas para polipéptidos y polinucleótidos, las diferencias en los residuos de aminoácidos o en los residuos de nucleótidos en comparación con una secuencia referencial, son caracterizadas comúnmente como "mutaciones", y pueden ser configuradas en varias combinaciones en la secuencia para formar a una permutación del conjunto de mutaciones. Las permutaciones incluyen a mutaciones individuales, así como a cada combinación posible de mutaciones de un conjunto definido.

El término "Polinucleótidos" u "Oligonucleótidos" se refiere a polímeros u oligómeros de bases nitrogenadas en los cuales las bases nitrogenadas están conectadas por medio de enlaces de fosfato de azúcar (estructura de azúcar-fosfato). Poli- y oligonucleótidos de ejemplo incluyen a polímeros de 2'deoxirribonucleótidos (ADN) y polímeros de ribonucleótidos (ARN). Un polinucleótido puede estar compuesto enteramente de ribonucleótidos, enteramente de 2'deoxirribonucleótidos o sus combinaciones.

El término “Polinucleótidos” o “Análogo de oligonucleótido” se refiere a polímeros u oligómeros de bases nitrogenadas en los cuales las bases nitrogenadas están conectadas por una estructura de fosfato de azúcar que comprende a uno o más análogos de fosfato de azúcar. Análogos comunes de fosfato de azúcar incluyen, pero no se limitan a, alquilfosfonatos de azúcar, fosforamiditas de azúcar, alquilfosfotriésteres sustituidos o no sustituidos de azúcar, fosfotioatos de azúcar, fosforoditioatos de azúcar, fosfatos de azúcar y análogos de fosfato de azúcar en los cuales el azúcar es uno diferente a 2'-deoxirribosa o ribosa, polímeros de bases nitrogenadas que tienen enlaces positivamente cargados de azúcar-guadinilo tales como aquellos descritos en la patente de Estados Unidos número 6,013,785 y la patente de Estados Unidos número 5,696,253 (refiérase también a Dagani 1995, Chem. Eng. News (Noticias) 4-5:1153; Dempey et al., 1995, J Am Chem Soc 117:6140-6141). Aquellos análogos cargados positivamente en los cuales el azúcar es 2'-deoxirribosa son denominados “DNGs”, mientras que aquellos en los cuales el azúcar es ribosa son denominados “RNGs”. Incluidos específicamente dentro de la definición de análogos de poli- y oligonucleótidos son los ácidos nucleicos bloqueados (LNAs - locked nucleic acids; refiérase a, por ejemplo, Elayadi et al., 2002, Biochemistry (Bioquímica) 41:9973-9981; Koshkin et al., 1998, J Am Chem Soc 120:13252-3; Koshkin et al., 1998, Tetrahedron Letters (Cartas de Tetraedro) 39:4381-4384; Jumar et al., 1998, Bioorganic & Medicinal Chemistry Letters (Cartas de Química Inorgánica y Medicinal) 8:2219-2222; Singh y Wengel, 1998, Chem. Commun., 12:1247-1248; WO 00/56746; WO 02/28875; y, WO 01/48190).

El término “Cebadores” se refiere a oligonucleótidos que tienen secuencias complementarias a una secuencia objetivo, generalmente denominada como una secuencia enlazadora de cebadores. La porción complementaria de un cebador puede ser de cualquier longitud que sea compatible con una hibridación específica y estable entre el cebador y la secuencia objetivo bajo las condiciones de reacción. Los cebadores pueden ser de entre 5 y 60 nucleótidos de largo, de alrededor de entre 10 a 35 nucleótidos de largo, o pueden ser de desde, y en particular, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, y/o 20 nucleótidos de largo. Generalmente, los cebadores para replicaciones por polimerasas son capaces de apoyar extensiones por las polimerasas cuando el cebador es recosido a la secuencia objetivo. El término “cebador de amplificación” se refiere a un cebador de oligonucleótidos utilizado para la amplificación de una secuencia objetivo de ácidos nucleicos.

El término “Cebador directo” y “Cebador inverso” se refiere a un conjunto de cebadores de amplificación, donde un cebador se recosa a un extremo 3' del objetivo (hebra plantilla) mientras que el otro cebador se recosa al extremo 3' de la hebra objetiva complementaria para amplificar un amplicón.

El término “Cebador degenerado” tal como se utilizan este documento se refiere a un cebador que incluye a por lo menos un codón degenerado en una posición de una diferencia definida. Por lo tanto, aunque el término “cebador degenerado” es singular se refiere a un conjunto de oligonucleótidos que tienen secuencias idénticas excepto en la posición de una diferencia definida. Adicionalmente, una “cebador degenerado” como se utiliza en este documento puede incluir a más de un codón degenerado (por ejemplo, NNT/WWG) y/o una mezcla de codones específicos (por ejemplo, TGG/TTC, o NNTNWG/TGG) en la posición de la diferencia definida. Cebadores degenerados pueden ser utilizados en reacciones PCR para generar a “amplicones degenerados” que pueden ser utilizados entonces para ensamblar y replicar a “variantes de polinucleótidos degenerados” de acuerdo a los métodos de esta divulgación (por ejemplo, variantes de polinucleótidos que codifican a una biblioteca de variantes de polipéptidos que tienen a todos los 20 aminoácidos naturales en cada una de algunas posiciones definidas de secuencias de aminoácidos).

El término “Próximo” se refiere a la distancia de nucleótidos desde una base definida (por ejemplo, una primera mutación de nucleótidos) a una 2ª base definida (por ejemplo, una 2ª mutación de nucleótidos), donde la primera y 2ª mutaciones pueden ser acomodadas en un solo cebador de oligonucleótidos utilizado para propósitos de amplificación (por ejemplo, cebador directo o inverso). Por lo tanto, en algunas realizaciones, el término “próximo” es determinado en relación a la longitud del cebador. En algunas realizaciones, 2 mutaciones pueden ser próximas si son separadas por 1, 2, 3, 4, 5, 6, 8, 10, 12, 14, 16, 18, 20, o 25 bases nitrogenadas y están dentro del cebador. En algunas realizaciones, la ubicación de las mutaciones en relación al extremo 3' del cebador es tal que un oligonucleótido recosido a una hebra plantilla puede experimentar una extensión por polimerasas, tal como se describe en mayor detalle más adelante.

Los términos “Proteína”, “Polipéptido”, “Oligopéptido” y “Péptido” se utilizan intercambiamente para denotar un polímero de por lo menos 2 aminoácidos enlazados covalentemente por un enlace de amidas, sin importar la longitud o modificación post-interpretación (por ejemplo, una glicosilación, una fosforilación, una lipidación, una miristoilación, una ubiquitinación, etc.). Incluidas con esta definición se encuentran a los D- y L-aminoácidos, y las mezclas de D- y L-aminoácidos.

El término “Recombinante” cuando se lo utiliza en referencia, por ejemplo, a una célula, a un ácido nucleico, a un polipéptido, a una cinta o vector de expresión, se refiere a un material, o un material correspondiente a la forma natural o nativa del material, que ha sido modificada por la introducción de una nueva partícula o alteración de una partícula existente por medio de técnicas recombinantes, o es idéntico a este pero producido o derivado de materiales sintéticos utilizando técnicas recombinantes. Por ejemplo, las células recombinantes se expresan a genes que no se encuentran dentro de la forma nativa (no recombinante) de la célula (es decir, “ácidos nucleicos

exógenos”) o expresan genes nativos que son expresados de otra forma en un nivel diferente, comúnmente, sub-expresados o no expresados en absoluto.

5 El término “Secuencia referencial” se refiere a una secuencia definida utilizada como una base para una comparación secuencial. Una secuencia referencial podría ser un subconjunto de una secuencia más grande, por ejemplo, un segmento de un gen completo o una secuencia completa de polipéptidos. Generalmente, una secuencia referencial tiene por lo menos 20 nucleótidos o residuos de aminoácidos de longitud, por lo menos 25 residuos de longitud, por lo menos 50 residuos de longitud, o la longitud completa del ácido nucleico o polipéptido. Puesto que 2 polinucleótidos o polipéptidos podrían cada uno (1) comprender a una secuencia (es decir, una porción de la secuencia completa) que es similar entre las 2 secuencias, y (2) podría comprender además una secuencia que es divergente entre las 2 secuencias, las comparaciones secuenciales entre 2 (o más) polinucleótidos o polipéptidos son realizadas comúnmente al comparar a secuencias de los 2 polinucleótidos o polipéptidos a lo largo de una “ventana de comparación” para identificar y comparar a regiones locales de similitud secuencial.

15 El término “Replicante” se refiere a copiar de una secuencia objetivo de polinucleótidos para sintetizar una copia complementaria inversa del polinucleótidos. Generalmente, una replicación es hecha por polimerasas que copian a un polinucleótido plantilla para sintetizar un polinucleótido que es un complemento reverso de la secuencia objetiva de polinucleótidos.

20 El término “Segmento” se refiere a una secuencia que es una porción de una secuencia más grande de polinucleótidos. La secuencia más grande de polinucleótidos puede ser dividida en una pluralidad de segmentos, donde la combinación de los segmentos comprende a la longitud completa de la secuencia más grande de polinucleótidos.

25 El término “Variante de polipéptidos” o “Análogo de polipéptidos” tal como se utiliza en este documento se refiere a polipéptidos que son comparados de un segmento que tiene una actividad funcional, con o sin la retención de cualquier propiedad mejorada, y tienen una identidad sustancial a una porción de un polipéptido referencial. En algunas realizaciones, polipéptidos análogos comprenden a sustituciones conservadoras o no conservadoras de aminoácidos, o la adición o eliminación de uno o más residuos de aminoácidos en relación a la secuencia referencial.

30 El término “Emparejamiento de bases de Watson/Crick” se refiere a un patrón de parejas específicas de bases nitrogenadas y de análogos que se enlaza entre sí a través de enlaces de hidrógeno de secuencias específicas, por ejemplo, A se empareja con T y U, y G se empareja con C.

35 El término “Sustitución” se refiere al reemplazo de uno o más nucleótidos de aminoácidos por nucleótidos o aminoácidos diferentes, respectivamente, en relación a una secuencia referencial, tal como, por ejemplo, una secuencia de tipo silvestre.

40 El término “Sustrato”, “Soporte”, “Soporte sólido”, “Portador sólido” o “Resina” son términos intercambiables y se refieren a cualquier material de fase sólida. El término sustrato también abarca a términos tales como “fase sólida”, “superficie”, y/o “membrana”. Un soporte sólido puede ser compuesto de polímeros orgánicos tales como, el poliestireno, el polietileno, el polipropileno, el polifluoroetileno, el polietilenooxi, y poliacrilamida, así como sus copolímeros e injertos. Un soporte sólido también puede ser inorgánico, tal como el vidrio, la sílice, vidrio de poros controlados (CPG - controlled pore glass), sílices de fase inversa o metales, tales como el oro o el platino. La configuración de un sustrato puede ser en la forma de micropartículas, esferas, partículas, gránulos, un gel, una membrana o una superficie. Las superficies pueden ser planas, sustancialmente planas, o no planas. Los soportes sólidos pueden ser porosos o no porosos, y pueden tener características de hinchamiento o de no hinchamiento. Un soporte sólido puede ser configurado en la forma de un pozo, una depresión, u otro tipo de contenedor, envase, característica o ubicación. Una pluralidad de soportes puede ser configurada en una formación en varias ubicaciones, a las que se les puede asignar direcciones referenciales para la entrega robótica de reactivos, o por medio de métodos y/o instrumentos de detección.

50 El término “Formación” se refiere a una configuración de agentes (por ejemplo, proteínas, anticuerpos, empaques genéticos replicables) en un sustrato en ubicaciones posicionalmente distintas. En algunas realizaciones, los agentes en la formación son codificados espacialmente para que la identidad de un agente pueda determinarse a partir de su ubicación en la formación. Una “micro-formación” generalmente se refiere a una formación en la cual la detección requiere el uso de detección microscópica para detectar a complejos formados con agentes en el sustrato. Una “ubicación” en una formación se refiere a un área localizada en la superficie de la formación que incluye a agentes, cada uno definido para que pueda ser distinguido de ubicaciones adyacentes (por ejemplo, siendo posicionados en la formación general, o que tienen algunas características detectables, que permiten que la ubicación pueda ser distinguida de otras ubicaciones). La ubicación puede tener cualquier forma conveniente (por ejemplo, en forma circular, rectangular, elíptica o de cuña). El tamaño o el área de la ubicación puede variar significativamente. Las formaciones pueden ser construidas en sustratos, tales como diapositivas de vidrio o de plástico de microscopios, y pueden ser configuradas en la forma de pozos, depresiones, gotas u otros contenedores,

o portadores de reacción, tal como un pozo de micro - placas. En general, no existen restricciones en el formato de la formación siempre y cuando los lugares individuales en los cuales los agentes están colocados puedan ser ubicados e identificados.

5 El término “Cámara de reacción” se refiere al entorno en el cual los agentes y/o los componentes de reacción están ubicados. Contenedores de reacción que están disponibles comercialmente contienen a por lo menos una cámara de reacción, pero pueden contener a 8, 24, 96 o 384 cámaras de reacción. Para propósitos de esta divulgación, el término “cámara o cámaras de reacción”, “pozo o pozos”, “lugar o lugares de reacción” son utilizados intercambiamente. Un ejemplo de una cámara de reacción es una de 96 pozos de micro - titulación en una placa de micro - titulación de 96 pozos.

10 El término “Formación de cebadores” se refiere a la formación de cebadores o de conjuntos de cebadores utilizados en una reacción de amplificación en ubicaciones posicionalmente distintas en un sustrato (por ejemplo, un sustrato de formaciones). Generalmente, el conjunto de cebadores comprende a una pareja de cebadores directos e inversos usada para amplificar a un amplicón.

15 El término “Formación de amplicones” se refiere a una configuración de polinucleótidos amplificados en ubicaciones posicionalmente distintas en un sustrato (por ejemplo, un sustrato de formaciones). En algunas realizaciones, la formación de amplicones puede tener la misma configuración posicional que la primera formación, para cuando la reacción de amplificación es ejecutada en la formación de cebadores para generar a polinucleótidos amplificados.

20 El término “Plásmidos”, “Vector” y “Cinta” se refiere a un elemento cromosómico extra que a menudo porta a genes, y usualmente en la forma de moléculas circulares de ADN de doble hebra. Aquellos elementos podrían ser secuencias autónomamente replicantes, secuencias integradoras de genoma, secuencias de fagos o de nucleótidos, lineares o circulares, de un ADN o ARN de una sola o de doble hebra, derivadas de cualquier fuente, en la cual un número de secuencias de nucleótidos ha sido juntado o recombinado en una construcción única que es capaz de introducir a un fragmento promotor y a una secuencia de ADN para un producto genético seleccionado junto con secuencias apropiadas no interpretadas de 3’ en una célula. El término “cinta de expresión” se refiere a un vector específico que contiene a un gen foráneo y tiene elementos adicionalmente al gen foráneo que permiten la expresión de un gen en un anfitrión.

5.2 Métodos para la síntesis de variantes de polinucleótidos

25 Esta divulgación suministra métodos para generar variantes de polinucleótidos que tienen un conjunto definido de diferencia secuencial en comparación con una secuencia referencial de polinucleótidos. En algunas realizaciones, los métodos son aplicables para generar a polinucleótidos que codifican a variantes de polipéptidos que tienen diferencias definidas en secuencias de aminoácidos en comparación a un polipéptido referencial. En algunas realizaciones, las variantes de polinucleótidos tienen un conjunto definido de diferencias de nucleótidos en regiones no codificadoras, por ejemplo, mutaciones silenciosas. Los polinucleótidos son generados eficientemente al utilizar bibliotecas de fragmentos de polinucleótidos, donde los miembros de la biblioteca codifican a una o más de las diferencias de aminoácidos en comparación a una secuencia referencial de polipéptidos, y los fragmentos de polinucleótidos son diseñados para tener regiones adyacentes superpuestas para que la selección de un conjunto apropiado de fragmentos, con y sin mutaciones, permita su ensamblaje a una variante de polinucleótidos, tal como un polinucleótido que codifica a una variante deseada de polipéptidos.

30 En algunas realizaciones, el método para generar a un polinucleótidos que codifica a un polipéptido que tiene una secuencia de aminoácidos con una o más diferencias definidas en residuos de aminoácidos, comprende (a) seleccionar una pluralidad de diferencias definidas de residuos de aminoácidos en relación a una secuencia referencial de aminoácidos; (b) definir segmentos superpuestos de una secuencia de polinucleótidos que codifican al polipéptido con la secuencia diferente de aminoácidos, u opcionalmente el polipéptido referencial, siendo enlazado cada segmento por un conjunto de secuencias enlazadoras de cebadores directos e inversos, donde una diferencia secuencial de polinucleótidos que codifica a cada una de las pluralidades de diferencias de residuos de aminoácidos es abarcada en las secuencias de los cebadores directos y/o inversos que enlazan a las secuencias enlazadoras de cebadores; (c) amplificar a cada segmento con el conjunto de cebadores directos e inversos, donde el cebador seleccionado directo y/o inverso contiene a las diferencias secuenciales de polinucleótidos, para generar una biblioteca de amplicones que comprende a miembros que codifican a las diferencias definidas de aminoácidos y donde la biblioteca comprende a miembros suficientes para ensamblar a 2 o más permutaciones diferentes de secuencias de aminoácidos de las diferencias definidas de residuos de aminoácidos; (d) ensamblar a partir de la biblioteca a un conjunto de amplicones que tienen regiones complementarias adyacentes que codifican en conjunto al polipéptido con una permutación definida de secuencias de aminoácidos que tienen uno o más diferencias definidas de residuos de aminoácidos; y (e) replicar un conjunto de amplicones ensamblados para sintetizar a polinucleótidos que codifican al polipéptido. Una biblioteca de amplicones que contiene a todas las diferencias definidas de aminoácidos debería permitir la síntesis de una pluralidad de polinucleótidos que codifica a todas las permutaciones posibles de secuencias de aminoácidos.

Tal como será aparente para una persona con conocimiento en la industria, seleccionar la pluralidad de diferencias definidas de residuos de aminoácidos puede obtenerse de varias fuentes. En algunas realizaciones, las posiciones y las mutaciones correspondientes de residuos de aminoácidos para un polipéptido definido pueden obtenerse de estudios de mutagénesis aleatoria, tales como aquellos descritas en Crameri et al., 1998, "DNA shuffling of a family of genes from diverse species accelerates directed evolution" ("La mezcla aleatoria de ADN de una familia de genes de especies diversas aceleran la evolución dirigida"), *Nature (Naturaleza)* 391:288-291; Crameri et al., 1997, "Molecular evolution of an arsenate detoxification pathway by DNA shuffling" ("La evolución molecular de un sendero de desintoxicación de arseniato por medio de mezclas aleatorias de ADN", *Nature (Naturaleza) Biotech* 15:436-438; Zhang et al., 1997, "Directed evolution of an effective fructosidase from a galactosidase by DNA shuffling and screening" ("Evolución dirigida de una fructosidasa efectiva proveniente de una galactosidasa por medio de mezclas aleatorias y revisiones de ADN"), *Proc Natl Acad Sci EE.UU.* 94:45-4-4509; Crameri et al., 1996, "Improved green fluorescent protein by molecular evolution using DNA shuffling" ("Proteína mejorada de fluorescencia verde por medio de evolución molecular utilizando mezclas aleatorias de ADN", *Nature (Naturaleza) Biotech* 14:315-319; Stemmer, 1994, "Rapid evolution of a protein in vitro by DNA shuffling" ("Evolución rápida de una proteína in vitro por medio de mezclas de ADN"), *Nature (Naturaleza)* 370:389-391; Stemmer, 1994, "DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution" ("Mezclas aleatorias de ADN por medio de fragmentaciones y re-ensamblajes aleatorios: Recombinación in vitro para evolución molecular"), *Proc Natl Acad Sci Estados Unidos de América* 91:10747-10751; WO 95/22625; WO 97/0078; WO 97/35966; WO 98/27230; WO 00/42651; WO 01/75767 y U.S. Pat. 6,537,746.

