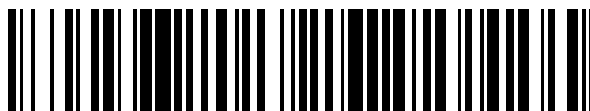


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 575 005**

51 Int. Cl.:

C12N 15/10 (2006.01)

C12N 15/11 (2006.01)

C07H 21/00 (2006.01)

C12Q 1/68 (2006.01)

C40B 50/06 (2006.01)

C40B 40/06 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **11.06.2009 E 09763625 (2)**

97 Fecha y número de publicación de la concesión europea: **09.03.2016 EP 2285958**

54 Título: **Método de síntesis de variantes de polinucleótidos**

30 Prioridad:

13.06.2008 US 61581 P

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

23.06.2016

73 Titular/es:

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

72 Inventor/es:

**COLBECK, JEFFREY;
MIJTS, BENJAMIN;
GIVER, LORRAINE, JEAN y
FOX, RICHARD,, J.**

74 Agente/Representante:

IZQUIERDO BLANCO, María Alicia

ES 2 575 005 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).

Método de síntesis de variantes de polinucleótidos

Descripción

1. ANTECEDENTES

[0001] Diversas técnicas de evolución dirigida basadas en *in silico* e *in vitro* de la función de proteína han permitido la generación de proteínas con propiedades novedosas. Por ejemplo, las enzimas del citocromo P450 han sido evolucionadas para tener actividad contra los sustratos normalmente no reconocidos por la enzima de origen natural (véase, por ejemplo, Landwehr et al, 2007, Chem Biol 14 (3): 269-78; Kubo et al., 2006, Chemistry 12(4):1216-20.). Típicamente, para la generación de tales nuevas enzimas, un polinucleótido que codifica un polipéptido de referencia, tal como una enzima de tipo silvestre, se somete a mutagénesis para generar polinucleótidos que codifican variantes de polipéptidos con cambios en la secuencia de aminoácidos. El examen de las variantes para una propiedad deseada, tal como una mejora en una estabilidad de la enzima o actividad frente a 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 la población de variantes seleccionadas. Por ejemplo, una mutación asociada con la estabilidad térmica de una enzima no puede ser encontrada en asociación con una mutación asociada con un cambio en la especificidad de sustrato. Este sesgo en la población puede surgir de varios factores, incluyendo, entre otros, la secuencia de aminoácidos parental codificada por el polinucleótido usado para la mutagénesis, es posible la selección en contra de la combinación durante la propagación *in vitro* del polinucleótido, y el sesgo en la técnica utilizada para la mutagénesis (por ejemplo, uso de polimerasas de introducir errores). EE.UU. 6,153,410 muestra un método para la mutagénesis *in vitro* y la recombinación de secuencias de polinucleótidos basadas en extensión de polimerasa catalizada de oligonucleótidos cebadores.

[0002] Debido a que las mutaciones en las posiciones de residuos de aminoácidos definidas de una secuencia de polipéptido de referencia pueden proporcionar una gran cantidad de información acerca de las actividades biológicas del polipéptido, una vez que se han identificado mutaciones inicialmente, es deseable preparar varias combinaciones de las mutaciones no se encuentran en la inicial conjunto de variantes seleccionados que pueden ser probados para la propiedad deseada. La selección basada *In silico* de mutaciones definidas o conjuntos de mutaciones proporcionan un marco para la generación de un gran número de posibles combinaciones de mutaciones. por ejemplo, las mutaciones que afectan a la especificidad de sustrato se pueden combinar con mutaciones que afectan otras propiedades de las enzimas, incluyendo, entre otros, la actividad enzimática, estabilidad térmica, y el inhibidor de la resistencia. Típicamente, el enfoque para la generación de estos polipéptidos que tienen nuevas combinaciones de mutaciones es la sintetización de especies individuales (es decir, la síntesis de cada polinucleótido que codifica el gen mutante). Esto se puede lograr mediante síntesis química y/o enzimática del polinucleótido en combinación con técnicas de recombinación estándar. Tales técnicas de síntesis *de novo* requieren la síntesis de la totalidad de genes de cada variante de polinucleótido y/o síntesis de un gran número de cebadores de oligonucleótidos que luego se utilizan para sintetizar la variante de polinucleótido entero (por ejemplo, a través de ION-PCR). Estas técnicas requieren más síntesis de oligonucleótidos y dan como resultado menores rendimientos de variantes que tienen la secuencia correcta. En consecuencia, si el conjunto de datos de mutaciones es grande, el costo y la eficiencia de la generación de las combinaciones de mutación pueden limitar la capacidad para detectar un gran número de nuevas combinaciones. Por lo tanto, los métodos eficaces y rentables de generar polinucleótidos que codifican combinaciones de mutaciones definidas son deseables.

2. RESUMEN

[0003] La presente invención proporciona un método de síntesis de una pluralidad de variantes de polinucleótido que tienen cada uno al menos una diferencia de nucleótidos definida en relación con una secuencia de polinucleótidos de referencia, comprendiendo el método: (a) amplificar por separado una plantilla de polinucleótido de referencia con cada uno de una pluralidad de pares de cebadores directo e inverso, en el que la pluralidad de pares de cebadores directo e inverso comprende la pluralidad de diferencias de nucleótidos definidas y en el que cada par genera un amplicón que comprende una secuencia capaz de unirse a una secuencia de solapamiento adyacente de al menos otro amplicón, generando de este modo una biblioteca direccionable de amplicones en el que: (i) la biblioteca de amplicones comprende elementos para el montaje de dos o más diferentes variantes de polinucleótido que comprende diferencias de nucleótidos definidas en relación con la secuencia de polinucleótidos de referencia; y (ii) los miembros de la biblioteca están dispuestos en lugares distintos posicionalmente sobre un sustrato; (b) ensamblar por separado una pluralidad de conjuntos de amplicones, en el que cada conjunto comprende amplicones que tienen secuencias adyacentes superpuestas capaces de unirse para formar la longitud completa de la secuencia de polinucleótidos de referencia; (c) la replicación de la pluralidad de conjuntos de amplicones reunidos, generando de ese modo una biblioteca direccionable de variantes de polinucleótido en el que los miembros de la biblioteca están dispuestos en lugares distintos posicionalmente sobre un sustrato.