Comúnmente, una biblioteca mutada de polinucleótidos es expresada y los polipéptidos expresados son examinados para detectar un rasgo característico deseado, y la mutación asociada con los cambios en la propiedad deseada son identificados. Una gran cantidad de mutaciones que afectan a una función de polipéptidos puede obtenerse fácilmente utilizando estas técnicas.

En algunas realizaciones, las opciones de diferencias de residuos de aminoácidos pueden obtenerse a partir de comparaciones de secuencias de aminoácidos de proteínas relacionadas, tales como aquellas encontradas en una base de datos de secuencias. Las comparaciones de secuencias pueden identificar posiciones, por ejemplo, residuos conservados que podrían ser importantes para la función proteínica, los cuales pueden ser usados entonces como objetivos para cambios definidos de aminoácidos. Refiérase, por ejemplo, Wankhade et al., 2000, *J. Biol. Chem.* 275(38):29701-29708 y Reddy et al., 2001, *Proteins: Structure, Function, and Genetics (Proteínas: Estructura, función y genética)* 42:148-163.

En algunas realizaciones, la pluralidad de diferencias definidas de residuos de aminoácidos puede basarse en diferencias secuenciales encontradas en la naturaleza, tales como polimorfismos encontrados en un gen específico. En algunas instancias, los polimorfismos son asociados con efectos específicos biológicos y los fenotipos asociados. Refiérase a, por ejemplo, Bidwell et al., 1999, *Genes and Immunity (Genes e inmunidad)* 1:3-19; Chen et al., 2003, *Mol. Biol. Evo.* 18:1771-1788. Colecciones de polimorfismos pueden formar la base para definir una pluralidad de diferencias de residuos de aminoácidos para formar polipéptidos variantes de una secuencia referencial de aminoácidos. Combinaciones de diferentes polimorfismos de aminoácidos pueden ser utilizadas para examinar la función de una proteína específica.

En los casos en los que la pluralidad de diferencias de residuos de aminoácidos es definida en relación a una secuencia referencial, el polinucleótido que codifica al polipéptido referencial o a las variantes de polipéptidos puede ser utilizado como una base para identificar a los segmentos para definir a los amplicones que van a ser generados para crear a la biblioteca de fragmentos de polinucleótidos (es decir, amplicones). En algunas realizaciones, la secuencia de polinucleótidos no necesita ser restringida a ninguna secuencia particular mientras que codifique, o pueda ser utilizada como una base para generar a polinucleótidos que codifican, a la secuencia de aminoácidos de interés. El polinucleótido puede basarse en una secuencia que ocurre naturalmente (por ejemplo, de tipo silvestre) o una secuencia optimizada para la expresión en un organismo particular de interés (por ejemplo, optimizada en lo que se refiere a codones). Por ejemplo, si el polipéptido de interés va a ser expresado en *E. coli*, una secuencia de polinucleótidos en la cual los codones están optimizados para su expresión en *E. coli* puede ser utilizada. Las técnicas de optimización de codones están bien dentro de la destreza de aquellas personas con conocimiento en la industria.

Tal como será aparente para una persona con conocimiento en la industria, el dividir el polinucleótido en segmentos definidos para su amplificación puede lograrse utilizando técnicas bien conocidas en la industria. En algunas realizaciones, puesto que los segmentos son definidos por las secuencias enlazadoras de cebadores, que son en sí utilizadas para introducir a mutaciones al amplicón, la división del polinucleótido en segmentos puede tomar en cuenta inicialmente la ubicación de las mutaciones en el polinucleótido. La división del polinucleótido en segmentos también puede tomar en cuenta la longitud total del polinucleótido, la eficiencia de la replicación (por ejemplo, la amplificación de segmentos), y el número deseado de rincones para ensamblaje. Otras consideraciones serán aparentes para una persona con conocimiento en la industria.

Reacciones de amplificación pueden ser afectadas por la secuencia, por el tipo de polimerasa utilizada, por

la eficiencia de los cebadores, y por reacciones colaterales no deseadas (por ejemplo, dímeros de cebadores). Por lo tanto, en algunas realizaciones, dependiendo de la longitud total del polinucleótido que va a ser ensamblado, las solicitudes de los segmentos pueden ser de 200 bases o menos, de 1500 bases o menos, de 1200 bases o menos, de 1000 bases o menos, de 900 bases o menos, de 800 bases o menos, de 700 bases o menos, de 600 bases o menos, de 500 bases o menos, de 400 bases o menos, de 300 bases o menos, de 250 bases o menos, de 200 bases o menos, o de alrededor de 100 o tan pocas como alrededor de 50 bases de longitud. Generalmente, la longitud de los segmentos es desde alrededor de 50 a alrededor de 1000 bases, desde alrededor de 200 a alrededor de 1000 bases, desde alrededor de 300 a alrededor de 700 bases, o desde alrededor de 400 a 600 bases, con alrededor de 500 bases siendo el promedio útil de longitud dada la eficiencia de las polimerasas utilizadas en las reacciones de amplificación. En varias realizaciones, los segmentos están superpuestos de tal forma que los amplicones de ahí producidos también tendrán regiones adyacentes superpuestas (es decir, regiones complementarias superpuestas) para ensamblar al polinucleótido.

En algunas realizaciones, las regiones superpuestas adyacentes deberán tener una longitud y complementariedad suficientes para permitir la formación de amplicones recosidos establemente (es decir, hibridados) durante el ensamblaje del polinucleótido. Por lo tanto, en algunas realizaciones, la longitud de la superposición puede ser de 4 o más nucleótidos, de 5 o más nucleótidos, de 6 o más nucleótidos, de 8 o más nucleótidos, de 10 o más nucleótidos, de 15 o más nucleótidos, de 20 o más nucleótidos, de 25 o más nucleótidos, de 30 o más nucleótidos, de 40 o más nucleótidos, de 50 o más nucleótidos, y 100 o menos, 90 o menos, 80 o menos, 70 o menos, 60 o menos nucleótidos de longitud tal como lo permita la capacidad para formar amplicones recosidos estables. Puesto que las regiones de superposición generalmente incluyen a las secuencias enlazadoras de cebadores utilizadas para generar a los amplicones, la longitud de superposición puede tomar en cuenta cualquier diferencia en la secuencia del cebador (por ejemplo, directo y/o inverso) utilizado para generar las diferencias de polinucleótidos que codifican a la mutación que va a ser introducida.

En algunas realizaciones, los segmentos son enlazados por las secuencias enlazadoras de cebadores a las cuales los cebadores directos/ inversos se recosen. Cuando sea apropiado, las secuencias enlazadoras de cebadores que definen a los segmentos también pueden abarcar la posición del polinucleótido que codifica a una diferencia de secuencia de aminoácidos. La secuencia enlazadora de cebadores puede ser de cualquier longitud suficiente para recoserse al cebador (directo o inverso) durante la reacción de amplificación. Asimismo, la secuencia enlazadora de cebadores puede ser de 100 bases o menos, de 90 bases o menos, de 80 bases o menos, de 70 bases o menos, de 60 bases o menos, de 50 bases o menos, de 40 bases o menos, de 30 bases o menos, de 20 bases o menos, de 15 bases o menos, hasta alrededor de 8 bases o 10 bases. En algunas realizaciones, la longitud de las secuencias enlazadoras de cebadores puede comprender a desde alrededor de 8 a 50 bases, desde alrededor de 8 a 40 bases, desde alrededor de 10 a 30 bases, desde alrededor de 15 a 25 bases. Los cebadores comúnmente pueden comprender a longitudes complementarias de las secuencias enlazadoras de cebadores ya descritas. Asimismo, en algunas realizaciones, la longitud de los cebadores directos / inversos puede ser de alrededor de 60 nucleótidos o menos, de 50 nucleótidos o menos, de 40 nucleótidos o menos, de 30 nucleótidos o menos, de 20 nucleótidos o menos, de 15 nucleótidos o menos, hasta alrededor de 10 nucleótidos o incluso 8 nucleótidos. En algunas realizaciones, la longitud de los cebadores directos / inversos puede ser de desde alrededor de 8 a 50 nucleótidos, de desde alrededor de 8 a 40 nucleótidos, de desde alrededor de 10 a 30 nucleótidos, o de desde alrededor de 15 a 25 nucleótidos.

En los casos en los que el cebador contiene una secuencia que codifica a una diferencia definida de aminoácidos, la mutación puede ser ubicada en una región del cebador que no interfiere con la extensión del cebador. En algunas realizaciones, la mutación es ubicada en aproximadamente el medio del cebador mutagénico, donde el cebador tiene un Tm que es suficiente para recoserse al ácido nucleico plantilla y sirve como un cebador para la reacción de extensión regulada por la polimerasa. En algunas realizaciones, las diferencias de secuencias de polinucleótidos pueden estar ubicadas, dependiendo de la longitud del cebador, a alrededor de 5 bases, 6 bases, 8 bases, 10 bases, 12 bases, 15 bases, 20 bases, 25 bases del extremo 3' del cebador. Asimismo, en algunas realizaciones la longitud de los cebadores directos / inversos puede ser de desde alrededor de 8 a 50 nucleótidos, de desde alrededor de 8 a 40 nucleótidos, de desde alrededor de 10 a 30 nucleótidos, o de desde alrededor de 15 a 25 nucleótidos, y comprenden además a la diferencia de secuencias de nucleótidos en aproximadamente el medio del cebador. Por lo tanto, en algunas realizaciones los cebadores directos / inversos tienen una longitud de alrededor de 50 nucleótidos con una diferencia de nucleótidos a alrededor de 25 nucleótidos a partir del extremo 3', alrededor de 40 nucleótidos de longitud con una diferencia de nucleótidos a alrededor de 20 nucleótidos a partir del extremo 3', alrededor de 30 nucleótidos de longitud con una diferencia de nucleótidos a alrededor de 15 nucleótidos a partir del extremo 3', alrededor de 25 nucleótidos de longitud con una diferencia de nucleótidos a alrededor de 12 nucleótidos a partir del extremo 3', o alrededor de 20 nucleótidos de longitud con una diferencia de nucleótidos a alrededor de 10 nucleótidos a partir del extremo 3'.

La estabilidad de los cebadores de oligonucleótidos, por ejemplo, la temperatura de derretimiento térmico, es una función de la fortaleza iónica, de la temperatura, del contenido G/C, y de la presencia de agentes caotrópicos y puede calcularse utilizando métodos conocidos para predecir las temperaturas de derretimiento (refiérase a, por ejemplo, Baldino et al., *Methods Enzymology* (Enzimología métodos) 168:761-777; Bolton et al., 1962, *Proc. Natl.*

Acad. Sci. Estados Unidos de América 48:1390; Bresslauer et al., 1986, Proc. Natl. Acad. Sci Estados Unidos de América 83:8893-8897; Freier et al., 1986, Proc. Natl. Acad. Sci Estados Unidos de América 83:9373-9377; Kierzek et al., Biochemistry (Bioquímica) 25:7840-7846; Rychlik et al., 1990, Nucleic Acids (Ácidos nucleicos) Res 18:6409-6412 (erratum (errata), 1991, Nucleic Acids (Ácidos nucleicos) Res 19:698); Sambrook et al., 2001, Molecular Cloning: A Laboratory Manual (Clonación molecular: Un manual de laboratorio), 3ª Ed., Cold Spring Harbor Laboratory Press, NY; Suggs et al., 1981, In Developmental Biology Using Purified Genes (En biología del desarrollo utilizando a genes purificados) (Brown et al., eds.), pp. 683-693, Academic Press; y Wetmur, 1991, Crit Rev Biochem Mol Biol 26:227-259. Todas las publicaciones son incorporadas este documento por referencia).

Para generar la biblioteca de amplicones, cebadores directos e inversos que se recosen a las secuencias enlazadoras de cebadores de cada segmento del polinucleótido son utilizados en una reacción de amplificación para generar a amplicones. En los casos en los que el amplicón tiene una diferencia de polinucleótidos que codifica a un cambio definido de aminoácidos en relación a la secuencia referencial, la secuencia de los cebadores directos y/o inversos el diseñada para introducir a la secuencia diferente (es decir, a la mutación) en la reacción de amplificación.

En algunas realizaciones, los conjuntos de cebadores directos e inversos pueden ser almacenados en una formación, por ejemplo, en una formación de cebadores, para qué puedan ser accedidos fácilmente cuando los amplicones son necesitados para la síntesis de un polinucleótido que codifica a una permutación definida de secuencias de aminoácidos. Tal como será apreciado en la industria, los cebadores de oligonucleótidos pueden ser utilizados para introducir a cualquier tipo de mutación seleccionada en la pluralidad definida de diferencias de residuos de aminoácidos, incluyendo, entre otras, a inserciones, eliminaciones y sustituciones de aminoácidos. Las sustituciones pueden ser mutaciones conservadoras o no conservadoras, tal como lo dicte la pluralidad escogida de diferencias de residuos de aminoácidos.

Una ventaja de los métodos aquí presentados para sintetizar una pluralidad de variantes de polinucleótidos es que puede suministrar bibliotecas muy grandes de diferencias diversas, pero bien definidas de nucleótidos que permiten una exploración más eficiente de espacio de diversidad secuencial. Por ejemplo, bibliotecas grandes pueden ser generadas donde cada variante tiene únicamente una sola diferencia definida de nucleótidos. Aquellas bibliotecas suministran un estudio amplio de diversidad secuencial que puede ser examinado y analizado teniendo mayor seguridad de que las relaciones estructuras-funciones identificadas son precisas (por ejemplo, menos positivos falsos y negativos falsos). La capacidad de los métodos para generar bibliotecas grandes de secuencias de variantes de polinucleótidos con una alta precisión (por ejemplo, el 75%, el 85%, el 95% o más de secuencias correctas de longitudes completas) mejora enormemente las ventajas de generar y analizar a aquellas bibliotecas. Por lo tanto, en algunas realizaciones, la pluralidad de variantes de polinucleótidos sintetizados utilizando los métodos puede comprender a por lo menos 10, 20, 30, 40, 50, 100, 200, o más variantes diferentes de polinucleótidos, donde cada variante comprende a una diferencia definida de nucleótidos en una de 10, 20, 30, 40, 50, 100, 200, o más posiciones seleccionadas diferentes (es decir, 10, 20, 30, 40, 50, 100, 200, o más mutaciones distintas en un solo lugar).

En algunas realizaciones, más de una diferencia de secuencias de aminoácidos pueden estar presentes en la misma posición de residuos de aminoácidos en una secuencia de polipéptidos. En estas realizaciones, diferentes amplicones del mismo segmento superpuesto pueden ser generados, donde cada amplicón es preparado con parejas de cebadores directos e inversos para cada mutación definida en la posición residual idéntica. Para preparar a un polinucleótido que codifica a una permutación específica de secuencias en la posición específica de residuos de aminoácidos, uno de los amplicones que contienen a la mutación deseada (es decir, la diferencia definida de nucleótidos) es escogido y ensamblado como un miembro del conjunto de amplicones para generar al polinucleótido que codifica a un polipéptido que contiene a la mutación o mutaciones deseadas en la posición especificada de residuos de aminoácidos.

En algunas realizaciones, más de una pareja de cebadores (por ejemplo, un conjunto de cebadores degenerados) pueden ser utilizadas para generar a un conjunto de amplicones (es decir, fragmentos de polinucleótidos) que puede ser utilizado para ensamblar a un conjunto de variantes de polinucleótidos que codifica a polipéptidos que tienen más cambios de residuos de aminoácidos (por ejemplo, sustituciones) en una posición definida especificada. Las variantes de polinucleótidos ensambladas a partir de los amplicones elaborados utilizando a cebadores degenerados pueden ser secuenciadas antes o después de que su polipéptido codificado sea expuesto a ensayos para determinar la secuencia específica en la posición de interés.

En algunas realizaciones, una pluralidad de cebadores degenerados (por ejemplo, cada cebador que tiene un cordón diferente en una posición seleccionada para mutaciones) puede ser amplificado por separado (por ejemplo, una reacción de la ronda 1) para generar una pluralidad de "amplicones degenerados". Cada pozo separado (o "grupo") de amplicones así generados comprende a una pluralidad de amplicones diferentes con secuencias idénticas excepto en la posición de diferencias definidas las cuales son establecidas por las secuencias de los cebadores degenerados utilizados. Estos amplicones degenerados (que tienen una pluralidad de diferencias definidas de nucleótidos en una o más posiciones) pueden ser ensamblados entonces por separado con uno o más amplicones superpuestos (ya sean amplicones "homogéneos" o grupos de amplicones degenerados) y replicados

(por ejemplo, en una reacción SOE-PCR de la ronda 2) para generar a un pozo separado de “variantes degeneradas de polinucleótidos” de longitud completa, donde las variantes de polinucleótidos comprenden a secuencias variantes que comprenden a todas las diferencias definidas diseñadas en los amplicones degenerados utilizados en el primer paso.