[0004] La invención también proporciona una biblioteca direccionable de amplicones que comprenden una pluralidad de amplicones, en el que cada miembro de la biblioteca de los amplicones comprende al menos una diferencia de nucleótidos definida en relación con una secuencia de polinucleótidos de referencia y una región adyacente superpuesta capaz de unirse a la región adyacente superposición de al menos otro amplicón en la biblioteca, en el

que la pluralidad de amplicones comprenden una pluralidad de conjuntos de amplicones capaces de unirse para formar la longitud completa de la secuencia de polinucleótidos de referencia, en el que la pluralidad de amplicones comprende miembros para el montaje de dos o más diferentes variantes de polinucleótido de longitud completa que comprende diferencias de nucleótidos definidas con respecto a la secuencia de polinucleótido de referencia, en el que: (i) al menos uno de la pluralidad de conjuntos de amplicones comprende miembros para el montaje de una variante de polinucleótido que comprende al menos 5 diferencias de nucleótidos definidas en relación con la secuencia de polinucleótidos de referencia; (ii) al menos uno de la pluralidad de conjuntos de amplicones comprende al menos 3 diferentes amplicones; o (iii) la longitud de la secuencia de polinucleótidos de referencia es de al menos 500 pb, y en el que los miembros de la biblioteca están dispuestos en ubicaciones distintas posicionalmente sobre un sustrato.

[0005] La presente descripción se refiere a métodos de generación eficiente de polinucleótidos que tienen diferentes combinaciones de cambios de secuencia definidos (por ejemplo, mutaciones de aminoácidos deseadas) en comparación con una secuencia de polinucleótido de referencia. Los métodos se basan en el uso de una biblioteca de fragmentos de polinucleótidos (es decir, amplicones) que tienen regiones adyacentes se superponen de manera que un conjunto de los fragmentos de polinucleótidos pueden ser ensamblados para generar una pluralidad de variantes de polinucleótido que tienen cada uno un conjunto definido de cambios de secuencia. Cebadores delanteros y traseros seleccionados se utilizan para introducir los cambios de secuencia mediante la amplificación de una plantilla de polinucleótido de referencia y por lo tanto generar fragmentos de polinucleótido que comprende la secuencia de los cambios definidos. La biblioteca está diseñada para tener suficientes fragmentos de polinucleótidos para ensamblar al menos dos secuencias variantes de polinucleótidos diferentes. En algunas realizaciones, la biblioteca de fragmentos de polinucleótidos contiene miembros que tienen todas las diferencias definidas en la secuencia de polinucleótido (por ejemplo, cambios de nucleótidos deseados) en comparación con una secuencia de referencia de tal manera que todas las permutaciones de la secuencia se pueden montar. En algunas realizaciones, los polinucleótidos pueden ser diseñados para codificar polipéptidos que tienen diferencias definidas en la secuencia de aminoácidos en comparación con una secuencia de aminoácidos de referencia.

[0006] Los procedimientos de la presente descripción son capaces de producir grandes bibliotecas de secuencias variantes de polinucleótidos que tienen diferencias de nucleótidos definidas (por ejemplo, bibliotecas de 10, 50, 100, 150, o más variantes, cada uno con 1, 2, 3, 5, 9, 12, 15, o más, cambios deseados), con relativamente pocos (por ejemplo, en comparación con los métodos de síntesis de genes enteros), y oligonucleótidos relativamente cortos (por ejemplo, 35-mer o menos), y en el que el porcentaje medio de secuencias correctas es sorprendentemente alto (por ejemplo, al menos 65%, 75%, 85%, 95%, o más).

[0007] En algunas realizaciones, el método de formación de los polinucleótidos que codifican variantes de polipéptidos puede comprender: la selección de una pluralidad de diferencias de restos de aminoácidos definidos relativos a una secuencia de aminoácidos de referencia; la definición de segmentos de superposición de una secuencia de polinucleótido que codifica el polipéptido con la diferente secuencia de aminoácidos o la secuencia de aminoácidos de referencia, cada segmento estando limitada por un conjunto de avance y retroceso secuencias de cebador de unión y en el que una diferencia en la secuencia de polinucleótidos que codifica cada una de la pluralidad de diferencias de residuos de aminoácido se engloban en la secuencia de unión a cebador; amplificando cada segmento con el conjunto de cebadores directo e inverso, en el que cebadores delanteros y/o traseros seleccionados contienen las diferencias de secuencia de polinucleótidos, para generar una biblioteca de amplicones que comprenden miembros que codifican diferencias de residuo de aminoácido, en las que la biblioteca comprende elementos para el montaje de dos o más diferentes permutaciones de secuencias de aminoácidos de las diferencias de aminoácidos definidas; el montaje de la biblioteca de un conjunto de amplicones que tienen regiones adyacentes solapadas complementarias, en las que el conjunto de amplicones juntos codifican el polipéptido con una permutación de secuencia de aminoácidos definida que tiene una o más diferencias de residuos de aminoácidos; y replicar el conjunto de amplicones ensamblados para sintetizar el polinucleótido que codifica el polipéptido.

[0008] Además se describen aquí métodos de generación de la biblioteca de fragmentos de polinucleótidos, donde el método comprende: (a) la generación de una pluralidad de permutaciones de secuencias de aminoácidos que difieren de una secuencia de aminoácidos de referencia basado en una pluralidad de diferencias de restos de aminoácidos definida a partir de la secuencia de referencia de aminoácidos, y para cada permutación; (i) la determinación de una secuencia de polinucleótidos que codifica la secuencia de aminoácidos de permutación basada en una secuencia de polinucleótidos de referencia; (ii) la identificación de un cambio en la secuencia de polinucleótidos que codifica una diferencia de residuo de aminoácido en comparación con una secuencia de aminoácidos de referencia, y la determinación de la proximidad de un cambio de vecino más próximo en la secuencia de polinucleótido que codifica otra diferencia de residuo de aminoácido en la secuencia de aminoácidos de permutación; (iii) seleccionar un cebador de oligonucleótido delantero que tiene una secuencia que codifica la diferencia de residuos de aminoácidos y, opcionalmente, incluyendo el cambio del vecino más próximo en la secuencia de polinucleótido en el mismo cebador de oligonucleótidos delanteros si la proximidad del primer cambio en la secuencia de polinucleótido; (iv) la identificación del siguiente cambio en la secuencia de polinucleótido o hasta que se alcance el extremo del polinucleótido, y la selección de un cebador inverso de oligonucleótido para amplificar un fragmento de polinucleótido con los cebadores delanteros de oligonucleótidos, en el que el cebador inverso codifica opcionalmente el próximo cambio en diferencias de residuo de aminoácidos; (v) la reiteración de pasos (ii) a

(iv) para cada cambio en la secuencia de polinucleótido que codifica una diferencia de residuos de aminoácidos de tal manera que todos los cambios en la secuencia de polinucleótidos están presentes en los cebadores de oligonucleótidos; y (b) amplificar con cada conjunto de cebadores delanteros y traseros de oligonucleótidos para generar la biblioteca de amplicones que tienen miembros que codifican diferencias de residuos de aminoácidos superpuestos.

[0009] En otro aspecto, la presente descripción proporciona biblioteca de tales fragmentos de polinucleótidos (es decir, amplicones) para el montaje de los polinucleótidos. En algunas realizaciones, la pluralidad de fragmentos de polinucleótido comprende fragmentos de polinucleótidos con la superposición de las regiones adyacentes, cada fragmento de polinucleótido de ser limitada por las secuencias de unión del cebador para los cebadores delanteros y traseros, en el que la pluralidad de fragmentos de polinucleótidos tienen miembros que codifican en las secuencias de unión a una diferencia de residuo de aminoácido específica a partir de una pluralidad definida de las diferencias de residuos de aminoácidos con respecto a una secuencia de aminoácidos de referencia de tal manera que la pluralidad de fragmentos de polinucleótido codifica toda una pluralidad seleccionada de diferencias de restos de aminoácidos de la pluralidad definida de diferencias de restos de aminoácidos; y en el que la pluralidad de fragmento de polinucleótido comprende miembros para el montaje de dos o más diferentes permutaciones de secuencias de aminoácidos de las diferencias de aminoácidos definidas. En algunas realizaciones, la pluralidad de fragmentos de polinucleótido comprende miembros suficientes para el montaje de la totalidad de las permutaciones de secuencia de aminoácidos posible de la pluralidad seleccionada de las diferencias de residuo de aminoácido.

[0010] En otro aspecto, la presente descripción proporciona un método para sintetizar una pluralidad de variantes de polinucleótido que tienen cada una una diferencia de nucleótidos definida en relación con una secuencia de polinucleótidos de referencia, en el que el método comprende: (a) la amplificación por separado de una plantilla de polinucleótido de referencia con cada uno de una pluralidad de pares de cebadores delanteros y traseros, en el que la pluralidad de pares de cebadores delanteros y traseros comprende la pluralidad de diferencias de nucleótidos definidas y en el que cada par genera un amplicón que comprende una secuencia capaz de unirse a una secuencia de solapamiento adyacente de al menos otro amplicón; (b) ensamblar por separado una pluralidad de conjuntos de amplicones, en el que cada conjunto comprende amplicones que tienen secuencias adyacentes superpuestas capaces de unirse para formar la longitud completa de la secuencia de polinucleótidos de referencia; y (c) la replicación de la pluralidad de conjuntos de amplicones ensamblados, sintetizando de ese modo una pluralidad de variantes de polinucleótidos.

[0011] En otra realización, la descripción proporciona un método para sintetizar una pluralidad de variantes de polinucleótido que tienen cada uno una diferencia de nucleótidos definida en relación con una secuencia de polinucleótidos de referencia, comprendiendo el método: (a) la selección de una pluralidad de diferencias de nucleótidos definida en relación con la secuencia de polinucleótidos de referencia; (b) que define una pluralidad de segmentos de la secuencia de polinucleótidos de referencia, en el que cada segmento se solapa con al menos un segmento adyacente y está delimitada por un par de secuencias de unión de cebadores delanteros y traseros, en el que los cebadores delanteros y/o traseros comprenden al menos una de la pluralidad de diferencias de nucleótidos definidas; (c) amplificar por separado una plantilla de polinucleótido de referencia con cada uno de una pluralidad de pares de cebadores delanteros y traseros, en el que cada par de cebadores comprende al menos uno de la pluralidad de diferencias de nucleótidos definidas, generando de este modo una biblioteca direccionable de amplicones cada uno correspondiente a una segmento de la secuencia de polinucleótidos de referencia que tiene un diferencias de nucleótidos definidas; (d) ensamblar por separado una pluralidad de conjuntos de amplicones de la biblioteca direccionable de amplicones, en el que cada conjunto comprende amplicones que corresponde a la superposición de los segmentos adyacentes que comprenden la longitud completa de la secuencia de polinucleótidos de referencia; y (e) la replicación de la pluralidad de conjuntos de amplicones ensamblados, sintetizando de ese modo una pluralidad de variantes de polinucleótidos.

[0012] En algunas realizaciones de los métodos de síntesis de una pluralidad de variantes de polinucleótido se describe aquí, los métodos pueden llevarse a cabo en el que el polinucleótido de referencia codifica un polipéptido de referencia y cada uno de la pluralidad de variantes de polinucleótido codifica un polipéptido que tiene al menos una diferencia de secuencia de aminoácido. En realizaciones adicionales, los métodos pueden ser llevados a cabo en el que comprende además las etapas de: (i) la clonación de cada una de la pluralidad de variantes de polinucleótido en un vector de expresión; (ii) la transformación de células con los vectores de expresión; (iii) la selección de las células transformadas para la actividad de los polipéptidos codificados por las variantes de polinucleótidos; o (iv) el aislamiento de al menos un polipéptido codificado por las variantes de polinucleótidos. Además, los métodos pueden llevarse a cabo en los que cada variante de polinucleótido se monta en una posición conocida en una matriz.

[0013] En algunas realizaciones de los métodos de síntesis de una pluralidad de variantes de polinucleótido se describe aquí, los métodos pueden llevarse a cabo en donde las secuencias de la pluralidad de secuencias de cebadores delanteros y traseros se generan por las etapas de: (i) la identificación de una primera diferencia definida en la secuencia variante de polinucleótido en comparación con la secuencia de referencia, y la determinación de la proximidad de una diferencia de vecino más cercana se define en la secuencia de polinucleótido; (ii) la selección de un cebador directo que tiene una secuencia que comprende la primera diferencia de nucleótidos definida, y,

opcionalmente, incluyendo cualquier diferencia de vecino más cercano se define en el mismo cebador delantero si la proximidad de la primera diferencia de nucleótidos definida; (iii) identificar una siguiente diferencia definida en la secuencia variante de polinucleótido en comparación con la secuencia de referencia, y la determinación de la proximidad de una diferencia definida de vecino más próxima en la secuencia de polinucleótidos, o la identificación de que el final de la variante de polinucleótido ha sido alcanzada; (iv) la selección de un cebador trasero que tiene una secuencia que comprende la siguiente diferencia de nucleótidos definida, y, opcionalmente, incluyendo cualquier diferencia de vecino más cercana definida en el mismo cebador trasero si la proximidad de la siguiente diferencia de nucleótidos definida; y (v) repetir las etapas (iii) a (iv) para cada diferencia definida en la secuencia variante de polinucleótido tal que toda diferencia definida están presentes en los cebadores.

[0014] En algunas realizaciones de los métodos de síntesis de una pluralidad de variantes de polinucleótido se describe aquí, la pluralidad de variantes de polinucleótidos comprende al menos 10, 25, 35, 50, 75, 90, 120, 150, 180 o incluso más diferentes variantes de polinucleótidos .