5
 10
 15
 20
 25
 30

Por ejemplo, cebadores degenerados pueden ser diseñados para que tengan diferencias definidas de nucleótidos que comprenden a codones degenerados (por ejemplo, NNK, o el conjunto de 23 codones de NNT, VVG y TGG) que representan a 20 aminoácidos en 3 posiciones en cada uno de los 3 polipéptidos codificados. Estos cebadores degenerados pueden ser utilizados con una plantilla referencial de polinucleótidos en 6 reacciones separadas de PCR en la ronda 1 para generar 6 grupos de amplicones degenerados (es decir, una pareja de amplicones degenerados superpuestos para cada una de las 3 posiciones de la diferencia definida de nucleótidos). El ensamblar y replicar por separado a cada una de las 3 parejas de los amplicones degenerados superpuestos en 3 reacciones separadas de SOE-PCR de la ronda 2 resulta en 3 agrupaciones separadas de variantes de longitud completa de polinucleótidos donde cada grupo contiene a polinucleótidos que codifican a un conjunto de variantes degeneradas de polipéptidos que tienen a todos los 20 aminoácidos diferentes en cada una de las 3 posiciones seleccionadas usadas como objetivos para mutaciones. Aunque los 3 grupos separados pueden mantenerse como una biblioteca de variantes de polinucleótidos a la cual se le pueden asignar direcciones referenciales con 3 miembros que tienen una degeneración completa en cada una de las 3 posiciones seleccionadas, pueden lograrse ventajas sorprendentes de eficiencia en lo que se refiere a la transformación y la revisión (particularmente con bibliotecas más grandes) al destruir la capacidad de asignar direcciones referenciales y combinar a todos los 3 miembros de la biblioteca en una sola agrupación. Asimismo, los 3 grupos de variantes degeneradas de polinucleótidos son combinados lo cual resulta en una biblioteca agrupada de polinucleótidos que codifica a 60 polipéptidos comprendiendo a todos los aminoácidos posibles en cada una de las 3 posiciones objetivo. Esta biblioteca agrupada comprende a una diversidad de secuencias completas de aminoácidos que codifican a variantes de polinucleótidos de un conjunto bien definido (por ejemplo, mutagénesis de saturación) en las 3 posiciones objetivo. La biblioteca agrupada puede ser clonada entonces en un sistema de expresión, transformada en una sola reacción, colocada en placas, recogida y examinada. Puesto que el método de síntesis que involucra a los pasos de la ronda 1 y de la ronda 2 resulta en un nivel alto de precisión (por ejemplo, el 75%, el 85%, el 95% o más de las variantes deseadas de longitud completa) un número relativamente pequeño de colonias puede ser escogido y examinado para acceder al 75%, al 85%, al 95%, al 99% o incluso una cobertura más grande de revisión de la biblioteca de las variantes de polipéptidos.

35
 40

Tal como será aparente para una persona con conocimiento en la industria, en algunas realizaciones, los métodos que utilizan a cebadores degenerados (y los amplicones degenerados resultantes) pueden ser utilizados para preparar a variantes de polinucleótidos que tienen diferencias definidas en cada posición. En algunas realizaciones, se contempla que el método sea utilizado para suministrar bibliotecas de “mutagénesis de saturación” donde cada miembro de polinucleótidos variantes tienen una secuencia diferente que comprende a 1 de 23 codones diferentes en posiciones seleccionadas diferentes de la secuencia referencial de polinucleótidos. Por lo tanto, en algunas realizaciones, los métodos son capaces de suministrar una pluralidad de variantes de polinucleótidos que comprenden a por lo menos 23, 46, 69, 92, 115, 230, 460, 920 o más variantes diferentes de polinucleótidos donde cada una comprende a uno de 23 codones diferentes en una de 1, 2, 3, 4, 5, 10, 20, 40, o más posiciones seleccionadas diferentes. Una persona con conocimiento normal en la industria reconocerá que, en otras realizaciones, más o menos codones pueden ser utilizados en posiciones de diferencias definidas de nucleótidos y/o en varias posiciones seleccionadas y cada variante puede ser mutada utilizando los métodos aquí descritos.

45
 50
 55

Esta divulgación también indica en algunas realizaciones que el número general de variantes puede ser controlado mediante la mezcla de más o menos degeneración en los cebadores en distintas posiciones de diferencias definidas. Por lo tanto, el método permite una mutagénesis de saturación completa que puede ser ejecutada en una o más posiciones seleccionadas mientras que diferencias definidas individuales o subconjuntos más pequeños de degeneración son introducidos en otras posiciones seleccionadas del polinucleótido. Por ejemplo, en la misma secuencia referencial de polinucleótidos que comprende a diferencias definidas en 3 posiciones, en la primera posición seleccionada para mutaciones podría existir solamente una diferencia definida (por ejemplo, G cambiada a T), en la 2ª posición seleccionada podría existir un codón degenerado que codifica a 15 aminoácidos diferentes (por ejemplo, NNT), y en la 3ª posición seleccionada podrían existir 2 codones degenerados y un codón específico utilizado representando a 23 codones que codifican a 20 aminoácidos (por ejemplo, NNT, VVG y TGG).

60
 65

En algunas circunstancias, puede ser útil el generar una biblioteca de variantes de polinucleótidos que comprende a permutaciones diversas de diferencias múltiples de nucleótidos en un conjunto de posiciones seleccionadas. Asimismo, en algunas realizaciones de la divulgación, un método para sintetizar variantes de polinucleótidos puede ejecutarse donde el paso de PCR de la ronda 1 utiliza una mezcla de cebadores mutagénicos, que comprende a una diferencia definida en una posición seleccionada, y cebadores no mutagénicos, que no contienen diferencias en la posición seleccionada (es decir, la misma secuencia en la posición seleccionada que el polinucleótido referencial). Al utilizar una mezcla de cebadores mutagénicos y no mutagénicos que tienen una tasa definida en la amplificación PCR de la ronda 1, los amplicones resultantes comprenden a una mezcla de secuencias con o sin la diferencia definida en la posición objetivo. Al ensamblar y replicar esta mezcla de amplicones en una reacción SOE-PCR agrupada de la ronda 2, una biblioteca de variantes de polinucleótidos es generada la cual tiene

una mezcla estocástica de las diferencias definidas de nucleótidos en el conjunto de posiciones seleccionadas. Por ejemplo, con una diferencia definida (por ejemplo, una sustitución específica de codones) diseñada en cada una de las 13 posiciones objetivo y una tasa de 1:2 de cebadores mutagénicos en relación a los no mutagénicos, el grupo resultante de variantes de polinucleótidos comprende a una mezcla estocástica de polinucleótidos que tienen a
5 desde de cero a 13 diferencias definidas. Esta realización de los métodos aquí presentados suministra, por lo tanto, una forma rápida y fácil para generar bibliotecas combinatorias de variantes de polinucleótidos con combinaciones de diferencias definidas en las posiciones objetivo.

En una realización de este método para generar bibliotecas combinatorias, las pluralidades de mezclas de
10 cebadores son amplificadas por separado en una pluralidad de reacciones PCR de la ronda 1 para generar a una pluralidad de grupos de amplicones que tienen regiones superpuestas (por ejemplo, N+1 grupos de amplicones para N posiciones objetivo cuando una diferencia definida es incluida por cebador). La pluralidad de grupos de amplicones generados de esta forma es combinada entonces en una sola reacción SOE-PCR de la ronda 2. El grupo resultante de variantes de polinucleótidos comprende a una mezcla estocástica de las diferencias definidas
15 (por ejemplo, cambios de nucleótidos que codifican a cambios de aminoácidos) en las posiciones objetivo.

Alternativamente, la pluralidad de mezclas de cebadores puede ser combinada y amplificada en una sola
20 reacción PCR de la ronda 1 para generar a un solo grupo de amplicones mezclados. No se necesitan más mezclas de los amplicones, y el ensamblaje y la replicación SOE-PCR de la ronda 2 de los amplicones es ejecutada para generar al grupo que comprende a la mezcla estocástica de variantes de polinucleótidos.

Tal como será aparente para personas con conocimiento normal en la industria, en algunas realizaciones,
25 un segmento superpuesto definido para una secuencia de polinucleótidos podría no tener ninguna mutación asociada. Adicionalmente, el mismo segmento podría, en una permutación de secuencias de aminoácidos, abarcar a una mutación especificada, pero en algunas permutaciones de secuencias podría no tener ninguna mutación asociada con el segmento. Por lo tanto, en algunas realizaciones, la biblioteca de amplicones puede contener a miembros que no tienen ninguna diferencia de secuencias de polinucleótidos en comparación a la secuencia referencial para un segmento específico. Éstos polinucleótidos de conexión, que no tienen cambios asociados en la
30 secuencia en comparación a la secuencia referencial, pueden ser utilizados como conectores para ensamblar al polinucleótido completo.

Con una selección apropiada de segmentos, la biblioteca de amplicones comprende a miembros que
35 pueden ser utilizados para ensamblar a por lo menos 2 o más permutaciones diferentes de secuencias de aminoácidos de las diferencias definidas de aminoácidos en relación a la secuencia referencial. Por ejemplo, una pluralidad de mutaciones definidas por las diferencias A y B de aminoácidos puede tener las siguientes permutaciones: solamente A, solamente B, o A y B. Por lo tanto, la biblioteca de amplicones tiene suficientes miembros para generar una permutación de secuencias de aminoácidos que tienen independientemente a una mutación A o una mutación B. En algunas realizaciones, la biblioteca de amplicones tiene suficientes miembros para
40 generar a todas las permutaciones de secuencias de aminoácidos de las diferencias definidas de residuos de aminoácidos en relación a la secuencia referencial. Por lo tanto, para el ejemplo dado, la biblioteca de amplicones tiene suficientes miembros para generar permutaciones de secuencias de aminoácidos que tienen independientemente una mutación A, una mutación B, o una mutación A + B. Puesto que el tamaño de los amplicones corresponderá al tamaño de los segmentos, los amplicones pueden ser de 2000 bases o menos, de 1500 bases o menos, de 1200 bases o menos, de 1000 bases o menos, de 900 bases o menos, de 800 bases o
45 menos, de 700 bases o menos, de 600 bases o menos, de 500 bases o menos, de 400 bases o menos, de 300 bases o menos, de 250 bases o menos, o de 200 bases o menos hasta alrededor de 100, o tan pocas como alrededor de 50 bases de longitud. Generalmente, la longitud de los amplicones es desde alrededor de 50 a alrededor de 1000 bases, desde alrededor de 200 a alrededor de 1000 bases, desde alrededor de 300 a alrededor de 700 bases, o desde alrededor de 400 a 600 bases, con alrededor de 500 bases o menos como una longitud útil
50 dada la eficiencia de las polimerasas utilizadas en las reacciones de amplificación. En algunas realizaciones, los amplicones son de alrededor de 400 bases o menos de longitud.

Generalmente, la reacción de amplificación puede utilizar cualquier enzima utilizada para reacciones de
55 extensión reguladas por polimerasas, tales como la polimerasa Taq, la polimerasa Pfu, la polimerasa Pwo, la polimerasa Tfl, la polimerasa rTth, la polimerasa Tli, la polimerasa Tma, y el fragmento Klenow. Las condiciones para amplificar a un segmento de polinucleótidos utilizando una reacción en cadena de polimerasas puede seguir las condiciones estándar conocidas en la industria. Refiérase a, por ejemplo, Sambrook et al., 2001, Molecular Cloning: A Laboratory Manual (Clonación molecular: Un manual de laboratorio), 3ª Ed., Cold Spring Harbor Laboratory Press, NY y Ausubel et al., 1989, Current Protocols in Molecular Biology (Protocolos actuales en biología molecular),
60 Greene Publishing Associates y Wiley Interscience, N.Y (actualizaciones hasta 2008).

En algunas realizaciones, la amplificación de cada amplicón puede ejecutarse en reacciones separadas,
65 minimizando, por lo tanto, la necesidad de aislar a un producto de amplicones de otro amplicón. Sin embargo, las reacciones de amplificación para 2 o más amplicones pueden ejecutarse en una sola reacción y los productos pueden ser aislados, tal como un medio de electroforesis y cromatografía. En algunas realizaciones, los productos

de la reacción de amplificación pueden ser tratados con varias combinaciones de exonucleasas y fosfatasa para remover cebadores remanentes y nucleótidos libres (por ejemplo, una combinación de exonucleasas I y fosfatasa alcalinas).

5 Para generar al polinucleótido que codifica al polipéptido con la permutación definida de secuencias de aminoácidos, un conjunto de amplicones que tienen regiones complementarias superpuestas es seleccionado y ensamblado bajo condiciones que permiten el recosido de las regiones complementarias superpuestas entre sí. Por ejemplo, los amplicones pueden ser desnaturalizados y se les puede permitir entonces recoserse para formar a un complejo de amplicones que juntos codifican al polipéptido con una permutación definida de secuencias de aminoácidos que tiene una o más de las diferencias de residuos de aminoácidos en relación a una secuencia referencial. Generalmente, el ensamblaje de cada conjunto de amplicones puede ejecutarse por separado para que el polinucleótido que codifica a una permutación de secuencias de aminoácidos sea distinguida fácilmente de otro polinucleótido que codifica a una permutación diferente de secuencias de aminoácidos. En algunas realizaciones, el ensamblaje puede ser ejecutado en ubicaciones a las que se les puede asignar direcciones referenciales en un sustrato (por ejemplo, una formación) para que una pluralidad de polinucleótidos que codifican a una pluralidad de permutaciones definidas de secuencias de aminoácidos pueda ser generada simultáneamente.

En algunas realizaciones, los ensamblajes pueden ser preparados para que varios (es decir, 2 o más) amplicones sean representados para el mismo fragmento. El producto resultante de esta reacción de ensamblaje contendrá a una mezcla de polinucleótidos que contienen a diferentes permutaciones de las diferencias definidas de secuencias de aminoácidos. Esta mezcla puede ser donada directamente y las variantes pueden ser secuenciadas antes o después de que los polipéptidos codificados sean expuestos a ensayos.

El amplicón ensamblado es replicado utilizando una polimerasa para sintetizar al polinucleótido que codifica al polipéptido de interés. En algunas realizaciones, las condiciones de reacción pueden utilizar las mismas condiciones y las polimerasas utilizadas para la reacción de amplificación. Los amplicones ensamblados actúan como cebadores para que una sola ronda de replicación cree a un duplicado de los amplicones ensamblados. Generalmente, en el paso de replicación, los cebadores que se recosen a las secuencias enlazadoras de cebadores que flanquean al polinucleótido (es decir, la región terminal 5' y la región terminal 3') pueden agregarse para amplificar al producto de polinucleótidos al ejecutar reacciones adicionales de amplificación. En algunas realizaciones, estos cebadores que flanquean puede incorporar a secuencias de reconocimiento para enzimas de restricción para facilitar la clonación del producto sintetizado de polinucleótidos en plásmidos o vectores, tales como vectores de expresión.

En algunas realizaciones, los cebadores flanqueadores pueden tener secuencias que permiten una expresión directa in vitro utilizando un sistema acoplado de transcripción-interpretación para la síntesis del producto proteínico sin la necesidad de la transformación en un organismo anfitrión. Por lo tanto, algunos cebadores flanqueadores pueden incorporar a secuencias de control para controlar la expresión de la región codificadora del polipéptido. Reacciones de amplificación que utilizan a aquellos cebadores flanqueadores pueden enlazar operativamente a las secuencias de control a la región codificadora del polipéptido de interés.

En algunas realizaciones, la pluralidad de diferencias de aminoácidos es por lo menos 2. En algunas realizaciones, la pluralidad de diferencias de aminoácidos es por lo menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, o más. Asimismo, el número de diferencias definidas de nucleótidos puede variar desde 2 a 45, o más. El número de permutaciones para "n" pluralidades de diferencias definidas de residuos de aminoácidos es dada por la fórmula $n!/(k!(n-k)!)$, donde n es el número de mutaciones que no son mutuamente exclusivas, y k es el número de diferencias de aminoácidos y n! denota el operador factorial. En algunas realizaciones, el tamaño de la biblioteca de amplicones, por ejemplo, para un mínimo de 2 diferencias de residuos de aminoácidos es un tamaño de biblioteca que contiene a por lo menos 3 amplicones diferentes. En algunas realizaciones, el tamaño de la biblioteca es por lo menos 5, 6, 7, 8, 9, 10, 11, o incluso más amplicones diferentes. Por ejemplo, para una pluralidad de variantes que comprenden a por lo menos 10 diferencias definidas, asumiendo que ninguna de las diferencias está ubicada en una proximidad de tal forma que más de una puede ser incluida por cebador, variantes ensamblados con 10 diferencias definidas usa hasta 11 amplicones por reacción de ensamblaje. Asumiendo que una pluralidad de mutaciones diferentes es deseada en cualquiera de las pluralidades de posiciones que tienen diferencias definidas, se pueden utilizar bibliotecas mucho más grandes de amplicones. Asimismo, en algunas realizaciones, la biblioteca de amplicones puede comprender a por lo menos 5, 10, 20, 30, 40, 50, 75, 100, o más amplicones diferentes.

Una vez que la biblioteca de amplicones ha sido sintetizada, cualquier polinucleótido que codifica a una permutación especificada de secuencias de aminoácidos que se basa en una pluralidad de diferencias de residuos de aminoácidos puede elaborarse utilizando la biblioteca de amplicones. En algunas realizaciones, el método para generar a un polinucleótido que codifica a un polipéptido que tiene a una secuencia de aminoácidos con una o más diferencias definidas de residuos de aminoácidos en comparación a la secuencia referencial de polipéptidos puede comprender los pasos de: (a) ensamblar un conjunto de amplicones que tengan regiones complementarias adyacentes superpuestas, donde el conjunto ensamblado de amplicones comprende a una secuencia de polinucleótidos que codifica a una secuencia de aminoácidos con una o más diferencias definidas de residuos de

aminoácidos en comparación a una secuencia referencial, donde los amplicones son seleccionados de una biblioteca de amplicones que tienen miembros que codifican a una pluralidad de diferencias de aminoácidos, y (b) replicar el conjunto de fragmentos ensamblados de polinucleótidos superpuestos para sintetizar al polinucleótido de interés.

5 En algunas realizaciones, la biblioteca de amplicones puede ser utilizada para generar a polinucleótidos que codifican a cualquier permutación de una pluralidad definida de diferencias establecidas de aminoácidos, donde el método comprende: (a) generar permutaciones de secuencias de aminoácidos que difieren de una secuencia referencial de aminoácidos basándose en una pluralidad de diferencias definidas de residuos de aminoácidos en
10 comparación con una secuencia referencial de aminoácidos, (b) seleccionar una permutación definida de secuencias de aminoácidos y determinar una secuencia correspondiente de polinucleótidos que se basa en una secuencia referencial, (c) seleccionar un conjunto de fragmentos superpuestos de polinucleótidos que codifica a las permutaciones definidas de secuencias de aminoácidos, donde por lo menos cada fragmento superpuesto de polinucleótidos que codifica a una diferencia de aminoácidos proviene de una pluralidad de fragmentos de polinucleótidos que codifica a diferencias distintas de residuos de aminoácidos, donde la pluralidad de fragmentos
15 tiene miembros suficientes para ensamblar a polinucleótidos que codifican a por lo menos 2 permutaciones diferentes de secuencias de aminoácidos, (d) ensamblar el conjunto de fragmentos de polinucleótidos que tienen regiones complementarias adyacentes superpuestas, y (e) replicar el conjunto de fragmentos ensamblados superpuestos para sintetizar el polinucleótido que codifica al polipéptido. Para cada permutación deseada de secuencias de aminoácidos, los pasos de (b) hasta (e) pueden ser repetidos.

Un proceso de ejemplo para generar a los amplicones para "n" número de variantes es mostrado en la figura 4. En la realización ilustrada, el proceso comprende: (a) importar una secuencia referencial y una lista de mutaciones asociadas con la secuencia, (b) crear una lista de permutaciones basándose en la lista de mutaciones,
25 (c) seleccionar una permutación definida de la secuencia de aminoácidos (es decir, la variante 1), (d) identificar fragmentos superpuestos de polinucleótidos de una biblioteca de amplicones (por ejemplo, tal como se preparó en la figura 5), (e) determinar el número de variantes y si el número de variantes es menor que el número total de variantes deseadas, realizar nuevamente los pasos (a) al (d).