[0015] En algunas realizaciones de los métodos de síntesis de una pluralidad de variantes de polinucleótido se describen aquí, al menos una de la pluralidad de variantes de polinucleótidos comprende al menos 2, 3, 6, 9, 12, 15, 18, 21, 24, 27 , 30, o diferencias incluso más definidas de nucleótidos con respecto a la secuencia de referencia de polinucleótidos. En algunas realizaciones, dos o más, o en algunas formas de realización cada una de la pluralidad de variantes de polinucleótidos comprende al menos 1, 2, 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, o diferencias incluso más definidas de nucleótidos respecto a la secuencia de polinucleótidos de referencia.

[0016] En algunas realizaciones de los métodos de síntesis de una pluralidad de polinucleótidos de variantes se describe aquí, al menos una de la pluralidad de conjuntos de amplicones comprende al menos 3, al menos 5, al menos 7, al menos 10, o más amplicones diferentes. En algunas realizaciones, dos o más, o en algunas formas de realización cada una de la pluralidad de conjuntos de amplicones comprende al menos 3, al menos 5, al menos 7, al menos 10, o más amplicones diferentes.

[0017] En algunas realizaciones de los métodos de síntesis de una pluralidad de variantes de polinucleótido se describe aquí, la longitud de la secuencia de polinucleótidos de referencia es al menos 500 pb, 750 pb, 1000 pb, 1250 pb, 1500 pb, o incluso más largo.

[0018] En algunas realizaciones de los métodos de síntesis de una pluralidad de variantes de polinucleótidos divulgadas en el presente documento, la pluralidad de pares de cebadores delanteros y traseros comprende 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 pares de cebadores delanteros y traseros comprende de 6 a aproximadamente 200, de 6 a aproximadamente 150, de 6 a aproximadamente 100, de 6 a aproximadamente 50, 6 a aproximadamente 40, 6 a aproximadamente 30, 6 a aproximadamente 25, 6 a aproximadamente 20, 6 a aproximadamente 15, o incluso menos oligonucleótidos diferentes, y en el que 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.

[0019] En algunas realizaciones de los métodos de síntesis de una pluralidad de variantes de polinucleótidos descritas aquí, el porcentaje medio de la pluralidad de variantes de polinucleótidos sintetizados que comprende la secuencia correcta es al menos aproximadamente 65%, 75%, 85%, o 95%, o más.

[0020] En algunas realizaciones, la descripción proporciona métodos para sintetizar una pluralidad de variantes de polinucleótido en las que cualquiera de los parámetros descritos anteriormente (por ejemplo, número de variantes, número de diferencias de nucleótidos definidas, la longitud de la secuencia de polinucleótidos de referencia, número de pares de cebadores delanteros e traseros, longitud de cebadores de oligonucleótidos, y/o porcentaje de secuencias perfectas de longitud completa) se combinan.

[0021] Además de los métodos anteriores, la descripción también proporciona una biblioteca de variantes direccionables de polinucleótido que comprende una pluralidad de variantes de polinucleótidos o amplicones sintetizados según cualquiera de los métodos anteriores. Por consiguiente, en algunas realizaciones, la descripción proporciona una biblioteca direccionable de amplicones, en el que cada miembro de la biblioteca de los amplicones comprende al menos una diferencia de nucleótidos definidas en relación con una secuencia de polinucleótidos de referencia y una región adyacente superpuesta capaz de unirse a la región adyacente de solapamiento de por lo menos otro amplicón en la biblioteca, y en el que la pluralidad de amplicones comprenden al menos un conjunto de amplicones capaces de unirse para formar la longitud completa de la secuencia de polinucleótido de referencia. En algunas realizaciones, la biblioteca direccionable de amplicones comprende miembros para el montaje de dos o más diferentes variantes de polinucleótido que comprende diferencias de nucleótidos definidas con respecto a la secuencia de polinucleótido de referencia. En algunas realizaciones de la biblioteca direccionables de amplicones, la secuencia de polinucleótidos de referencia codifica un polipéptido de referencia y la pluralidad de amplicones comprende miembros suficientes para el montaje de todas las posibles diferencias de nucleótidos que codifica una pluralidad seleccionada de diferencias de restos de aminoácidos.

[0022] También se describe en este documento métodos implementadas en ordenador para llevar a cabo diversos pasos de los métodos aquí descritos.

3. BREVE DESCRIPCIÓN DE LAS FIGURAS

[0023]

FIG. 1 muestra una técnica estándar para la generación de polinucleótidos que codifican variantes de polipéptidos definidos (a la izquierda) en comparación con el método descrito en el presente documento para el uso de bibliotecas de la superposición de fragmentos de polinucleótidos (a la derecha).

FIG. 2 proporciona un esquema de flujo de trabajo de ejemplo para la generación de bibliotecas de la superposición de fragmentos de polinucleótidos basadas en la generación de cebadores de oligonucleótidos para la superposición de los segmentos de un polinucleótido y el uso de los oligonucleótidos en reacciones de PCR para generar bibliotecas de la superposición de fragmentos de polinucleótidos.

FIG. 3 muestra los geles de agarosa de 96 muestras (y 8 controles) resultantes de ensamblaje de los fragmentos de polinucleótidos se solapan y la reproducción de los fragmentos de polinucleótidos ensamblados para sintetizar el variantes de polinucleótido que codifican la variante de polipéptido deseado. Casi todos los geles mostraron una única banda fuerte que indica la secuencia de longitud esperada estaba presente.

FIG. 4. La figura 4 muestra un diagrama de flujo para la generación de polinucleótidos que codifican variantes mediante el uso de la biblioteca de la superposición de fragmentos de polinucleótidos.

FIG. 5 muestra un diagrama de flujo para la generación de una biblioteca de cebadores de oligonucleótidos para la generación de una biblioteca de la superposición de amplicones de polinucleótidos para cada secuencia de aminoácidos de permutación.

FIG. 6 muestra un diagrama de flujo de instrucciones para la creación y selección de cebadores de oligonucleótidos y fragmentos de oligonucleótidos superpuestos automatizado.

4. DESCRIPCIÓN DETALLADA

[0024] Tal como se utiliza en esta memoria descriptiva y las reclamaciones adjuntas, las formas singulares "un", "una" y "el" incluyen referentes plurales a menos que el contexto indique claramente lo contrario. Así, por ejemplo, la referencia a "una proteína" incluye más de una proteína, y la referencia a "un compuesto" se refiere a más de un compuesto.

[0025] Además, el uso de "o" significa "y/o" a menos que se indique lo contrario. Del mismo modo, "comprender", "comprende", "que comprende" "incluir", "incluye" e "incluyendo" son intercambiables y no pretenden ser limitantes.

[0026] Se ha de entender además que en la descripción de varias formas de realización utilizan el término "que comprende", los expertos en la técnica entenderán que, en algunos casos específicos, una forma de realización puede describirse, alternativamente, usando el lenguaje "que consiste esencialmente en" o "que consiste de."

[0027] Los títulos de las secciones que aparecen aquí son únicamente para fines de organización y no deben interpretarse como una limitación del objeto descrito. En la presente memoria, los siguientes términos están destinados a tener los siguientes significados.

4.1 Definiciones

[0028] "Amplificar" y "amplificación", como se usa en el presente documento, incorporan su uso común y se refieren al uso de cualquier metodología de amplificación adecuada para generar o detectar cualquier polinucleótido, recombinante o expresado de forma natural, susceptible a la amplificación *in vivo* o *in vitro*, tal como mediante reacción en cadena de la polimerasa (PCR).

[0029] "Amplicón" se refiere al producto de la reacción de amplificación generado a través de la extensión de uno o ambos de un par de cebadores de amplificación. Un amplicón puede contener exponencialmente ácidos nucleicos amplificados si ambos cebadores utilizados se hibridan a una secuencia diana. Como alternativa, los amplicones se pueden generar mediante amplificación lineal si uno de los cebadores utilizados no se hibrida a la secuencia diana. por lo tanto, este término se utiliza genéricamente en la presente memoria y no implica la presencia de ácidos nucleicos de manera exponencial amplificados.

[0030] "Recocado" o "hibridación" se refiere a las interacciones de apareamiento de bases de un polímero de nucleobase con otra que resulta en la formación de una estructura de doble cadena, una estructura triplex o una

estructura cuaternaria. El recocido o hibridación puede ocurrir a través de interacciones de apareamiento de bases Watson Crick, pero puede mediarse por otras interacciones de enlace de hidrógeno, tales como el emparejamiento de bases de Hoogsteen.

[0031] "Montaje" se refiere a la reunión de una pluralidad de fragmentos de polinucleótidos (por ejemplo, amplicones) en condiciones en las que las regiones complementarias entre polinucleótidos pueden reconocerse para formar un complejo de hibridación, por ejemplo, que tenga una doble región hibridada hebra con voladizos para las regiones no complementarias. Una pluralidad de polinucleótidos puede ser ensamblada para formar un polinucleótido más grande que codifica un polipéptido de interés.

[0032] "Polinucleótido puente" se refiere a un polinucleótido que tiene regiones complementarias a las regiones terminales de tal manera que un polinucleótido se puede hibridar con una región terminal y otro polinucleótido puede hibridarse a la otra región terminal del polinucleótido de puente.

[0033] "Secuencia de codificación" se refiere a aquella porción de un ácido nucleico (por ejemplo, un gen) que codifica una secuencia de aminoácidos de una proteína.

[0034] "Optimizado por codones" se refiere a cambios en los codones del polinucleótido que codifica una proteína a los utilizados preferentemente en un organismo particular de tal manera que la proteína codificada se expresa de manera eficiente en el organismo de interés. Aunque el código genético es degenerado en que la mayoría de los aminoácidos están representados por varios codones, llamados "sinónimos" o codones "sinónimas", es bien conocido que el uso de codones por organismos particulares no es aleatorio y sesgado hacia determinados tripletes de codones. Este sesgo del uso de codón puede ser mayor en referencia a un gen dado, genes de la función o de origen ancestral común, proteínas altamente expresadas frente a las proteínas de bajo número de copias, y las regiones de codificación de proteínas agregadas del genoma de un organismo.

[0035] "Complementario" se refiere a la hibridación o la base de emparejamiento entre los nucleótidos o ácidos nucleicos, tales como, por ejemplo, entre las dos hebras de una molécula de ADN de doble hebra o entre un cebador de oligonucleótido y un sitio de unión a cebador en un polinucleótido monocatenario a ser secuenciado o amplificado. Nucleótidos complementarios son, en general, A y T (o A y A), o C y G. Se dice que dos moléculas de ARN o ADN de hebra doble son sustancialmente complementarias cuando una hebra de polinucleótido (ARN o ADN) se hibridará bajo condiciones de hibridación selectivas a su complemento. Típicamente, la hibridación selectiva se producirá cuando existe al menos aproximadamente 65% complementarias en un tramo de al menos 14 a 25 nucleótidos, preferiblemente al menos aproximadamente 75%, más preferiblemente al menos aproximadamente 90% complementarias. Véase, por ejemplo, M. Kanehisa, 1984, Nucleic Acids Res 12:203. "Complementario a" se utiliza en la presente memoria para significar que la secuencia complementaria es sustancialmente idéntica o idéntica a el complemento trasero de la totalidad o una porción de una secuencia de referencia o polinucleótido que cada nucleótido en una hebra es capaz de formar un par de bases con un nucleótido o análogo de la misma en la cadena opuesta.

[0036] "Sustituciones conservadoras de aminoácidos" se refieren a la intercambiabilidad de residuos que tienen cadenas laterales similares, y por lo tanto típicamente implica la sustitución del aminoácido en el polipéptido con aminoácidos dentro de la misma o similar clase definida de aminoácidos. A modo de ejemplo y no de limitación, un aminoácido con una cadena lateral alifática puede estar sustituido con otro aminoácido alifático, por ejemplo, alanina, valina, leucina, isoleucina y metionina; un aminoácido con la cadena lateral de hidroxilo está sustituido con otro aminoácido con una cadena lateral hidroxilo, por ejemplo, serina y treonina; un aminoácido que tenga cadenas laterales aromáticas estará sustituido con otro aminoácido que tiene una cadena lateral aromática, por ejemplo, fenilalanina, tirosina, triptófano e histidina; un aminoácido con una cadena lateral básica estará sustituido con otro aminoácido con una cadena lateral de base, por ejemplo, lisina, arginina, e histidina; un aminoácido con una cadena lateral ácida está sustituido con otro aminoácido con una cadena lateral ácida, por ejemplo, ácido aspártico o ácido glutámico; y un aminoácido hidrófobo o hidrófilo se sustituye con otro aminoácido hidrófobo o hidrófilo, respectivamente.

[0037] "Secuencia de control" se refiere a una secuencia de polinucleótido utilizado para efectuar la expresión de secuencias de codificación y de no codificación a las que están asociadas. La naturaleza de tales secuencias de control difiere dependiendo del organismo huésped. Las secuencias de control generalmente incluyen promotor, sitio de unión ribosomal, y la secuencia de terminación de la transcripción. El término "secuencias de control" pretende incluir componentes cuya presencia puede influir en la expresión, y también puede incluir componentes adicionales cuya presencia es ventajosa, por ejemplo, el secuencias líderes y secuencias de compañeros de fusión.

[0038] "Diferencia definida" como se usa en la presente memoria en el contexto de mutaciones a una secuencia de polinucleótido o polipéptido se refiere a un *a priori* especificado, seleccionado, y/o cambio deseado a la secuencia (por ejemplo, un cambio de nucleótido de c a g en una posición seleccionada de una secuencia de polinucleótidos que resulta en un aminoácido diferente en una posición deseada del polipéptido codificado).