30 Para la síntesis eficiente de las bibliotecas de amplicones, cebadores diseñados apropiadamente de oligonucleótidos son utilizados en una reacción de amplificación. En algunas realizaciones, el método para generar una biblioteca de fragmentos superpuestos de polinucleótidos puede comprender: (a) generar una pluralidad de permutaciones de secuencias de aminoácidos que difieren de una secuencia referencial de aminoácidos que se basa en una pluralidad de diferencias definidas de residuos de aminoácidos que proviene de una secuencia referencial de aminoácidos, y para cada permutación (i) determinar una secuencia de polinucleótidos que codifique a la secuencia de aminoácidos que se basa en una secuencia referencial de polinucleótidos; (ii) examinar una secuencia de polinucleótidos e identificar un cambio en la secuencia de polinucleótidos que codifica a una diferencia de residuos de aminoácidos, y opcionalmente determinar la proximidad de un próximo cambio en la secuencia de polinucleótidos que codifica a una siguiente diferencia de residuos de aminoácidos en la permutación de secuencias de aminoácidos; (iii) seleccionar un cebador directo de oligonucleótidos que tiene una secuencia que codifica a la diferencia de aminoácidos, y opcionalmente incluir el siguiente cambio en la secuencia de polinucleótidos en el mismo cebador directo si se encuentra cerca del cambio en secuencia de polinucleótidos; (iv) examinar una secuencia de polinucleótidos de la ubicación del cebador directo hasta que el próximo cambio en la secuencia de polinucleótidos sea identificado o hasta el final del polinucleótido, y seleccionar un cebador inverso de oligonucleótidos para amplificar un fragmento de polinucleótidos con el cebador hacia delante de oligonucleótidos, donde el cebador inverso tiene una secuencia que codifica opcionalmente al siguiente cambio en una diferencia de residuos de aminoácidos; (v) repetir los pasos (ii) al (iv) para cada cambio en la secuencia de polinucleótidos que codifica a una diferencia de residuos de aminoácidos hasta que todos los cambios en la secuencia de polinucleótidos estén presentes en los cebadores de oligonucleótidos y hasta cuando se alcanza el final de la secuencia de polinucleótidos; y (g) amplificar con cada conjunto de cebadores directos e inversos de oligonucleótidos para generar la biblioteca de amplicones superpuestos que tienen miembros que codifican a las diferencias de aminoácidos. En estas realizaciones, cuando la revisión de la secuencia de polinucleótidos encuentra el final del polinucleótido, los cebadores flanqueadores pueden ser utilizados en combinación con los cebadores internos para completar la generación de los amplicones.

55 Un proceso de ejemplo para seleccionar los cebadores directos e inversos es ilustrado en la figura 5. En la figura 5, el proceso para seleccionar a los cebadores de oligonucleótidos comprende: (a) seleccionar una variante (una permutación de secuencias de aminoácidos) y generar su secuencia correspondiente de polinucleótidos basándose en una secuencia referencial, (b) crear un cebador hacia delante de oligonucleótidos para un fragmento con una primera mutación, (c) examinar la secuencia de la primera mutación a la siguiente mutación o al final del gen y crear un cebador inverso de oligonucleótidos para la siguiente mutación, (d) y si la siguiente mutación es cercana a la primera mutación, colocar a la siguiente mutación en el mismo cebador hacia delante de oligonucleótidos, (e) repetir los pasos (b) al (d) hasta que los finales de las variantes n de polinucleótidos sean alcanzadas.

65

Tal como se mencionó anteriormente, en algunas realizaciones donde el polinucleótido ha sido separado en segmentos superpuestos definidos por un conjunto de cebadores directos e inversos, los cebadores directos e inversos podrían no tener mutaciones asociadas. Un contexto en el cual esto podría ocurrir es en casos en los cuales los segmentos de polinucleótidos deben ser restringidos en lo que se refiere al tamaño, por ejemplo, alrededor de menos de 1000 bases, debido a una necesidad de una síntesis eficiente de un amplicón, de tal forma que no todos los segmentos tengan cambios definidos en la secuencia de polinucleótidos. En algunas realizaciones, al preparar los oligonucleótidos que se basan en el método que se acaba de mencionar, la búsqueda de la secuencia puede limitarse a un tamaño específico "l", por ejemplo, por alrededor de 1200 bases en el paso (iv) para seleccionar un cebador inverso. En otras palabras, después de la identificación de un cebador directo que se basa en una diferencia de secuencias, una revisión es hecha en una dirección o en la otra dirección de la secuencia de polinucleótidos para determinar la distancia en nucleótidos hasta la siguiente mutación. Si la distancia excede el límite establecido, un segmento que no contiene a ninguna mutación puede ser creado para conectar a los 2 segmentos que contienen las 2 mutaciones distantes. El proceso de revisión puede ser reiterado en el punto de la siguiente mutación.

Tal como se mencionó anteriormente, los cebadores de oligonucleótidos, ya sean individualmente o en conjuntos (por ejemplo, oligonucleótidos directos e inversos) así como los amplicones correspondientes pueden ser ubicados en sustratos a los que se les puede asignar direcciones referenciales para fines de automatización y/o almacenamiento. Cebadores de oligonucleótidos en sustratos a los que se les pueda asignar direcciones referenciales, también descritos aquí como formaciones de cebadores, pueden ser accedidos robóticamente para sintetizar a cualquier biblioteca de amplicones para una pluralidad definida de diferencias de aminoácidos. Asimismo, los amplicones en los sustratos a los que se les pueda asignar direcciones referenciales, también descritos en este documento como formaciones de amplicones, pueden ser accedidos para generar a una secuencia de polinucleótidos que codifica a una permutación deseada de secuencias de aminoácidos basándose en la pluralidad definida de diferencias de residuos de aminoácidos. Un sustrato o soporte sólido para la formación puede componerse de polímeros orgánicos tales como poliestireno, polietileno, polipropileno, polifluoroetileno, polietilenooxi y poliacrilamida, así como sus co-polímeros e injertos. Un soporte sólido también puede ser inorgánico, tal como vidrio, sílice, video de poros controlados (CPG - controlled pore glass), sílice de fase inversa o metales, tales como el oro o el platino. La configuración de un sustrato puede ser en la forma de micropartículas, esferas, partículas, gránulos, un gel, una membrana o una superficie. Las superficies pueden ser planas, sustancialmente planas, o no planas. Los soportes sólidos pueden ser porosos o no porosos, y pueden tener características de hinchamiento o de no hinchamiento un soporte sólido puede ser configurado en la forma de un pozo, una depresión, o de otro contenedor, frasco, característica o ubicación. Una pluralidad de soportes puede ser configurada en un ensayo en varias ubicaciones, a la cual se le puede asignar direcciones referenciales para la entrega robótica de reactivos, o por medio de métodos y/o instrumentos de detección. En algunas realizaciones, el sustrato es una cámara de reacción. Contenedores de reacción que están comercialmente disponibles tienen por lo menos una cámara de reacción, pero pueden contener a 8, 24, 96 o 384 cámaras de reacción. Un ejemplo de una cama de reacción es una placa de 96 pozos de micro - titulación.

En algunas realizaciones, un sistema robótico y un sistema informático asociado capaz de tomar muestras de cebadores o de parejas de cebadores de las formaciones puede ser utilizado para entregarlas a una cámara de reacción. Los reactivos para amplificaciones mediadas por polimerasas también pueden ser entregados a cada conjunto de cebadores en la cámara de reacción seguido por la implementación de una rutina de amplificación (tal como en un termociclador automatizado). Esto permite la formación de un sustrato al que se pueden asignar direcciones referenciales que contiene a amplicones definidos basándose en segmentos superpuestos de una secuencia de polinucleótidos. El sistema robótico puede escoger el conjunto apropiado de amplicones basándose en la permutación deseada de la secuencia de aminoácidos, los cebadores flanqueadores para la amplificación del producto final de polinucleótidos, y la entrega de reactivos para el ensamblaje y la reacción de amplificación. Un sistema robótico de ejemplo es indicado en la figura 6. El sistema robótico en la figura 6 comprende a instrucciones para (a) seleccionar un segmento y un amplicón asociado para su amplificación, (b) identificar oligonucleótidos directos e inversos para el fragmento seleccionado (es decir, el amplicón), almacenar la información de datos de los oligonucleótidos en una lista de oligonucleótidos únicos (por ejemplo, una placa de micro - titulación de 96 pozos), y colocar a los oligonucleótidos en un primer sustrato al que se le puede asignar direcciones referenciales (c) almacenar la información de los datos en fragmentos sintetizados (por ejemplo, posición en una formación, secuencia, oligonucleótidos utilizados, etcétera) para listar a fragmentos únicos, y colocar al oligonucleótido en un 2º sustrato que se le puede asignar direcciones referenciales, (d) determinar el número de fragmentos seleccionados en contra del número total de fragmentos requeridos para ensamblaje, y repetir los pasos (a) al (d) hasta que todos los fragmentos hayan sido seleccionados, (e) colocar al gen ensamblado en un 3º sustrato al que se le puede asignar direcciones referenciales, y repetir los pasos (a) hasta el (d) hasta que todas las variantes deseadas hayan sido generadas.

En algunas realizaciones, esta divulgación también suministra una biblioteca de fragmentos de polinucleótidos (es decir, amplicones) para ensamblar una pluralidad de polinucleótidos que codifican a permutaciones diferentes de secuencias de aminoácidos. En algunas realizaciones, la pluralidad de polinucleótidos comprende: a fragmentos de polinucleótidos con regiones superpuestas adyacentes, donde cada fragmento de

5 polinucleótidos está enlazado por secuencias enlazadoras de cebadores para cebadores directos e inversos, donde la pluralidad de polinucleótidos tiene a miembros que codifican a las secuencias enlazadoras de cebadores de una diferencia específica de residuos de aminoácidos que proviene de una pluralidad definida de diferencias de residuos de aminoácidos en relación a una secuencia referencial de aminoácidos para que la pluralidad de fragmentos de polinucleótidos codifique a toda una pluralidad seleccionada de diferencias de residuos de aminoácidos a partir de la pluralidad definida de diferencias de residuos de aminoácidos; y donde la pluralidad de fragmentos de polinucleótidos comprende a miembros para ensamblar 2 o más permutaciones diferentes de secuencias de aminoácidos de las diferencias definidas de aminoácidos. En algunas realizaciones, la pluralidad de fragmentos de polinucleótidos comprende a miembros suficientes para ensamblar a todas las permutaciones posibles de secuencias de aminoácidos de la pluralidad seleccionada de diferencias de residuos de aminoácidos. En algunas realizaciones, los miembros de la pluralidad de amplicones son formados utilizando cebadores directos e inversos.

15 Tal como será aparente para una persona con conocimiento en la industria, los métodos aquí descritos pueden ser practicados utilizando técnicas estándar disponibles para personas con conocimiento en la industria, tal como aquel descrito en Sambrook et al., 2001, *Molecular Cloning: A Laboratory Manual* (Clonación molecular: Un manual de laboratorio), 3ª Ed., Cold Spring Harbor Laboratory Press, NY y Ausubel et al., 1989, *Current Protocols in Molecular Biology* (Protocolos actuales en biología molecular), Greene Publishing Associates y Wiley Interscience, N.Y (actualizaciones hasta 2008). Los oligonucleótidos pueden ser sintetizados utilizando metodologías químicas conocidas, tales como aquellas que se basan en metodologías de síntesis de fases sólidas de fosforamiditas (refiérase a, por ejemplo, Wright, et al., 1993, *Tetrahedron Letters* (Cartas de tetraedro) 34, 3373-3376; Caruthers, 1991, *Acc. Chem. Res.* 24, 278-284; y las referencias aquí citadas).

25 También se suministra en este documento a sistemas informáticos implementados en la forma de software informático para ejecutar los métodos ya descritos. En algunas realizaciones, el producto de programa informático comprende un medio de almacenamiento legible por máquinas que tiene instrucciones de un programa que comprende a códigos para cada uno de los pasos de: (a) importar una secuencia referencial y una lista de mutaciones asociadas con la secuencia, (b) crear una lista de permutaciones basándose en la lista de mutaciones, (c) seleccionar una permutación definida de la secuencia de aminoácidos, (d) identificar fragmentos superpuestos de polinucleótidos provenientes de una biblioteca de amplicones (por ejemplo, tal como se preparó en la figura 5), (e) determinar el número de variantes y si el número de variantes es menos que el número total de variantes deseadas, repetir los pasos (a) al (d).

35 En algunas realizaciones, el producto de programa informático contiene a un medio de almacenamiento legible por máquinas que tiene instrucciones de un programa que comprende a códigos para que cada uno de los pasos de: (a) seleccionar una variante (una permutación de secuencias de aminoácidos) y generar su secuencia correspondiente de polinucleótidos basándose en una secuencia referencial, (b) crear un cebador directo de oligonucleótidos para un fragmento con una primera mutación, (c) examinar la secuencia de la primera mutación a la siguiente mutación o al final del gen y crear un cebador inverso de oligonucleótidos para la siguiente mutación, (d) y si la siguiente mutación está cerca de la primera mutación, ubicar a la siguiente mutación en el mismo oligonucleótido hacia delante, (e) repetir los pasos desde (b) hasta (d) hasta los finales de la variante del polinucleótido n sean alcanzados.

45 Tal como se mostró en las ilustraciones de las figuras 4, 5 y 6, los programas implementados por computadora para la selección de amplicones, para la selección de cebadores de oligonucleótidos, y para el almacenamiento en formatos a los cuales se les puede asignar direcciones referenciales pueden ser integrados para permitir la automatización de varios pasos de los métodos de la divulgación.

50 Tal como se describe en este documento, en algunas realizaciones, los métodos pueden ser utilizados para sintetizar a polipéptidos que codifican a polinucleótidos que tienen un conjunto definido de mutaciones seleccionadas de la pluralidad de diferencias definidas en residuos de aminoácidos a partir de una secuencia referencial. Los métodos aquí descritos permiten una síntesis eficiente de varias permutaciones de secuencias de aminoácidos que se basan en las diferencias de residuos de aminoácidos. Una síntesis eficiente de polinucleótidos que codifican a varias permutaciones de secuencias de aminoácidos es útil para una variedad de aplicaciones de diseño proteínico. Refiérase a, por ejemplo, la publicación de la aplicación de Estados Unidos US20060195947; la publicación de la aplicación de Estados Unidos US20050153417; y la patente de Estados Unidos número 7,220,566. En algunas realizaciones, los métodos pueden ser utilizados para sintetizar a polinucleótidos que codifican a variantes de enzimas que tienen propiedades mejoradas que se basan en un conjunto de mutaciones conocidas por afectar a diferentes propiedades de la enzima. Por ejemplo, algunas mutaciones pueden afectar, entre otras cosas, la actividad enzimática, la estabilidad térmica, la especificidad de sustrato, la estereoselectividad, la estereoespecificidad, y la capacidad de refracción para producir inhibiciones. Aunque las técnicas tradicionales de mutagénesis aleatoria y las técnicas de evolución proteínica pueden conllevar a la identificación de mutaciones que afectan a estas varias propiedades enzimáticas, muchas de estas mutaciones podrían ocurrir independientemente de las otras. Utilizando los métodos aquí descritos pueden elaborarse varias permutaciones de mutaciones que afectan a diferentes rasgos, tales como la actividad enzimática, la especificidad de sustratos y la estabilidad térmica, y pueden examinarse para identificar a enzimas diseñadas que tienen varios rasgos alterados deseados.

Los métodos aquí indicados suministran una eficiencia y una precisión sorprendentes para generar bibliotecas grandes de variantes de polinucleótidos que comprenden a varias permutaciones de cambios de secuencias. Por ejemplo, la secuencia proteínica de la oxidasa de gulonolactona (GLO) (L-) de *Rattus norvegicus* (acceso: 92090602-sp-P10867.3-GGLO_RAT) puede ser interpretado de vuelta para suministrar una secuencia de ADN de 1.3 kb que puede ser utilizada como una plantilla para diseñar a 90 variantes de polinucleótidos, donde cada uno codifica a un polipéptido variante que tiene una combinación diferente de desde 3 a 5 sustituciones de aminoácidos. Por ejemplo, una lista de 90 permutaciones de desde 3 a 5 sustituciones de aminoácidos puede ser seleccionada de la siguiente lista de 10 posibles sustituciones: T28S, D95A, S156N, G175S, R212D, I251E, F302S, H330I, Y370G, y K423N. Las 90 mutaciones diferentes de sustituciones de aminoácidos codificadas por las variantes de polinucleótidos fueron las siguientes: D95A/F302S/H330I/K423N; D95A/F302S/Y370G; D95A/G175S/H330I; D95A/G175S/H330I/Y370G/K423N; D95A/G175S/R212D/F302S/Y370G; D95A/G175S/R212D/H330I; D95A/G175S/R212D/Y370G/K423N; D95A/I251E/F302S/K423N; D95A/I251E/H330I; D95A/I251E/K423N; D95A/I251E/Y370G; D95A/R212D/F302S; D95A/R212D/I251E/F302S; D95A/S156N/F302S/H330I/K423N; D95A/S156N/G175S; D95A/S156N/G175S/H330I/Y370G; D95A/S156N/G175S/I251E/F302S; D95A/S156N/I251E/H330I; D95A/S156N/I251E/K423N; D95A/S156N/K423N; D95A/S156N/R212D/I251E; F302S/H330I/K423N; G175S/F302S/Y370G/K423N; G175S/H330I/K423N; G175S/I251E/F302S; G175S/R212D/H330I; G175S/R212D/I251E/H330I; G175S/R212D/K423N; G175S/R212D/Y370G; G175S/R212D/Y370G/K423N; H330I/Y370G/K423N; I251E/H330I/Y370G; I251E/H330I/Y370G; I251E/Y370G/K423N; R212D/F302S/Y370G/K423N; R212D/H330I/K423N; R212D/I251E/F302S; R212D/I251E/F302S/H330I; R212D/I251E/Y370G; R212D/I251E/Y370G; S156N/F302S/H330I; S156N/F302S/K423N; S156N/F302S/Y370G; S156N/G175S/F302S/Y370G; S156N/G175S/I251E/F302S; S156N/G175S/K423N; S156N/G175S/K423N; S156N/G175S/R212D/F302S/H330I; S156N/I251E/F302S/H330I; S156N/I251E/H330I/Y370G; S156N/I251E/H330I/Y370G/K423N; S156N/I251E/Y370G; S156N/R212D/F302S/H330I/Y370G; S156N/R212D/K423N; T28S/D95A/G175S/F302S; T28S/D95A/G175S/F302S/Y370G; T28S/D95A/H330I; T28S/D95A/I251E; T28S/D95A/I251E/F302S/K423N; T28S/D95A/R212D; T28S/D95A/S156N/H330I/Y370G; T28S/D95A/S156N/R212D; T28S/D95A/S156N/R212D; T28S/D95A/S156N/R212D/Y370G; T28S/D95A/Y370G; T28S/D95A/Y370G/K423N; T28S/F302S/K423N; T28S/G175S/H330I; T28S/G175S/H330I/Y370G; T28S/G175S/I251E/F302S; T28S/G175S/I251E/F302S/Y370G; T28S/G175S/I251E/H330I; T28S/G175S/I251E/K423N; T28S/H330I/K423N; T28S/I251E/F302S/H330I/K423N; T28S/R212D/F302S/H330I; T28S/R212D/H330I; T28S/R212D/I251E/F302S; T28S/R212D/I251E/Y370G/K423N; T28S/R212D/Y370G; T28S/S156N/F302S/H330I/Y370G; T28S/S156N/F302S/Y370G; T28S/S156N/F302S/Y370G; T28S/S156N/G175S; T28S/S156N/G175S; T28S/S156N/G175S/I251E; T28S/S156N/G175S/I251E/K423N; T28S/S156N/R212D/I251E/H330I; T28S/S156N/R212D/I251E/K423N; y T28S/S156N/R212D/K423N

Se puede utilizar software (por ejemplo, tal como se describió en las figuras 4-6) para determinar un total de solamente 55 amplicones, que corresponden a fragmentos variantes de polinucleótidos con regiones superpuestas de secuencias. También puede utilizarse el software para ensamblar a 90 variantes de polinucleótidos en una reacción SOE-PCR de la ronda 2. Softwares también pueden ser utilizados para determinar que un total de solamente 22 cebadores de oligonucleótidos son necesarios en 55 reacciones PCR separadas de la ronda 1 con el polinucleótido referencial de 1.3 kb como plantilla para generar a los 55 amplicones necesarios. Los 22 cebadores de oligonucleótidos tienen una longitud de solamente 30 a 33 nucleótidos, e incluyen a cebadores mutagénicos que comprenden a cambios de nucleótidos en el medio de la secuencia (por ejemplo, en los nucleótidos 15-17).