[0039] "Deleción" con respecto a un polipéptido o polinucleótido se refiere a la eliminación de uno o más

aminoácidos o nucleótidos a partir del polipéptido de referencia o polinucleótido, respectivamente. Las deleciones puede comprender la eliminación de más aminoácidos más aminoácidos o nucleótidos 1 o, 2 o 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, de hasta 10% del número total de aminoácidos o nucleótidos, o hasta el 20% del número total de aminoácidos o nucleótidos que componen el polipéptido de referencia o polinucleótido. Las deleciones pueden ser dirigidas a las partes internas y/o porciones terminales del polipéptido o polinucleótido. En diversas realizaciones, la eliminación puede comprender un segmento continuo o puede ser discontinuo.

[0040] "Heterólogo" cuando se usa con referencia a un ácido nucleico o polipéptido, indica que una secuencia que comprende dos o más subsecuencias que no se encuentran en la misma relación entre sí como se encuentra normalmente en la naturaleza, o se ha diseñado de forma recombinante de modo que su nivel de expresión, o relación física con otros ácidos nucleicos u otras moléculas en una célula, o de la estructura, no se encuentra normalmente en la naturaleza. Por ejemplo, un ácido nucleico heterólogo típicamente se produce de forma recombinante, teniendo dos o más secuencias de genes no relacionados dispuestos de manera que no se encuentra en la naturaleza, por ejemplo, un marco de lectura abierta de ácido nucleico (ORF) de la invención operativamente unido a una secuencia de promotor insertado en un casete de expresión, por ejemplo, un vector.

[0041] "Inserción" o "Adición" se refiere a un cambio en una secuencia de nucleótidos o de aminoácidos mediante la adición de uno o más nucleótidos o residuos de aminoácidos, respectivamente, en comparación con una secuencia de referencia, tal como por ejemplo, una secuencia de tipo silvestre.

[0042] "Biblioteca" se refiere a un conjunto (por ejemplo, una pluralidad) de polipéptidos o ácidos nucleicos heterogéneos. Una biblioteca está compuesta por miembros, que tienen un único polipéptido o secuencia de ácido nucleico. En este sentido, "biblioteca" es sinónima de "repertorio". Las diferencias de secuencia entre miembros de la biblioteca son responsables de la diversidad presente en la biblioteca. La biblioteca puede tomar la forma de una simple mezcla de polipéptidos o ácidos nucleicos, o puede ser en los organismos de formulario o células, por ejemplo, bacterias, virus, animales o células vegetales y similares, transformados con una biblioteca de ácidos nucleicos.

[0043] "Sustitución no conservadora" se refiere a la sustitución de un aminoácido en el polipéptido con un aminoácido con propiedades significativamente diferentes de la cadena lateral. Las sustituciones no conservativas pueden utilizar aminoácidos entre, en lugar de dentro, los grupos definidos y afecta a (a) la estructura del esqueleto peptídico en el área de sustitución (por ejemplo, prolina para la glicina) (b) la carga o hidrofobicidad, o (c) el grueso de la cadena lateral. A modo de ejemplo y no de limitación, un ejemplar de sustitución no conservadora puede ser un aminoácido ácido sustituido con un aminoácido básico o alifático; un aminoácido aromático sustituido con un aminoácido pequeño; y un aminoácido hidrófilo sustituido con un aminoácido hidrófobo.

[0044] "De origen natural" o "de tipo silvestre" se refiere a la forma que se encuentra en la naturaleza. Por ejemplo, una de origen natural o un polipéptido de tipo silvestre o secuencia de polinucleótidos es una secuencia presente en un organismo que puede aislarse de una fuente en la naturaleza y que no ha sido modificado intencionadamente por manipulación humana.

[0045] "Nucleobase" o "base" se refiere a aquellos de origen natural y fragmentos heterocíclicos sintéticos comúnmente conocidos por los que utilizan la tecnología de ácido o polinucleótido nucleico o utilizan poliamida o tecnología de ácido nucleico peptídico para generar de ese modo los polímeros que pueden hibridar con los polinucleótidos de una manera específica de secuencia. Los ejemplos no limitantes de nucleobases adecuadas incluyen: adenina, citosina, guanina, timina, uracilo, 5-propinilo-uracilo, 2-tio-5-propinil-uracilo, 5-metilcitosina, pseudoisocitosina, 2-tiouracilo y 2-tiotimina, 2-aminopurina, N9-(2-amino-6-cloropurina), N9-(2,6-diaminopurina), hipoxantina, N9-(7-deaza-guanina), N9-(7-deaza-8-aza-guanina) y N8-(7-deaza-8-aza-adenina). Otros ejemplos no limitantes de nucleobases adecuadas incluyen las nucleobases ilustradas en las Figuras 2 (A) y 2 (B), de Buchardt et al. (WO 92/20702 o WO 92/20703).

[0046] "Polímero de nucleobase" o "Oligómero" se refiere a dos o más nucleobases que están conectadas por enlaces que permiten el polímero resultante nucleobase u oligómero para hibridar con un polinucleótido que tiene una secuencia de nucleobase complementaria. Polímeros u oligómeros nucleobásicos incluyen, pero no se limitan a, poli- y oligonucleótidos (por ejemplo, polímeros de ADN y ARN y oligómeros), análogos poli- y oligonucleótidos e imitador poli- y oligonucleótidos, tales como ácidos nucleicos de poliamida o de péptidos. polímeros u oligómeros nucleobásicos pueden variar en tamaño desde unas pocas nucleobases, de 2 a 40 nucleobases, a varios cientos de bases nitrogenadas, a varios miles de nucleobases, o más.

[0047] "Vinculado operablemente" se refiere a una relación funcional entre dos o más segmentos de ácidos nucleicos (por ejemplo, ADN). En algunas realizaciones, se refiere a la relación funcional de una secuencia reguladora de la transcripción de una secuencia transcrita. Por ejemplo, un promotor (definido a continuación) está unido operativamente a una secuencia codificante, tal como un ácido nucleico de la invención, si estimula o modula la transcripción de la secuencia codificante en una célula huésped apropiada u otro sistema de expresión. En

general, el promotor de las secuencias reguladoras de la transcripción que se unen operativamente a una secuencia transcrita están físicamente contiguas a la secuencia transcrita, es decir, son de actuación cis. Sin embargo, algunas secuencias reguladoras de la transcripción, tales como potenciadores, no necesitan ser físicamente contiguas o situado muy cerca de las secuencias codificantes cuya transcripción potencian, en algunos formas de realización, la secuencia reguladora es una secuencia reguladora de la traducción unida a una secuencia de codificación.