Por lo tanto, de acuerdo a los métodos aquí descritos, la construcción de las 90 variantes diferentes de polinucleótidos requieren la síntesis de únicamente 22 oligonucleótidos relativamente cortos (30-mer a 33-mer), una primera reacción PCR de la ronda 1 para crear a los 55 amplicones (es decir, fragmentos variantes de polinucleótidos), y una 2ª reacción SOE-PCR de la ronda 2 en la cual los 55 amplicones son agrupados en varias combinaciones (con cebadores flanqueadores directos e inversos) para permitir un ensamblaje SOE-PCR de las 90 variantes de polinucleótidos. Al preparar las reacciones SOE-PCR de la ronda 2, cada uno de los 55 amplicones pueden ser re - utilizados en un promedio de 7.8 veces, con ciertos fragmentos siendo utilizados únicamente 1 o 2 veces, y otros tanto como 36 veces.

El flujo de trabajo de las reacciones de la ronda 1 y de la ronda 2 pueden ser controladas por listas de trabajo generadas por medio de software (por ejemplo, tal como las figuras 4 y 6) las cuales son utilizadas para ejecutar a los sistemas robóticos de Tecan para el manejo de líquidos. Las listas de trabajo para la construcción de esta biblioteca ilustrativa de 90 variantes sólo requirieron 110 operaciones de manejo de líquidos para las reacciones PCR de la ronda 1 para generar 55 amplicones utilizando a 22 cebadores, y solamente 430 operaciones de manejo de líquidos para las reacciones de ensamblaje SOE-PCR de la ronda 2 para preparar las 90 variantes de polinucleótidos de longitud completa a partir de los 55 amplicones.

La precisión de las secuencias variantes de polinucleótidos suministrada por los métodos aquí presentados puede determinarse por pasos adicionales de clonación y de secuenciación de cada una de las pluralidades de las estructuras provenientes de las reacciones de la ronda 2. Tal como se ilustra por los ejemplos (más adelante), los métodos aquí presentados resultan en un nivel sorpresivamente alto de secuencias correctas (secuencias perfectas de longitud completa (FLP - full-length perfect) - es decir, secuencias que tienen los cambios de nucleótidos

deseados en relación al polinucleótido de referencia.

Por lo menos algunas de las ventajas sorprendidas de los métodos aquí descritos se refieren a la mayor precisión de las bibliotecas grandes de variantes de polinucleótidos producidas. En algunas realizaciones, los métodos pueden ser utilizados para preparar una biblioteca a la cual puede asignarse direcciones referenciales de por lo menos 10 variantes diferentes de polinucleótidos, donde cada una comprende a por lo menos una diferencia definida de secuencias en relación a una secuencia referencial de polinucleótidos, donde por lo menos un promedio del 75% de las secuencias variantes de polinucleótidos son secuencias correctas (por ejemplo, las secuencias que comprenden a la secuencia referencial de longitud completa con las diferencias definidas de nucleótidos introducidas por los cebadores utilizados en el método). En algunas realizaciones, los métodos suministran una biblioteca a la cual se le pueden asignar direcciones referenciales de por lo menos 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200 o más variantes diferentes de polinucleótidos, donde cada una comprende a por lo menos una diferencia definida de secuencias en relación a una secuencia referencial de polinucleótidos, donde por lo menos el 75%, el 80%, el 85%, el 90%, el 95% o más de las secuencias variantes de polinucleótidos son correctas - por ejemplo, FLP por medio de análisis secuencial.

En ciertas realizaciones, los métodos aquí descritos comprenden a una pluralidad de reacciones PCR de la ronda 1 utilizando a una plantilla referencial de polinucleótidos y a una pluralidad de reacciones de ensamblaje de amplicones SOE-PCR de la ronda 2 puede ser utilizada para preparar una biblioteca a la que se le pueden asignar direcciones referenciales de 10 o más variantes de polinucleótidos de un polinucleótido referencial de por lo menos 500 bp, 750 bp, 1000 bp, 1250 bp, 1500 bp o más, donde cada variante comprende a desde alrededor de 1-30, 1-25, 1-20, 1-15, 3-30, 3-20, o 3-15 cambios de nucleótidos en relación al polinucleótido referencial, donde las reacciones PCR de la ronda 1 comprenden a alrededor de 6-300, 6-200, 6-100, 6-50, 6-40, 6-30, 6-25, 6-20, 6-15, o tan pocos como 6-10 cebadores diferentes de oligonucleótidos, y por lo menos el 75%, el 80%, el 85%, el 90%, el 95% o más de las secuencias variantes de polinucleótidos son perfectas y de longitudes completas.

En algunas realizaciones, varias permutaciones de fragmentos de polinucleótidos (por ejemplo, seleccionados de una biblioteca a la que se le puede asignar direcciones referenciales) pueden ensamblarse en una biblioteca de variantes de polinucleótidos a la cual puede asignarse direcciones referenciales donde cada variante de polinucleótidos codifica a un polipéptido variante diferente que tiene una diferencia definida de residuos de aminoácidos. Esta biblioteca de clones a la cual se pueden asignar direcciones referenciales puede ser transformada a células (por ejemplo, E. Coli), para su interpretación, y colocación automatizada en placas y selección de colonias (es decir, transformantes viables). Las secuenciaciones pueden ser ejecutadas entonces para confirmar la combinación de mutaciones en cada secuencia de polipéptidos variantes que han sido generados de esa forma. Ensayos (por ejemplo, por medio de revisiones de alto caudal) de los polipéptidos variantes para detectar rasgos alterados deseados pueden ejecutarse en todos los polipéptidos variantes, u opcionalmente solamente en aquellos polipéptidos variantes que a través de secuenciaciones se ha confirmado que tienen la combinación deseada de mutaciones.

Aunque en algunas realizaciones este invento suministra métodos que resultan en una biblioteca de variantes de polinucleótidos a la cual se le pueden asignar direcciones referenciales que tienen diferencias definidas en posiciones objetivo, en algunas realizaciones de los métodos es preferible combinar (o agrupar) a las variantes de polinucleótidos. El agrupar a las bibliotecas bien definidas y diversas de variantes de polinucleótidos suministrada por los métodos aquí descritos resulta en una ventaja contra-intuitiva y sorprendente para permitir una transformación más eficiente y una revisión de una diversidad más amplia del rango de secuencias de polinucleótidos y la identificación de variantes de interés (por ejemplo, "coincidencias" que tienen propiedades mejoradas), antes de ejecutar secuenciaciones intensas en lo que se refiere a recursos y que además consumen más tiempo. Por lo tanto, este método permite al investigador crear y examinar un espacio de diversidad secuencial grande y bien definido de la biblioteca de variantes de polinucleótidos, y enfocarse en secuenciar únicamente a las variantes de interés. Otra ventaja de agrupar antes de la transformación es que se permite una revisión eficiente de bibliotecas con una diversidad extremadamente grande, por ejemplo, se puede obtener un estudio de una biblioteca de variantes de polipéptidos que codifican a variantes de polinucleótidos que tienen un conjunto completo de 20 diferencias de residuos de aminoácidos en cada posición, y únicamente aquella diferencia definida individual en cada miembro de la biblioteca.

La ventaja inesperada de agrupar una biblioteca de variantes a la cual se le pueden asignar direcciones referenciales antes de la transformación se cree que surge en parte de la capacidad del método aquí descrito para sintetizar a variantes de polinucleótidos para producir las variantes deseadas con una precisión con un promedio del 85% o más (es decir, la producción porcentual de las variantes de polinucleótidos de longitud completa que tienen las diferencias definidas en las posiciones objetivo). Otros métodos para generar bibliotecas de polinucleótidos diversos, tales como PCR propenso a errores, resultan en una distribución rápidamente amplia de mutaciones (por ejemplo, 37% sin mutaciones, el 37% con una mutación deseada, y el 26% con 2 o más mutaciones). Además, los métodos de esta divulgación son capaces de acceder a todas las 19 sustituciones de aminoácidos en cada posición de un polipéptido codificado. Un análisis comparativo de mutagénesis aleatoria usando PCR propenso a errores muestra que solamente se accede a alrededor del 45% de estas mutaciones que

son accedidas por los métodos de esta divulgación. Debido a esta precisión mucho mayor, y a la accesibilidad a la diversidad (variantes precisas individuales y de varios lugares), las bibliotecas agrupadas de variantes de polinucleótidos hechas por los métodos de esta divulgación pueden examinarse más eficientemente. Por ejemplo, variantes de posiciones individuales que comprenden a todos los 20 aminoácidos en cada una de las 20 posiciones de un polinucleótido referencial de 1.2 kb que codifican a un polipéptido fueron sintetizadas y agrupadas, y la agrupación resultante de variantes clonadas, fueron transformadas y puestas en 19 placas (88 pozos de datos/placa). La revisión de aquellas 19 placas genera variantes de polinucleótidos que codifican a 361 polipéptidos únicos que tienen a una sola diferencia de aminoácidos en relación al polipéptido referencial. Este resultado representa el 95% de la cobertura de la revisión de la biblioteca agrupada en solamente 19 placas.

Asimismo, en algunas realizaciones de los métodos de esta divulgación, la pluralidad resultante de variantes de polinucleótidos que tienen a por lo menos una diferencia definida en relación a un polinucleótido referencial son combinadas. En algunas realizaciones, las variantes agrupadas son clonadas entonces en un sistema de expresión, creando, por lo tanto, a una biblioteca de clones agrupados. Esta biblioteca de clones agrupados puede ser transformada (por ejemplo, en un sólo paso de transformación) a células para su interpretación, colocación en placas y selección de colonias (es decir, transformantes viables). Una formación de colonias que proviene de esta biblioteca de clones agrupados puede ser ejecutada (por ejemplo, por medio de revisiones de alto caudal) antes de secuenciarlas para identificar polipéptidos que codifican a variantes de polinucleótidos que tienen rasgos alterados deseados. Una de aquellas "coincidencias" para un rasgo alterado es identificada, puede ser secuenciada para determinar la combinación específica de mutaciones presentes en la secuencia variante de polinucleótidos. Opcionalmente, aquellos polipéptidos que codifican a variantes que no tienen los rasgos alterados deseados buscados en la formación no necesitan ser secuenciados. Asimismo, el método de la biblioteca de clones agrupados puede suministrar más eficiencia al requerir únicamente una sola transformación en vez de un conjunto de reacciones de transformación en paralelo.

En una realización alternativa, el método puede ser ejecutado de tal forma que una biblioteca a la que se le puede asignar direcciones referenciales de variantes de polinucleótidos que tienen a por lo menos una diferencia definida en relación a un polinucleótido referencial es clonada por separado en un sistema de expresión - por ejemplo, usando una reacción mega - cebadora (refiérase a, por ejemplo, Tyagi et al., BMC Biotechnology (Biotecnología) 2004, 4:2; doi:10.1186/1472-67504-2). Esta biblioteca de clones a la que se le puede asignar direcciones referenciales (por ejemplo, vectores de expresión que contienen cada uno a una variante de polinucleótidos de la biblioteca a la que se le puede asignar direcciones referenciales) puede entonces ser agrupada y transformada en un sólo paso en células para su interpretación, colocación en placas, selección de colonias, y revisión tal como se describió anteriormente.

Asimismo, el método también puede ser utilizado para generar a varias permutaciones de combinaciones de mutaciones para examinar las características estructurales de proteínas biológicamente importantes. Por ejemplo, los receptores involucrados en la transducción de señales de moléculas extracelulares actúan por medio de la interacción con otros receptores, así como varias proteínas intracelulares. Este complejo de interacciones puede afectar diferentes procesos de señalización celular del mismo tipo que la molécula receptora. La señalización negativa y positiva puede ser iniciada por el mismo receptor. Un ejemplo específico son los receptores proteínicos G- acoplados, que interactúan con las proteínas $\beta\gamma$, G_s α y $G_i\alpha$. Refiérase a, por ejemplo, Morris et al., 1999, *Physiol. Rev.* 79:1373 1430. Debido a que mutaciones en diferentes dominios del receptor pueden tener diferentes efectos, los métodos aquí descritos suministran un proceso eficiente para generar permutaciones diferentes de combinaciones de mutaciones que son conocidas por afectar diferentes aspectos de la función receptora, permitiendo, por lo tanto, estudios acerca de las funciones biológicas estructurales y asociadas de la proteína de interés.

Mientras que los métodos para generar a permutaciones diferentes de una secuencia de polinucleótidos han sido ilustrados para generar a polinucleótidos que codifican a varias permutaciones de un polipéptido a partir de un conjunto definido de diferencias de residuos de aminoácidos, se debe entender que los métodos pueden ser adaptados en general para generar a permutaciones de secuencias de polinucleótidos. Por ejemplo, los métodos aquí descritos pueden ser utilizados para generar a permutaciones diferentes de polinucleótidos funcionales, tales como genes para ARNs de ribosomas. Varios ARNs forman a complejos de nucleoproteínas que participan en la síntesis proteínica en procariontes y eucariontes. Muchos antibióticos funcionan al interrumpir la función de los ribosomas y son conocidos por interactuar con regiones definidas de ciertos ARNs. Varias mutaciones han sido identificadas que afectan a la síntesis proteínica y estas regiones son correlacionadas con lugares de interacción con antibióticos. Refiérase a, por ejemplo, Yassin et al., 2005, *Proc Natl Acad Sci. Estados Unidos de América* 102(46):16620-16625.

Utilizando los métodos aquí descritos, varias permutaciones de mutaciones conocidas que afectan a la función ribosómica del ARN pueden ser sintetizadas y los efectos de ciertas combinaciones de mutaciones pueden ser examinadas. Otras aplicaciones serán aparentes para aquellas personas con conocimiento en la industria.

6. EJEMPLOS

Ejemplo 1: Preparación de Amplicones

Preparación de oligonucleótidos. Los cebadores de oligonucleótidos con una concentración de 200 μM son diluidos a 4 μM con agua estéril en una placa Axygen HalfDeep 96 (1.1 mililitros). Para la mayoría de posiciones en la placa de micro - titulación, el agregar 10 μl de oligonucleótidos a 490 μl de dH_2O será suficiente. Para las posiciones A01 y D01 en la placa de oligonucleótidos, se podría requerir un volumen grande de cebadores comunes directos e inversos. Los volúmenes máximos de aspiración en el informe resultante acerca de la sección de aspiración y de los volúmenes de repartición es verificada antes del siguiente paso. Los cebadores pueden ser agrupados en tasas equimolares y no equimolares después de la dilución y antes del uso para la formación de amplicones (refiérase a, por ejemplo, los ejemplo 6-8).

Ronda 1 - formación de amplicones por medio de PCR. Los sistemas robóticos de Tecan son utilizados en porciones de 5 μl de cada cebador de oligonucleótidos directo e inverso en placas BioRad HardShellPCR96 (salida de comandos de Tecan) y se agregan 40 μl de una mezcla maestra. Se realiza el PCR de la ronda 1 y se verifica la amplificación utilizando a un 2% de un e-gel de 96 pozos. El reactivo para el PCR es de la siguiente forma: 5 microlitros de 10x de amortiguador de Herculesa, 1 μl de 40 mM de dNTPs, 1 μl de 100 ng/microlitro de plantilla SOE.; 5 unidades de polimerasa de Herculesa (Stratagene, La Jolla, CA, Estados Unidos de América). El PCR es ejecutado de la siguiente forma: 2 minutos de desnaturalización a 95 °C seguido por una ciclación a 95 °C durante 30 segundos, 56 °C durante 30 segundos, 72 °C durante un minuto/Kb. El número de ciclos es 17.

Tratamiento con ExoSAP-it. Para la ronda 1 del PGR, se transfieren 25 μl del producto de reacción a una placa fresca de 96 pozos, y se agregan 2 μl de ExoSAP-it (USB Corp., Ohio, USA) junto con 0.5 microlitros de DpnI y se realiza una ciclación (transferencia manual a 37 °C durante una hora y a 80 °C durante 15 minutos). Las muestras son diluidas a un volumen final de 100 μl por medio de la adición de 73 μl de dH_2O y agrupadas a otra placa BioRad HardShellPCR96 utilizando un guion de comandos Tecan.

Ejemplo 2: Ensamblaje de amplicones y análisis de productos

Ronda 2 - ensamblaje y SOE-PCR. Sistemas robóticos Tecan son usados para dividir agrupaciones de 15 μl de fragmentos (es decir, los amplicones de la ronda 1) en placas BioRad HardShellPCR96 (resultado de los guiones de comandos de Tecan) y agregar 35 μl de la mezcla maestra (5 μl de 10x de amortiguador de Herculesa, 1 μl de 40 mM de dNTPs, 0.2 microlitros de cebadores hacia delante, 0.2 microlitros de cebadores inversos, 2.5 unidades de la enzima de Herculesa, y 28. 1 μl de dH_2O). Los amplicones pueden ser agrupados en montos equimolares o no equimolares (refiérase, por ejemplo, los ejemplos 6-8). El PCR es realizado y la amplificación es verificada utilizando un 2% de E-gel de 96 pozos. El PCR es ejecutado de la siguiente forma: 2 minutos de desnaturalización a 95 °C seguido por una ciclación a 95 °C durante 30 segundos, 56 °C durante 30 segundos, 72 °C durante un minuto/Kb. El número de ciclos es 17.

Purificación de placas de 96 pozos. Purificar todas las muestras utilizando placas de 96 pozos de limpieza PCR Zymo ZR-96 (protocolo de los fabricantes con las siguientes modificaciones (Zymo Research, CA, Estados Unidos de América). Se realizan todos los pasos de centrifugación a 2800 revoluciones por minuto durante 10 minutos. Para eluir el ADN, se aplican 25 microlitros de dH_2O a 55 °C directamente a la membrana sílice, se centrifuga durante 10 minutos y se repite. Al utilizar este método se recuperan 48-50 μl del producto. En este punto, los productos pueden ser agrupados en combinaciones deseadas antes de la digestión y de la clonación.

Digestión de la enzima BglI de restricción. La digestión BglI es para la clonación en el lugar BgLL de un vector de expresión. Se transfieren 30 μl de cada producto purificado o agrupación de productos en una placa PCR fresca semi-bordeada, se agregan 20 μl de la mezcla maestra de digestión de BglI (5 μl de 10X del amortiguador NEB 3, 13 μl de dH_2O , 20 U de BglI (New England Biolabs, MA, Estados Unidos de América)) a todas las muestras y se incuban a 37 °C durante 4 horas.

Purificación de placas de 96 pozos. Se purifican a todas las muestras utilizando placas de 96 pozos de limpieza PCR Zymo ZR-96 de acuerdo al protocolo del fabricante con las siguientes modificaciones: (1) realizar todos los pasos de centrifugación a 2800 revoluciones por minuto durante 10 minutos; (2) para eluir el ADN, aplicar 25 μl de dH_2O a 55 °C directamente a la membrana sílice, se centrifuga durante 10 minutos, y se repite. Utilizando este método, se recuperan 48-50 μl del producto. Se verifica la recuperación del producto utilizando un 2% de E-gel de 96 pozos.

Ligaciones para el vector de expresión. Se transfieren 3 μl de cada inserción purificada a una placa fresca y se agregan 27 μl de la mezcla maestra de ligación (3 μl de 10X de amortiguador de ligasas (New England Biolabs, MA, Estados Unidos de América), 1 μl del vector digerido por BglI (50 ng/microlitro), 400 unidades de ligasas T4 (New England Biolabs, MA, Estados Unidos de América), 22 μl de dH_2O) a las muestras. Se incuban durante 14 horas a 16 °C seguido por 15 minutos a 65 °C, seguido por 8 °C y se mantiene a esa temperatura.

Transformación HTP. Se transfieren 2 μl de cada reacción de ligación a 20 μl de células químicamente

competentes de TSS y se incubaba en hielo en un bloque metálico durante por lo menos 15 minutos. Se aplica un golpe de calor a 42 °C durante 35 segundos y se regresa al bloque metálico durante 2 minutos. Se agregan 80 µl de medio de SOC a 37 °C a cada muestra. Se incubaba a 37 °C durante una hora antes de colocarse en placas.

5 Colocación en placas. Se utiliza a Tecan para colocar en placas a 40 µl por pozo de la mezcla de transformación en bandejas Q divididas en 48 pozos. Se suministran 3 esferas de 5 mm a cada pozo utilizando un dispensador de micropartículas. Se cultiva a los transformantes durante la noche a 37 °C.

10 Selección y cultivo. Para el análisis de la formación y/o de las secuencias se generan placas maestras al escoger a colonias individuales de las bandejas Q para inocular los pozos de las placas de fondo plano de Nunc que contienen LB, CAM, y 1% de glucosa. En los casos de placas verificadas por secuencias, 2 colonias por pozo de la bandeja Q son escogidas en 2 placas separadas de fondo plano de Nunc. En los casos de placas de variantes verificadas no secuenciales, 3 colonias por cada bandeja Q son empacadas en 3 placas separadas de fondos planos de Nunc.

15 PCR de colonias. El PCR de colonias es ejecutado en por lo menos una placa maestra al agregar 2 µl de aquel cultivo a la mezcla estándar maestra de PCR de colonias y se realiza el PCR. Se utiliza ExoSAP-it para limpiar al producto PCR de la siguiente forma: (1) se transfieren 5 µl de la muestra de PCR a una nueva placa de PCR que contiene 2 µl de ExoSAP-it; (2) se incubaba a 37 °C durante 15 minutos y a 80 °C durante 15 minutos; y (3) se diluyen las muestras a un volumen final de 40 µl al agregarse 33 µl de dH₂O.

20 Secuenciación de los productos PCR. Se agrega 4 µl de 1 mM de cebador de secuenciación a la placa de secuenciación. Se agregan 4 µl de la muestra limpiada de PCR para crear una mezcla plantilla/cebador que entonces es utilizada para secuenciaciones estándar en ciclos.

25 Ejemplo 3: Generación de un conjunto de 190 variantes diferentes de polinucleótidos donde cada una codifica a un polipéptido que tiene un solo cambio de aminoácidos en relación a un polipéptido referencial.

30 Diseño experimental: Se seleccionó a un polinucleótido referencial de 1359 bp (que codifica a una enzima de 453 aminoácidos). Un total de 190 diferencias de residuos de aminoácidos en comparación con la secuencia referencial fueron seleccionadas basándose en los cambios secuenciales vistos en enzimas homólogas. Las 190 variantes fueron hechas como polinucleótidos individuales que codifican a 190 proteínas diferentes que serían expresadas y probadas. Cada una de las 190 variantes de polinucleótidos fueron ensambladas al combinar a 2 amplicones que comprenden al cambio deseado en un solo codón en su región superpuesta (tal como se preparó en la ronda 1 más adelante) en una reacción SOE (ronda 2 más adelante).

35 Preparación de oligonucleótidos: Un total de 382 cebadores de oligonucleótidos para PCR fueron diseñados y sintetizados de acuerdo a métodos estándar. Los oligonucleótidos fueron generalmente de 31 nucleótidos (nt) de longitud con los cambios deseados al codón de interés ubicado en el medio (aproximadamente en la base 15) de cebador del oligonucleótido. Todos los oligonucleótidos fueron diluidos a 4 µM con agua estéril en una placa Axygen HalfDeep 96 (1.1 mL).

40 Ronda 1 - Formación de amplicones por medio de PCR: Cada amplicón, correspondiente a un fragmento variante de polinucleótidos, fue generado en una reacción PCR utilizando un vector que comprende a un polinucleótido referencial de 1359 nt usado como una plantilla y utilizando a un cebador mutagénico en combinación con un cebador flanqueador común (se recose al vector corriente arriba o corriente abajo a partir del gen que no contiene ninguna mutación). El flujo de trabajo, las condiciones y limpieza de la reacción son las mismas que las descritas en el ejemplo 1.

45 Ronda 2 - Ensamblaje de amplicones. Los amplicones purificados de la ronda 1 fueron agrupados para que 2 amplicones con regiones superpuestas que comprenden al cambio deseado de secuencias de codones para cada una de las 190 variantes de polinucleótidos y cada grupo fue dividido en una placa de 96 pozos tal como se describió en los ejemplos 1 y 2. Cebadores comunes flanqueadores directos e inversos fueron agregados a cada grupo y se ejecutó un PCR tal como se describió en el ejemplo 2 resultando en el ensamblaje de los amplicones (es decir, fragmentos de polinucleótidos) para formar a las variantes de polinucleótidos de longitud completa. Las acciones de ensamblaje fueron inspeccionadas por medio de gel de agarosa y se encontró que contenían al producto del tamaño esperado (refiérase, por ejemplo, a la figura 3).

50 Análisis secuencial de las variantes de polinucleótidos. Después de la purificación utilizando la limpieza de PCR Zymo ZR-96, los productos fueron clonados a un vector de expresión, y cada ligación fue transformada en células anfitrionas de E. Coli. 2 colonias de cada transformación fueron escogidas y el ADN de plásmidos fue preparado para secuenciaciones de ADN. La muestra de cada transformación fue secuenciada utilizando cebadores de secuenciación que eran internos y flanqueaban al gen. De las 190 variantes de polinucleótidos, 160 (el 84%) demostraron tener solamente la secuencia perfecta de longitud completa (FLP – full length perfect) con los cambios deseados de secuencias de codones. Al secuenciar a la 2ª preparación de plásmidos de las 30 variantes que fueron

65

incorrectas, 25 secuencias adicionales correctas fueron identificadas. Esto provocó una tasa general de éxito del 97% de secuencias correctas (185 de los 190 nucleótidos deseados fueron identificados). Los polinucleótidos fueron expresados y los polipéptidos variantes fueron expuestos a ensayos.

- 5 Ejemplo 4: Generación de un conjunto de 96 variantes diferentes de polinucleótidos donde cada una codifica a un polipéptido que tiene 3 cambios de aminoácidos en relación a un polipéptido referencial.

10 Diseño experimental: Un polinucleótido referencial de 1359 nt fue seleccionado. 96 variantes fueron diseñadas, donde cada una contenía 3 mutaciones en relación a la secuencia referencial. Las 96 variantes fueron hechas como polinucleótidos individuales donde cada una codificaba a una de 96 proteínas diferentes que serían expresadas y probadas. Cada una de las 96 variantes de polinucleótidos fue ensamblada al combinar a 4 amplicones que comprenden a los cambios deseados de codones en sus regiones de superposición (tal como se preparó en la ronda 1 más adelante) en una reacción SOE (ronda 2 más adelante).

15 Preparación de oligonucleótidos: Un total de 130 cebadores de oligonucleótidos fueron diseñados y sintetizados de acuerdo a métodos estándar. Los oligonucleótidos fueron generalmente de una longitud de 31 nt con los cambios deseados en el codón de interés en la parte media (aproximadamente en la base 15) del oligonucleótido. Todos los oligonucleótidos fueron diluidos a 4 μM con agua estéril en una placa Axygen HalfDeep 96 (1.1 mL).

20 Ronda 1 - Formación de amplicones por medio de PCR: Cada amplicón, correspondiente a un fragmento variante de polinucleótidos, fue generado en una reacción PCR utilizando a un vector comprendido de un polipéptido referencial de 1359 nucleótidos usado como una plantilla y utilizando a un cebador mutagénico en combinación con otro cebador mutagénico o un cebador flanqueador común (que se recose al vector corriente arriba o corriente abajo desde el gen que no contiene mutaciones). El flujo de trabajo, las condiciones y limpieza de la reacción fueron las mismas que se describieron en el ejemplo 1.

25 Ronda 2 - Ensamblaje de amplicones. Los amplicones purificados de la ronda uno fueron agrupados para que 4 amplicones con regiones superpuestas que contenían a los cambios deseados de las secuencias de codones para cada una de las 96 variantes de polinucleótidos y cada agrupación fue dividida en la placa de 96 pozos tal como se describió en los ejemplos 1 y 2. Cebadores comunes flanqueadores directos e inversos fueron agregados a cada grupo y el PCR fue ejecutado tal como se describió en el ejemplo 2 resultando en un ensamblaje de los amplicones (es decir, fragmentos de polinucleótidos) para formar a la variante de polinucleótidos de longitud completa. Las reacciones de ensamblaje fueron inspeccionadas en un gel de agarosa y se encontró que contenían al producto del tamaño esperado. (Refiérase a, por ejemplo, la figura 3).

35 Análisis secuencial del producto de polinucleótidos. Después de una purificación utilizando la limpieza PCR Zymo ZR-96, los productos fueron clonados a un vector de expresión, y cada ligación fue transformada en células anfitrionas de E. Coli. 2 colonias de cada transformación fueron escogidas y el ADN de plásmidos fue preparado para secuenciaciones de ADN. Una muestra de cada transformación fue secuenciada utilizando cebadores de secuenciación internos y flanqueadores en relación al gen. De las 96 variantes de polinucleótidos, 82 (el 85%) se determinó que tenían sólo a la secuencia FLP correcta con los cambios deseados.

- 40 Ejemplo 5: Generación de un conjunto de 96 variantes diferentes de polinucleótidos donde cada una codifica a un polipéptido que tiene desde 1 a 6 cambios en relación a un polipéptido referencial.

45 Diseño experimental: Un polinucleótido referencial de 1056 nucleótidos fue seleccionado. 96 variantes fueron diseñadas, donde cada una contenía a desde una (1) a seis (6) mutaciones en comparación con la secuencia referencial. Las 96 variantes fueron hechas como polinucleótidos individuales donde cada una codificaba a una de las 96 proteínas diferentes que iban a ser expresadas y probadas. Cada una de las 96 variantes de polinucleótidos fue ensamblada al combinar a desde 2 (por ejemplo, para variantes que codifican a un solo cambio de aminoácidos) a 7 (por ejemplo, para variantes y codifican a 6 cambios de aminoácidos) amplicones que comprenden a los cambios deseados de codones y su región de superposición (tal como se preparó en la ronda 1 más adelante) en una reacción SOE (ronda 2 más adelante).

50 Preparación de oligonucleótidos: Un total de 108 cebadores de oligonucleótidos fueron diseñados y sintetizados de acuerdo a métodos estándar. Los oligonucleótidos fueron generalmente de 31 nucleótidos de longitud con los cambios deseados al codón de interés en la parte media (aproximadamente la base 15) del oligonucleótido. Los 2 cambios de aminoácidos fueron juntados de cerca, un oligonucleótido más largo fue diseñado para codificar a ambos cambios que van a ser incorporados. Todos los oligonucleótidos fueron diluidos a 4 μM con agua estéril en una placa Axygen HalfDeep 96 (1.1 mL).

55 Ronda 1 - Formación de amplicones por medio de PCR: Cada amplicón, correspondiente a un fragmento de variantes de polinucleótidos, fue generado en una reacción de PCR utilizando un vector que comprendía a un polinucleótido referencial de 1056 nucleótidos como una plantilla y usando a un cebador mutagénico en combinación

con otro cebador mutagénico o un cebador flanqueador común (recosiéndose al vector corriente arriba o corriente abajo a partir del gen que no contenía mutaciones). El flujo de trabajo, las condiciones y la limpieza de la reacción fueron tales como se describieron en el ejemplo 1.

5 Ronda 2 - Ensamblaje de amplicones. Los amplicones agrupados purificados (desde 2 a 7 amplicones/nucleótidos) fueron divididos en placas tal como se describió en el ejemplo 1. Cebadores comunes flanqueadores directos e inversos fueron agregados y el PCR fue ejecutado tal como se describió en el ejemplo 1.

10 Análisis secuencial del producto de polinucleótidos. Después de una purificación utilizando a la limpieza PCR de Zymo ZR-96, los productos fueron clonados a un vector de expresión, y cada legación fue transformada en células anfitrionas de E. Coli. 2 colonias de cada transformación fueron escogidas y se preparó ADN de plásmidos para su secuenciación de ADN. Una muestra de cada transformación fue secuenciada utilizando cebadores secuenciadores internos y flanqueadores en relación al gen. Tal como se mostró en la tabla 1 (más adelante), de las 96 variantes de polinucleótidos, 72 (el 75%) demostró tener solamente a la secuencia FLP correcta con los cambios deseados de 2-7 codones.

Tabla 1: 96 variantes construidas - 88 variantes requeridas para elaborar a la placa final

Conjunto de secuenciación	1 (n = 96)	2 (n = 96)	3 (n = 16)
Correctas totales	72	84	92
Secuencias fallidas	5	3	2
Contaminación cruzada	5	2	0
Codificación de mutaciones	8	2	0
Indeles	6	4	2

20
25 Ejemplo 6: Generación de un conjunto agrupado de 82 variantes específicas de polinucleótidos, donde cada una codifica a un polipéptido que contiene a una mutación de aminoácidos en comparación con un polinucleótido referencial

30 Diseño experimental: Un total de 82 residuos específicos en una enzima que tiene 413 aminoácidos fueron utilizados como objetivos para mutaciones. La enzima de tipo silvestre es codificada por una secuencia referencial de polinucleótidos de 1242 bp. 82 variantes específicas de polinucleótidos que contienen cada una a una de 82 mutaciones son preparadas por separado por medio del diseño de la síntesis de cebadores, de la generación de amplicones, y el ensamblaje de parejas de amplicones para sintetizar a un conjunto de variantes de polinucleótidos de longitud completa que comprenden a cada una de las mutaciones. Todas las variantes de polinucleótidos de longitud completa son agrupadas entre sí antes de la purificación, de la clonación y de la transformación. El agrupamiento previo a la transformación reduce la carga de trabajo (en comparación a hacer todos los pasos de clonación y de transformación por separado para cada variante genética). Después de la selección y de la expresión de las colonias, las variantes enzimáticas son examinadas para detectar variantes genéticas interesantes o secuencias para identificar la mutación de interés.

35
40 Preparación de cebadores de oligonucleótidos: Para cada una de las 82 mutaciones deseadas, 2 cebadores mutagénico de oligonucleótidos (uno directo y uno inverso) fueron diseñados e incluían al cambio deseado de secuencias de codones de polinucleótidos en relación a la secuencia referencial. Los cebadores fueron de 33 nucleótidos de longitud con una posición objetivo ubicada en el medio del oligonucleótido. Todos los oligonucleótidos fueron sintetizados de acuerdo a métodos estándar y diluidos a 4 µM con agua estéril en una placa de Axygen HalfDeep 96 (1.1 mililitros).

45
50 Ronda 1 - Formación de Amplicones por medio de PCR: Para cada una de las 82 posiciones, 2 amplicones fueron generados con regiones superpuestas comprendiendo al cambio deseado de codones en aquella posición. Cada amplicón fue generado en una reacción de PCR utilizando a un vector que contiene al polinucleótido referencial de 1142 nucleótidos como una plantilla y un cebador mutagénico en combinación con un cebador flanqueador común (recosiéndose al vector corriente arriba o corriente abajo desde el gen). El flujo de trabajo, las condiciones y la limpieza de la reacción fueron tal como se describieron en el ejemplo 1.

55
60 Ronda 2 - Ensamblaje de amplicones. Para cada una de las 82 variantes de polinucleótidos, los 2 amplicones purificados con regiones superpuestas que comprenden al cambio deseado de codones (de la reacción de la ronda uno) fueron combinados. Cebadores comunes flanqueadores directos e inversos fueron agregados a cada uno de los 82 pozos y se ejecutó el SOE-PCR tal como se describió en el ejemplo 2 resultando en un ensamblaje de los amplicones (es decir, fragmentos de polinucleótidos) para formar a las 82 variantes de polinucleótidos de longitud completa. Se ejecutó una inspección de las reacciones de ensamblaje en un gel de agarosa y todas las 82 reacciones demostraron contener al producto del tamaño esperado. Estos productos de longitud completa para todas las 82 reacciones fueron agrupados entre sí.

65 Clonaciones y transformaciones agrupadas de las variantes de polinucleótidos de longitud completa. Los

productos agrupados provenientes de las reacciones de ensamblaje fueron purificados utilizando un botiquín Qiagen, digeridos con enzimas de restricción, clonados en un vector de expresión, transformados en células anfitrionas, y puestos en placas en medios LB sólidos que contenían glucosa y cloroanfenicol.

5 Revisiones adicionales de los polipéptidos codificados por los polinucleótidos para identificar “coincidencias” (por ejemplo, polipéptidos que exhiben alguna propiedad o actividad mejorada en comparación con el tipo silvestre) son ejecutadas utilizando transformantes en ensayos estándar de alto caudal.

10 Análisis secuencial de variantes de polinucleótidos. 8 variantes fueron seleccionadas aleatoriamente para secuenciaciones. Cada variante demostró contener una mutación deseada. Ninguna mutación aleatoria fue observada.

Ejemplo 7: Generación de un conjunto agrupado de variantes de polinucleótidos que codifican a 912 polipéptidos, donde cada uno tiene un solo cambio de aminoácidos en relación a un polipéptido referencial.