[0048] "Zona de solapamiento" se refiere a una región de un primer polinucleótido que es complementario a un segundo polinucleótido, donde las zonas de solapamiento son capaces de hibridar entre sí para formar un complejo de hibridación. En general, el primer polinucleótido y el segundo polinucleótido parcialmente se solaparán de tal manera que los polinucleótidos tendrán regiones no complementarias que no recuecen entre los dos polinucleótidos.

[0049] "Permutaciones" se refiere a la disposición de los elementos (por ejemplo, mutaciones de sustitución) de un conjunto finito dado. En el contexto de las descripciones de la presente memoria para los polipéptidos y polinucleótidos, las diferencias en los residuos de aminoácidos o residuos de nucleótidos de una secuencia de referencia, típicamente caracterizadas como "mutaciones", se pueden organizar en varias combinaciones en la secuencia para formar una permutación de la mutación establecida. Permutaciones incluyen mutaciones individuales y todas las combinaciones de mutaciones posibles del conjunto definido.

[0050] "Polinucleótidos" o "Oligonucleótidos" se refieren a base nitrogenada polímeros u oligómeros en los que las nucleobases están unidas por enlaces de fosfato de azúcar (azúcar de la cadena principal de fosfato). Poli- y oligonucleótidos ejemplares incluyen polímeros de 2' desoxirribonucleótidos (ADN) y polímeros de ribonucleótidos (ARN). Un polinucleótido puede estar compuesto enteramente de ribonucleótidos, enteramente de 2' desoxirribonucleótidos o combinaciones de los mismos.

[0051] "Polinucleótido" o "Análogo de Oligonucleótido" se refiere a polímeros de nucleobase u oligómeros en los que las nucleobases están conectadas por una cadena principal de fosfato de azúcar que comprende uno o más análogos de fosfato de azúcar. Análogos típicos de fosfato de azúcar incluyen, pero no se limitan a, alquilofosfonatos de azúcar, fosforamiditas de azúcar, el alquilo de azúcar o sustituidos alquilofosfotriésteres, fosforotioatos de azúcar, fosforoditioatos de azúcar, fosfatos de azúcares y análogos de fosfato de azúcar en la que el azúcar es distinto de 2' desoxirribosa o ribosa, polímeros de nucleobases que tienen interrelaciones de carga positiva de guanidilo de azúcar tales como los descritos en la Patente de Estados Unidos N°. 6.013.785 y la Patente de Estados Unidos N° 5.696.253 (véase, también, Dagani 1995, Chem. Eng. News 4-5: 1153; Dempey et al, 1995, J Am Chem Soc 117:6140-6141). Tales análogos de carga positiva en la que el azúcar es 2'-desoxirribosa se conocen como "DNGs", mientras que aquellos en los que la ribosa de azúcar se les conoce como "generadores de números aleatorios." Se incluyen específicamente dentro de la definición de los análogos poli- y oligonucleótidos están ácidos nucleicos encerrados (LNAs; véase, por ejemplo Elayadi et al, 2002, Biochemistry 41: 9973-9981; Koshkin et al, 1998, J Am Chem Soc 120: 13252-3; Koshkin et al, 1998, Tetrahedron Letters 39:4381-4384; Jumar et al, 1998, Bioorganic & Medicinal Chemistry Letters 8:2219-2222; Singh y Wengel, 1998, Chem Commun, 12:1247-1248; WO 00/56746; El documento WO 02/28875; y, WO 01/48190).

[0052] "Cebadores" se refieren a oligonucleótidos que tienen una secuencia complementaria a una secuencia diana, se hace referencia generalmente como una secuencia de unión a cebador. La parte complementaria de un cebador puede ser de cualquier longitud que soporta la hibridación específica y estable entre el cebador y la secuencia diana en la reacción de condiciones. Los cebadores pueden ser de aproximadamente 5 a 60 nucleótidos de longitud, alrededor de 10 a 35 nucleótidos de longitud, o pueden ser de, y, en particular, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, y/o 20 nucleótidos de largo. En general, los cebadores para la replicación por la polimerasa son capaces de soportar la extensión por la polimerasa cuando el cebador se hibrida con la secuencia diana. "Cebador de amplificación" se refiere a un cebador oligonucleótido usado para la amplificación de una secuencia de ácido nucleico diana.

[0053] "Cebador delantero" y "Cebador trasero" se refieren a un conjunto de cebadores de amplificación, donde un cebador híbrido con el extremo 3' de la diana (cadena molde), mientras que el otro cebador hibrida con el extremo 3' de la cadena diana complementaria para amplificar un amplicón.

[0054] "Proximal" se refiere a la distancia de nucleótidos a partir de una base definido (por ejemplo, una primera mutación de nucleótidos) a una segunda base definida (por ejemplo, una segunda mutación de nucleótidos), donde la primera y segunda mutaciones pueden ser alojadas en un único cebador oligonucleótido usado para los propósitos de amplificación (por ejemplo, cebador delantero o trasero). por lo tanto, en algunas realizaciones, el término "proximal" se determina con respecto a la longitud del cebador. En algunas realizaciones, dos mutaciones pueden ser próxima si están separados por 1, 2, 3, 4, 5, 6, 8, 10, 12, 14, 16, 18, 20, o 25 bases de nucleótidos y están dentro de la imprimación. En algunas realizaciones, la ubicación de las mutaciones con respecto al extremo 3' del cebador es tal que un oligonucleótido hibridado con una hebra molde puede someterse a la extensión por la polimerasa, como se describe en detalle a continuación.

[0055] "Proteína", "polipéptido", "oligopéptido", y "péptido" se utilizan indistintamente para referirse a un polímero de al menos dos aminoácidos unidos covalentemente mediante un enlace amida, independientemente de la longitud o modificación postraducciona (por ejemplo, glicosilación, fosforilación, lipidación, miristilación, ubiquitinación, etc.). se

incluyen dentro de esta definición son los aminoácidos D y L, y las mezclas de aminoácidos D y L.

[0056] "Recombinante" cuando se usa con referencia a, por ejemplo, una célula, ácido nucleico, polipéptido, casete de expresión o vector, se refiere a un material, o un material correspondiente a la forma natural o nativa del material, que ha sido modificado por la introducción de un nuevo resto o alteración de un resto existente mediante técnicas recombinantes, o es idéntica a las mismas pero producidas o derivadas de materiales sintéticos utilizando técnicas recombinantes. por ejemplo, las células recombinantes expresan 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 se expresan de otra manera en un nivel diferente, por lo general, expresado en menor medida o no expresado en absoluto.