15 Diseño experimental: Se seleccionó a un polinucleótido referencial de 1383 bp (que codifica a una enzima de 460 aminoácidos). Residuos de aminoácidos en 48 posiciones diferentes de la enzima fueron seleccionados para una mutagénesis de saturación completa (es decir, la generación de todos los 19 cambios de residuos de aminoácidos en cada una de las 48 posiciones). Se utilizó un conjunto reducido de 23 codones, que comprenden a los conjuntos de codones reducidos de TGG, NNT y VWG, lo cual es capaz de codificar a todos los posibles cambios de aminoácidos. Para cada posición, todos los 23 polinucleótidos posibles que comprenden a 23 codones diferentes en aquella posición fueron ensamblados simultáneamente al combinar en una reacción SOE-PCR (reacción de la ronda 2 más adelante) a 2 conjuntos de amplicones (preparados en la reacción de la ronda 1), donde cada uno de los conjuntos de amplicones comprende a los 23 codones diferentes en la posición objetivo en su región superpuesta. La pluralidad resultante de variantes de polinucleótidos que están construidas incluyen a todos los 23 codones diferentes en cada una de las 48 posiciones diferentes. Estas variantes son combinadas entonces en un grupo antes de ligarse y transformarse para generar a clones. La secuenciación es ejecutada entonces en clones aleatorios para confirmar que los genes variantes de longitud completa fueron preparados y/o en “coincidencias” identificadas por medio de ensayos subsiguientes de revisión que fueron ejecutados en los clones.

20 Preparación de cebadores de oligonucleótidos: Para cada posición seleccionada de aminoácidos, 6 cebadores de oligonucleótidos fueron diseñados. Los oligonucleótidos fueron de 33 nucleótidos de longitud con el codón que va ser alterado ubicado en el medio de la secuencia (por ejemplo, en la posición 17). 3 oligonucleótidos fueron diseñados en la dirección ‘directa’, mientras que los otros 3 fueron complementos en la dirección ‘inversa’. 4 de los oligonucleótidos fueron degenerados y 2 fueron específicos. 2 de los cebadores fueron oligonucleótidos específicos incluyendo al codón “TGG” (o el complemento inverso “CCA”) que codifica al triptófano en el centro. Los otros 4 cebadores fueron conjuntos de oligonucleótidos degenerados incluyendo ya sea a un codón de cualquiera de los conjuntos de codones reducidos “VWG” o “NNT” (o sus complementos inversos “CWB” o “ANN”) en sus centros y aparte de eso fueron idénticos al polinucleótido referencial en todas las otras posiciones.

30 Todos los cebadores de oligonucleótidos fueron sintetizados de acuerdo a métodos estándar y diluidos a 4 µM con agua estéril en una placa Axygen HalfDeep 96 (1.1 mililitros). Los 3 cebadores ‘directos’ (que contenían a los codones “NNT”, “VWG” y “TGG”) fueron combinados en una tasa molar de 16:6:1, respectivamente. Los oligonucleótidos ‘inversos’ que contenían a los codones “ANN”, “CWB” y “CCA” fueron agrupados similarmente.

35 Ronda 1 - Formación de amplicones por medio de PCR: 2 agrupaciones de amplicones (es decir, conjuntos de productos PCR que comprendían a todos los 23 codones) fueron generadas para cada una de las 48 posiciones que fueron usadas como objetivo para mutagénesis de saturación. Cada grupo de amplicones fue generado en una reacción de PCR utilizando a un vector que contenía al polinucleótido referencial de 1383 nucleótidos como plantilla y los cebadores mutagénicos se agruparon (descritos anteriormente) en combinación con un cebador como un flaqueador (recosiéndose al vector corriente arriba o corriente abajo del gen). De otra forma el flujo de trabajo de la reacción, las condiciones y la limpieza y purificación de los amplicones fue tal como se describió en el ejemplo 1.

40 Ronda 2 - Ensamblaje de amplicones. Los grupos purificados de amplicones de la ronda 1 fueron combinados para que las parejas de amplicones con regiones superpuestas que comprenden a los 23 cambios deseados de codones para cada una de las 48 posiciones sean divididas en una placa de 96 pozos tal como se describió en los ejemplos 1 y 2. Cebadores comunes flaqueadores directos e inversos fueron agregados a cada pozo y se ejecutó el SOE-PCR tal como se describió en el ejemplo 2 resultando en un ensamblaje de los amplicones superpuestos (es decir, fragmentos de polinucleótidos) para formar agrupaciones de variantes de polinucleótidos de longitud completa (es decir, conjuntos de polinucleótidos que tienen 23 codones diferentes en una de las 48 posiciones). La inspección de estas reacciones de ensamblaje de SOE-PCR fue ejecutada en un gel de agarosa y se encontró que contenían productos de longitud completa del tamaño esperado.

45 Clonación y transformación agrupada de las variantes de polinucleótidos de longitud completa. Las 48 agrupaciones de variantes de polinucleótidos de longitud completa provenientes de las 48 reacciones de ensamblaje

fueron agrupadas entre sí. Este grupo final debería incluir a polinucleótidos de longitud completa donde cada uno tiene solamente a una de 48 posiciones y uno de 23 codones del conjunto reducido de codones. Por lo tanto, un total de 1056 variantes diferentes de polinucleótidos (22 cambios de codones X 48 posiciones) que codifican a 912 polipéptidos diferentes (19 cambios de aminoácidos X 48 posiciones) deben estar presentes.

5 Los polinucleótidos agrupados de longitud completa fueron purificados utilizando a un botiquín de Qiagen y luego fueron clonados a un vector de expresión, transformados en células anfitrionas, y puestos en placas en un medio LB sólido que contenía glucosa y cloroanfenicol.

10 Revisiones adicionales de los polipéptidos codificados por los oligonucleótidos para identificar a "coincidencias" (por ejemplo, polipéptidos que muestran alguna propiedad o actividad mejorada en comparación al tipo silvestre) son ejecutadas utilizando transformantes en ensayos estándar de alto caudal.

15 Análisis secuencial de variantes de polinucleótidos. 24 colonias fueron seleccionadas aleatoriamente para su análisis secuencial. Plásmidos preparados de cada transformante fueron seleccionados utilizando cebadores de secuenciación internos y flanqueadores en relación al gen. De las 4 variantes de polinucleótidos, 19 (el 80%) mostraron tener el resultado deseado de un gen de longitud completa que contiene a un cambio de aminoácidos en una posición objetivo y no existían mutaciones adicionales. 2 genes parecieron tener 2 cambios de aminoácidos que fueron atribuidos a transformaciones contaminadas (transformantes dobles). Un gen demostró tener 2 cambios de aminoácidos, uno de los cuales surgió de una mutación aleatoria en una posición no objetivo. 2 no demostraron tener ningún cambio de aminoácidos, pero contenían una mutación silenciosa. Estas variantes silenciosas son esperadas cuando se usan conjuntos de codones degenerados.

20 Ejemplo 8: Generación de un conjunto agrupado de variantes de polinucleótidos que comprenden a mutaciones definidas de mezclas estocásticas que codifican a desde 0 a 13 diferencias residuales de aminoácidos en un polipéptido referencial.

25 Diseño experimental: Un polinucleótido referencial de 1632 bp (que codifica a una enzima de 543 aminoácidos) fue seleccionado. Trece (13) cambios definidos de aminoácidos en la secuencia referencial fueron seleccionados para su preparación individual en varias combinaciones. Al mezclar a cebadores mutagénicos que codifican a mutaciones definidas con cebadores no mutagénicos, una mezcla de amplicones es preparada que comprende a las 13 mutaciones definidas individuales o en combinación, o sin mutaciones en absoluto. Todos los amplicones fueron combinados entonces en una sola reacción de ensamblaje para generar a un grupo de variantes de polinucleótidos de longitud completa (genes codificadores) que contiene a mezclas estocásticas de las mutaciones definidas. Esta mezcla de genes de longitud completa (la biblioteca genética) es ligada entonces al vector de expresión y transformada en el organismo anfitrión.

30 Preparación de cebadores de oligonucleótidos: Para 11 de las mutaciones definidas, una pareja de cebadores mutagénicos (uno directo y uno inverso) fue diseñada la cual incluía a las diferencias de nucleótidos que codifican al cambio deseado de aminoácidos. Adicionalmente, 2 cebadores no mutagénicos fueron diseñados para cada posición mutada. Los cebadores de oligonucleótidos fueron de 33 nucleótidos de largo y la posición de las diferencias de los nucleótidos (por ejemplo, el codón cambiado) estuvo ubicada en el medio del oligonucleótido. 2 de las diferencias definidas de residuos de aminoácidos en el polipéptido referencial fueron ubicadas muy cerca entre sí en las posiciones 173 y 176. Para estas 2 posiciones, 4 cebadores directos y 4 cebadores inversos fueron diseñados de la siguiente forma: un cebador que no tenía ninguna de las 2 mutaciones; 2 cebadores donde cada uno tenía una de las 2 mutaciones, y un cebador que tenía ambas mutaciones. Todos los cebadores fueron sintetizados de acuerdo a métodos estándar y diluidos a 4 µM con agua estéril en una placa Axygen HalfDeep 96 (una. 1 ml). Los 2 cebadores directos y los 2 cebadores inversos (uno que contenía la mutación, y otro no mutagénico) para cada posición, o los 4 cebadores directos y los 4 cebadores inversos (en el caso de las posiciones 173/176) fueron combinados en una taza molar de 1:2 (mutagénico: no mutagénico).

35 Ronda 1 - Formación de amplicones por medio de PCR: Un total de 13 mezclas de amplicones fue generada en una mezcla de reacción de PCR que contenía a un vector que tenía el polinucleótido referencial de 1632 nucleótidos como plantilla y la tasa molar de 1:2 de cebadores mutagénicos en relación a los no mutagénicos para las posiciones de interés. Los amplicones del extremo 5' y del extremo 3' fueron preparados utilizando cebadores mezclados para un extremo y un cebador común flanqueador (que se recose al vector corriente arriba o corriente abajo) para el otro. Todos los otros amplicones usaron cebadores mezclados para ambos extremos. La mayoría de amplicones contenían, por lo tanto, ninguna mutación, una mutación (por ejemplo, en el extremo 5' o en el extremo 3'), o 2 mutaciones. El flujo de trabajo, las condiciones y la limpieza de la reacción fueron tal como se describió en el ejemplo 1.

40 Ronda 2 - Ensamblaje de amplicones. Las 13 mezclas de amplicones purificados de la ronda 1 fueron agrupadas y se agregaron cebadores flanqueadores directos e inversos. Se ejecutó una SOE-PCR tal como se describe en el ejemplo 2 resultando en un ensamblaje de amplicones (es decir, fragmentos del oligonucleótido) en una muestra agrupada de variantes de polinucleótidos de longitud completa. Análisis de la reacción de ensamblaje

en un gel de agarosa indicó que contenía a productos de longitud completa del tamaño esperado.

Clonación y transformación agrupada de las variantes de polinucleótidos de longitud completa. La reacción de ensamblaje agrupada fue purificada utilizando un botiquín Qiagen, clonada en un vector de expresión, transformada en células anfitrionas, y colocada en placas en un medio LB sólido que contenía glucosa y cloroanfenicol.

Revisiones adicionales de los polipéptidos codificados por los polinucleótidos para identificar “coincidencias” (por ejemplo, polipéptidos que muestran alguna propiedad o actividad mejorada sobre el tipo silvestre) son ejecutadas utilizando los transformantes en ensayos estándar de alto caudal

Análisis secuencial de variantes de polinucleótidos. Una selección aleatoria de 31 clones fue secuenciada y se encontró que comprendía a la siguiente distribución de variantes de polinucleótidos de longitud completa que tenían una o más de las mutaciones definidas (o de tipo silvestre): 2 sin ninguna mutación; 4 con una de las mutaciones; 8 con 2 de las mutaciones; 8 con 3 de las mutaciones, 5 con 4 de las mutaciones; 3 con 5 de las mutaciones; una con 8 de las mutaciones. Todas las 13 mutaciones deseadas fueron observadas en por lo menos una de las variantes genéticas secuenciadas. Todas las variantes que contenían varias mutaciones deseadas tenían combinaciones diferentes de las mutaciones.

5
10
15
20
25
30
35
40
45
50
55
60
65

Reivindicaciones

- 5 **1.** Un método para sintetizar una pluralidad de variantes de polinucleótidos que tienen una mezcla estocástica de diferencias de nucleótidos definidas en relación a una secuencia de polinucleótido de referencia, en donde el polinucleótido de referencia codifica un polipéptido de referencia y cada una de la pluralidad de variantes del polinucleótido codifica un polipéptido que tiene por lo menos una diferencia de secuencia de aminoácidos en relación al polipéptido de referencia, el método comprendiendo:
- 10 (a) proporcionar una pluralidad de parejas de cebadores directos e inversos, en donde cada una de la pluralidad de cebadores directos e inversos comprende una mezcla de cebadores mutagénicos y no mutagénicos, cada cebador mutagénico comprendiendo una diferencia definida en una posición seleccionada y cada cebador no mutagénico no comprendiendo diferencia en la posición seleccionada, y en donde cada una de la pluralidad de parejas de cebadores directos e inversos comprende una mezcla que tiene una proporción definida de cebadores mutagénicos a no mutagénicos, y en donde cada pareja genera un amplicón que comprende una secuencia capaz de unirse a una secuencia superpuesta adyacente de por lo menos un otro amplicón;
- 15 (b) amplificar una plantilla de polinucleótido de referencia con cada una de la pluralidad de parejas de cebadores directos e inversos, generando de este modo una pluralidad de conjuntos de amplicones, en donde los amplicones resultantes comprenden una mezcla de secuencias con o sin las diferencias definidas en las posiciones seleccionadas, y en donde cada conjunto comprende amplicones que tienen secuencias superpuestas adyacentes capaces de unirse para formar la longitud completa de la secuencia de polinucleótidos de referencia;
- 20 (c) ensamblar y replicar la pluralidad de conjuntos de los amplicones, sintetizando de este modo una pluralidad de variantes de polinucleótido que tienen una mezcla estocástica de diferencias de nucleótidos definidas.
- 25 **2.** El método de la reivindicación 1, en donde la pluralidad de parejas de cebadores directos e inversos se combina antes de amplificar la plantilla de polinucleótidos de referencia, generando de este modo una pluralidad combinada de conjuntos de amplicones.
- 30 **3.** El método de la reivindicación 1, en donde la amplificación de la plantilla de polinucleótidos de referencia se lleva a cabo por separado con cada una de la pluralidad de las parejas de cebadores directos e inversos, y la pluralidad de conjuntos de amplicones se combina antes de ensamblar y replicar.
- 35 **4.** El método de cualquiera de las reivindicaciones 1-3, en el donde los cebadores se agrupan en proporciones equimolares antes de su uso para la formación de amplicones.
- 40 **5.** El método de cualquiera de las reivindicaciones 1-3, en donde los cebadores se agrupan en proporciones no equimolares antes de su uso para la formación de amplicones.
- 45 **6.** El método de cualquiera de las reivindicaciones 1-3, en donde la mezcla de cebadores mutagénicos y no mutagénicos comprende una proporción de mutagénicos a no mutagénicos de 1: 2.
- 50 **7.** El método de cualquiera de las reivindicaciones 1-6, en donde la pluralidad de variantes de polinucleótidos comprende por lo menos 10, 25, 35, 50, 75, 90, 120, 150, 180, 300, 500, 700, 900 o incluso más, variantes de polinucleótidos diferentes.
- 55 **8.** El método de cualquiera de las reivindicaciones 1-6, en donde:
- (i) por lo menos una de la pluralidad de variantes de polinucleótidos comprende por lo menos 2, 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, o incluso más, diferencias de nucleótidos definidas con respecto a la secuencia de polinucleótidos de referencia ; o
- (ii) dos o más, o cada una, de la pluralidad de variantes de polinucleótidos comprende por lo menos 1, 2, 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, o incluso más, diferencias de nucleótidos definidas con respecto a la secuencia de polinucleótidos de referencia.
- 60 **9.** El método de cualquiera de las reivindicaciones 1-6, en donde la pluralidad de variantes de polinucleótidos sintetizadas usando los métodos comprende por lo menos 10, 20, 30, 40, 50, 100, 200 o más, variantes de polinucleótidos diferentes, en donde cada variante comprende una diferencia de nucleótidos definida diferente en una de 10, 20, 30, 40, 50, 100, 200 o más, posiciones seleccionadas diferentes.
- 65 **10.** El método de cualquiera de las reivindicaciones 1-6, en donde:

(i) por lo menos una de la pluralidad de conjuntos de amplicones comprende por lo menos 3, por lo menos 5, por lo menos 7, por lo menos 10, o más amplicones diferentes; o

(ii) dos o más, o cada una, de la pluralidad de conjuntos de amplicones comprende por lo menos 3, por lo menos 5, por lo menos 7, por lo menos 10, o más amplicones diferentes.

5 **11.** El método de cualquiera de las reivindicaciones 1-6, en donde la longitud de la secuencia de polinucleótidos de referencia es de por lo menos 500 bp, 750 bp, 1000 bp, 1250 bp, o 1500 bp.

10 **12.** El método de un cualquiera de las reivindicaciones 1-6, en donde:

(i) la pluralidad de parejas de cebadores directos e inversos comprende 400 o menos, 300 o menos, 200 o menos, 100 o menos, 50 o menos, o 25 o menos oligonucleótidos diferentes; o

15 (ii) la pluralidad de parejas de cebadores directos e inversos comprende de 6 a aproximadamente 200, de 6 a aproximadamente 150, de 6 a aproximadamente 100, de 6 a aproximadamente 50, de 6 a aproximadamente de 40, de 6 a aproximadamente de 30, de 6 a aproximadamente 25, 6 a aproximadamente 20, 6 a aproximadamente 15, o incluso menos oligonucleótidos diferentes, y en donde las longitudes de los oligonucleótidos son de aproximadamente 20 a aproximadamente 50 nucleótidos, aproximadamente 20 a aproximadamente 40 nucleótidos, o aproximadamente 25 a aproximadamente 35 nucleótidos.

20 **13.** El método de cualquiera de las reivindicaciones 1-6, en donde la pluralidad de diferencias de nucleótidos definidas comprende por lo menos 5 posiciones seleccionadas del polinucleótido de referencia, y la pluralidad de variantes de polinucleótidos comprende por lo menos 15 variantes de polinucleótidos cada una teniendo una permutación diferente de las diferencias de nucleótidos definidas en las posiciones definidas.

25 **14.** El método de cualquiera de las reivindicaciones 1-6, en donde la pluralidad de diferencias de nucleótidos definidas comprende por lo menos 10 posiciones seleccionadas del polinucleótido de referencia, y la pluralidad de variantes de polinucleótidos comprende por lo menos 25 variantes de polinucleótidos cada una teniendo una permutación diferente de las diferencias de nucleótidos definidas en las posiciones definidas.

30

35

40

45

50

55

60

65

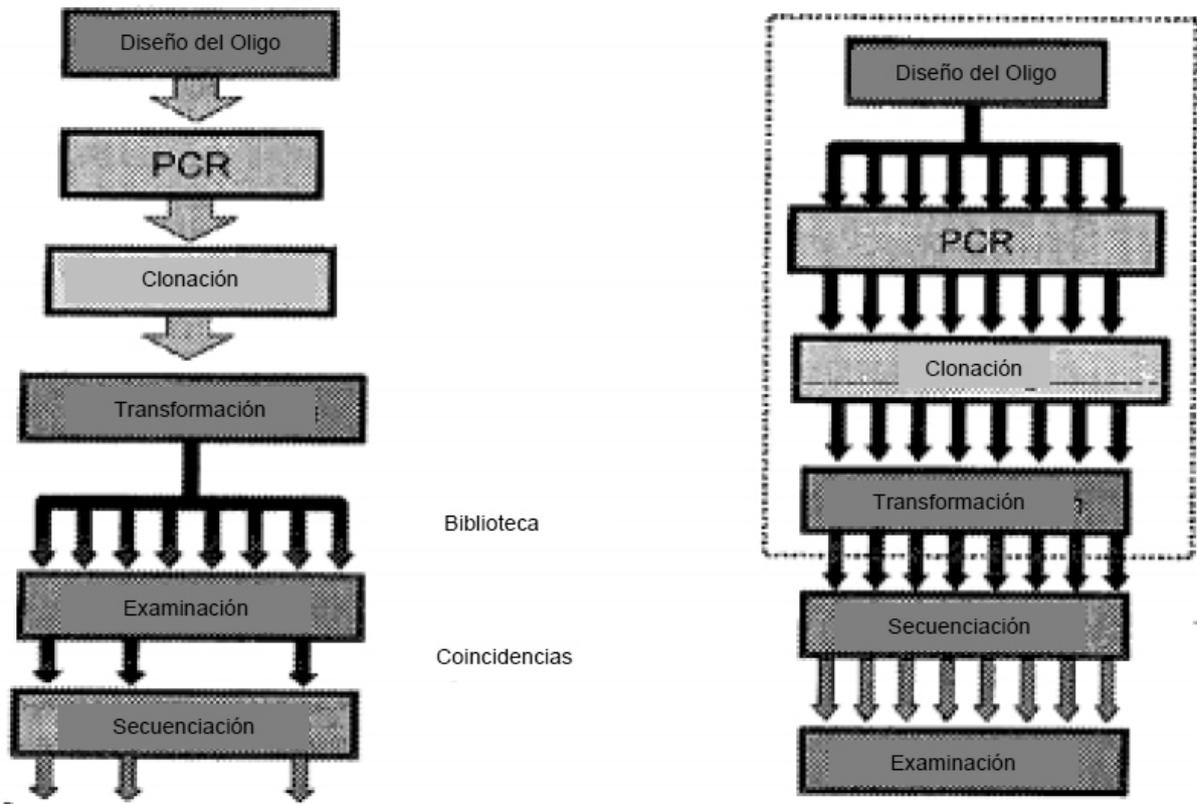


FIG. 1

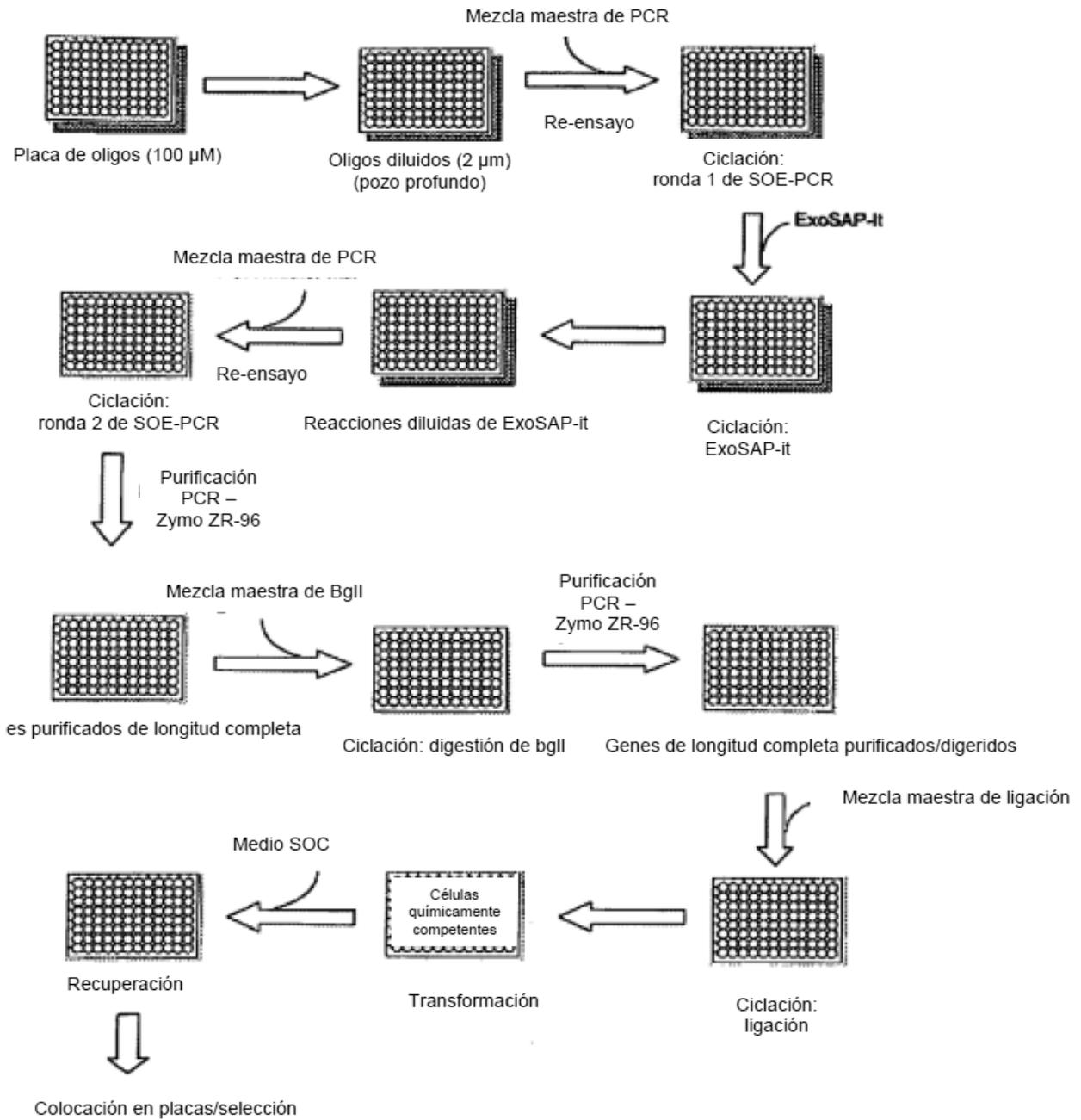


FIG. 2

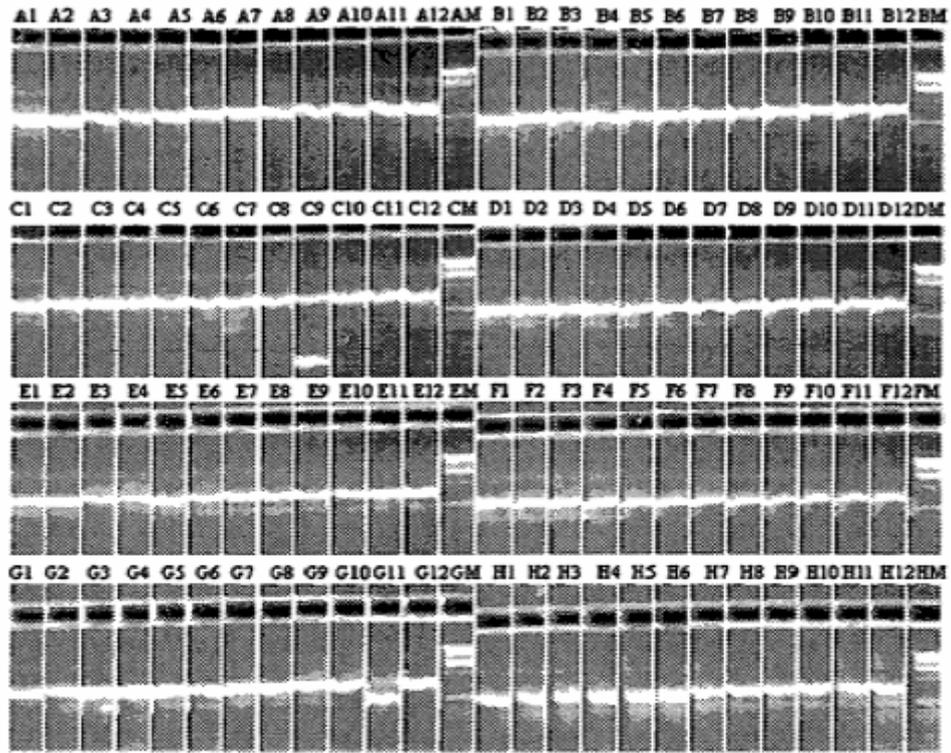


FIG. 3

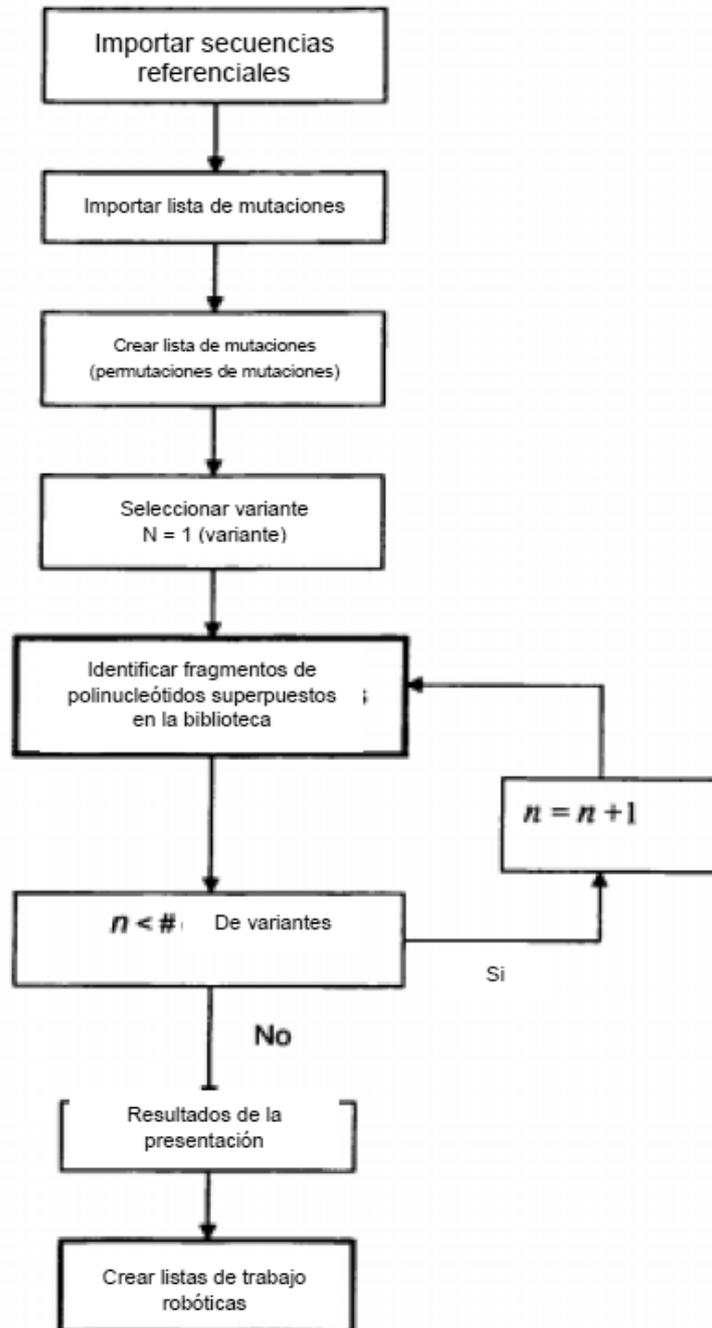


FIG. 4

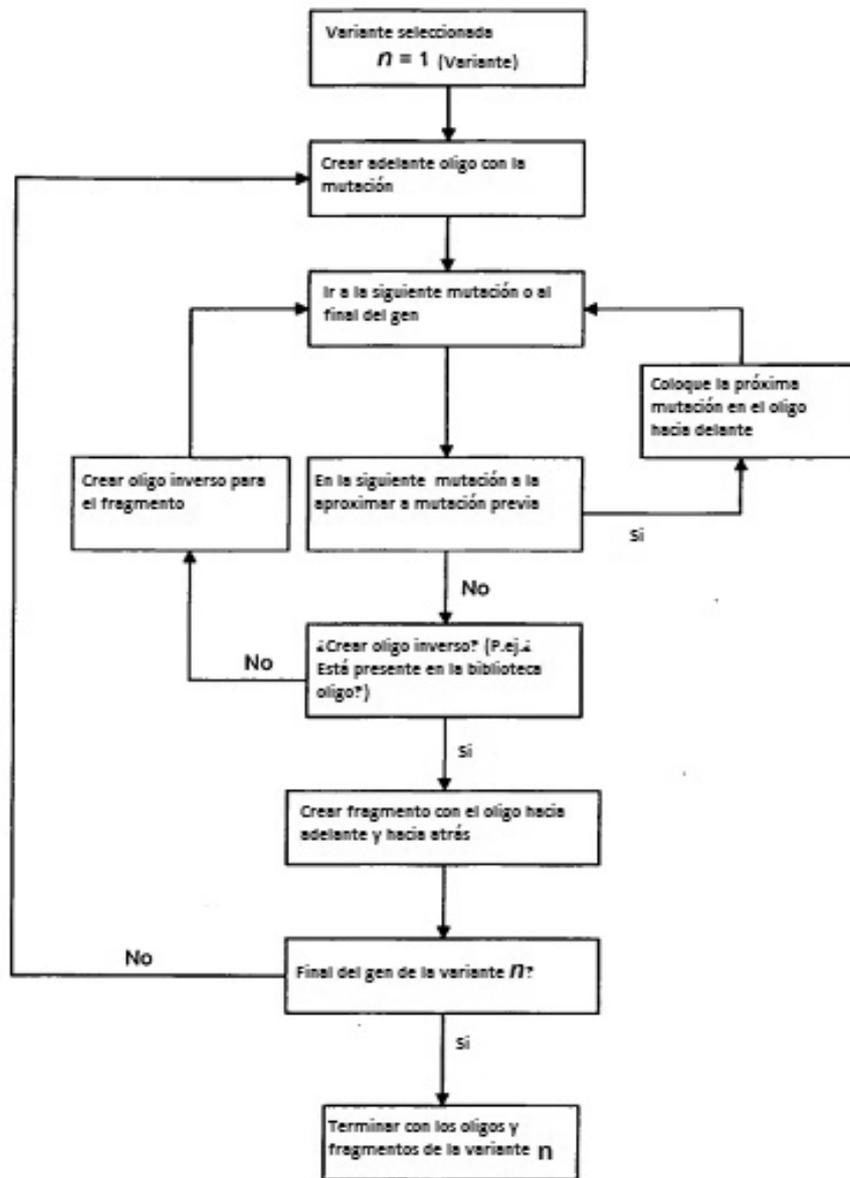


FIG. 5

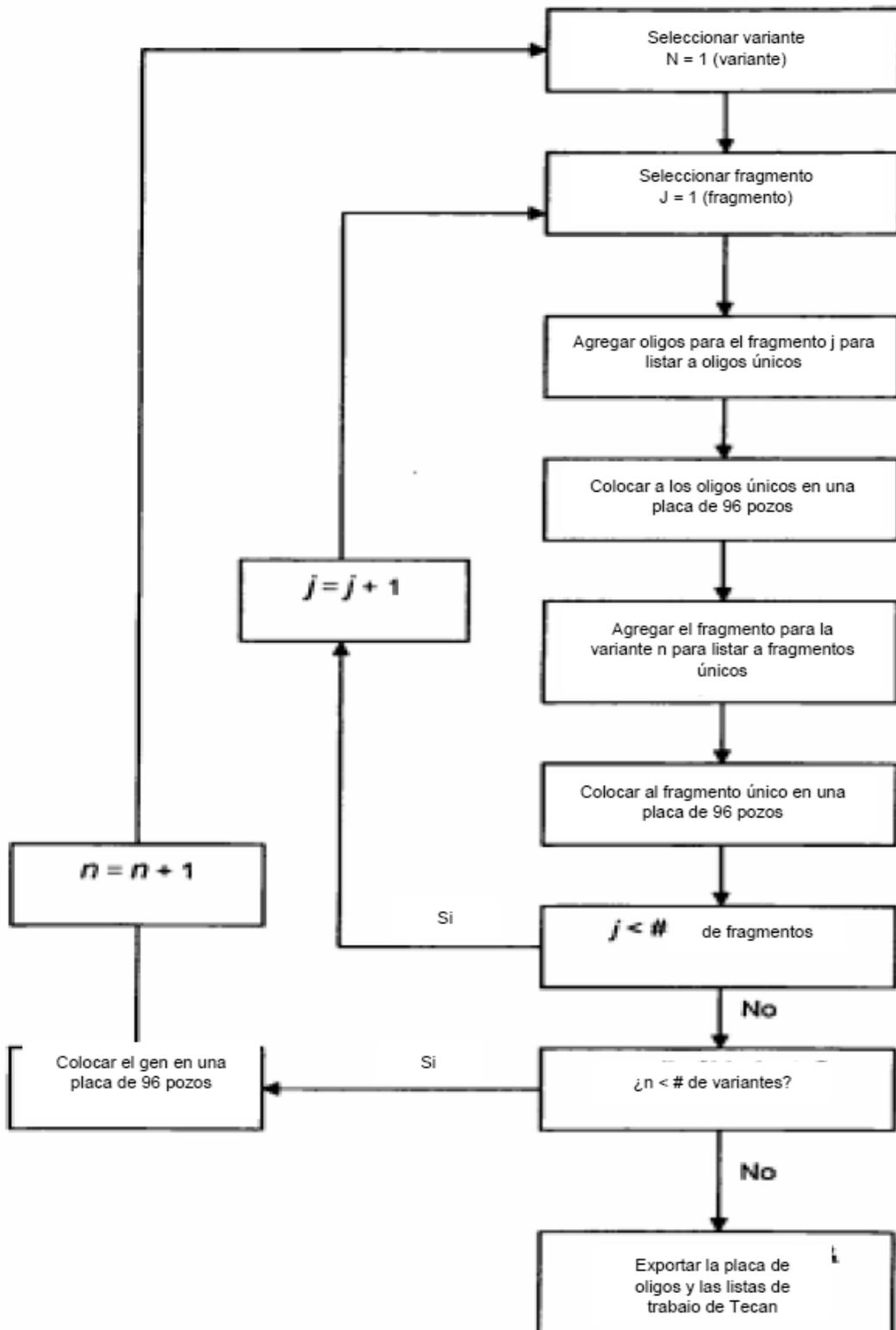


FIG. 6