[0057] "Secuencia de referencia" se refiere a una secuencia definida usada como base para una comparación de secuencias. Una secuencia de referencia puede ser un subconjunto de una secuencia mayor, por ejemplo, un segmento de un gen de longitud completa o secuencia de polipéptido. En general, una secuencia de referencia es de al menos 20 nucleótidos o aminoácidos residuos de longitud, al menos 25 residuos de longitud, al menos 50 residuos de longitud, o de toda la longitud del ácido nucleico o polipéptido. Dado que dos polinucleótidos o polipéptidos pueden cada uno (1) comprender una secuencia (es decir, una porción de la secuencia completa) que es similar entre las dos secuencias, y (2) puede comprender además una secuencia que es divergente entre las dos secuencias, las comparaciones de secuencias entre dos (o más) polinucleótidos o polipéptidos se realizan típicamente por la comparación de secuencias de los dos polinucleótidos o polipéptidos a través de una "ventana de comparación" para identificar y comparar regiones locales de similitud de secuencia.

[0058] "Replicar" se refiere a la copia de una secuencia de polinucleótido diana para sintetizar una copia complementaria inversa del polinucleótido. Generalmente, la replicación se realiza mediante polimerasas que copian un polinucleótido molde para sintetizar un polinucleótido que es un complemento trasero de la secuencia de polinucleótido diana.

[0059] "Segmento" se refiere a una secuencia que es una porción de una secuencia de polinucleótido más grande. La secuencia de polinucleótido más grande puede ser dividido en una pluralidad de segmentos, en el que la combinación de los segmentos comprende la longitud completa de la secuencia de polinucleótido más grande.

[0060] "Variante de polipéptido" o "análogo de polipéptido" tal como se utiliza aquí, se refiere a polipéptidos que se componen de un segmento que tiene actividad funcional, con o sin la retención de cualquier propiedad mejorada, y tiene una identidad sustancial con una porción de un polipéptido de referencia. En algunas realizaciones, polipéptidos análogos comprenden una sustitución conservadora o no conservadora de aminoácidos, o la adición o delección de uno o más residuos de aminoácidos con respecto a la secuencia de referencia.

[0061] "Watson/Crick base de sincronización" se refiere a un patrón de pares específicos de nucleobases y análogos que se unen juntos a través de enlaces de hidrógeno específicos de la secuencia, por ejemplo, A se aparea con T y U, y G se aparea con C.

[0062] "Sustitución" se refiere a la sustitución de uno o más nucleótidos o aminoácidos por diferentes nucleótidos o aminoácidos, respectivamente, con respecto a una secuencia de referencia, tal como, por ejemplo, una secuencia de tipo silvestre.

[0063] "Sustrato", "Soporte", "Soporte sólido", "Vehículo sólido", o "Resina" son términos intercambiables y se refieren a cualquier material en fase sólida. El sustrato también incluye términos como "fase sólida", "superficie", y/o "membrana". Un soporte sólido puede estar compuesto por polímeros orgánicos tales como poliestireno, polietileno, polipropileno, polifluoroetileno, polietileno y poliacrilamida, así como copolímeros y los injertos de los mismos. Un soporte sólido también puede ser inorgánico, tal como vidrio, sílice, vidrio de poro controlado (CPG), sílice de fase inversa o de metal, como el oro o el platino. La configuración de un sustrato puede estar en forma de perlas, esferas, partículas, gránulos, un gel, una membrana o una superficie. Las superficies pueden ser plana, sustancialmente plana, o no plana. Los soportes sólidos pueden ser porosos o no porosos, y pueden tener hinchazón o características no hinchazón. Un soporte sólido puede estar configurado en forma de un pozo, depresión, u otro recipiente, recipiente, característica o ubicación. Una pluralidad de soportes se puede configurar en una matriz en varios lugares, direccionable para la entrega robótica de reactivos, o por métodos y/o instrumentos de detección.

[0064] "Matriz" se refiere a una disposición de agentes (por ejemplo, proteínas, anticuerpos, paquetes genéticos replicables) sobre un sustrato en lugares posicionalmente distintos. En algunas realizaciones, los agentes de la matriz están espacialmente codificados de tal manera que la identidad de un agente puede determinarse a partir de su ubicación en la matriz. Una "micromatriz" generalmente se refiere a una matriz en la que la detección requiere el uso de detección microscópica para detectar los complejos formados con los agentes sobre el sustrato. Un "lugar" en una matriz se refiere a un área localizada en la superficie de la matriz que incluye agentes, definidos cada uno de modo que se puede distinguir de sitios adyacentes (por ejemplo, ser colocado en la matriz general, o tener alguna característica detectable, que permite la ubicación de distinguirse de otras ubicaciones). La ubicación puede tener cualquier forma conveniente (por ejemplo, circular, rectangular, elíptica o en forma de cuña). El tamaño o área de

una ubicación pueden variar significativamente. Las matrices se pueden construir sobre sustratos, tales como vidrio o plástico de microscopio diapositivas, y configurarse en forma de pozos, depresiones, gotitas, u otros recipientes, o recipientes de reacción, tal como un pocillo de la microplaca. En general, no hay ninguna restricción en el formato de la matriz proporciona los sitios individuales a los que están dispuestos pueden ser localizados e identificados los agentes.

[0065] Cámara de reacción "se entiende que el entorno en el que los agentes y/o de reacción los componentes se lleva a cabo. Recipientes de reacción comercialmente disponibles contienen al menos una cámara de reacción, pero puede contener cámaras de reacción 8, 24, 96 o 384. Para los fines de la presente descripción, "cámara(s) de reacción", "pocillo(s)", "sitio(s) de reacción," se utilizan indistintamente. un ejemplo de una cámara de reacción es uno de los 96 pocillos de microtitulación en una placa de microtitulación de 96 pocillos.

[0066] "Ensayo de cebador" se refiere a la serie de cebadores o conjuntos de cebadores usados en una reacción de amplificación en lugares posicionalmente distintos sobre un sustrato (por ejemplo, sustrato de matriz). En general, el conjunto de cebadores comprende un par de cebadores delantero y traseros utilizados para amplificar un amplicón.

[0067] "Ensayo de amplicón" se refiere a una disposición de polinucleótidos amplificados en lugares posicionalmente distintos sobre un sustrato (por ejemplo, sustrato de matriz). En algunas realizaciones, la matriz de amplicón puede tener la misma disposición posicional como ensayo de cebador, tal como cuando la reacción de amplificación es llevada a cabo en la matriz de cebadores para generar polinucleótidos amplificados.

[0068] "Plásmido", "vector", y "casete" se refieren a un elemento cromosómico adicional, que frecuentemente llevan genes, y por lo general en forma de moléculas de ADN de doble cadena circular. Tales elementos pueden ser secuencias de replicación autónoma, secuencias de integración de genoma, secuencias de fagos o nucleotídicas, lineales o circulares, de un ADN o ARN monocatenario o bicatenario, derivado de cualquier fuente, donde un número de secuencias de nucleótidos se han unido o recombinado en una construcción única que es capaz de introducir un fragmento promotor y la secuencia de ADN para un producto génico seleccionado junto con la secuencia no traducida apropiada 3' en una célula. "Casete de expresión" se refiere a un vector específico que contiene un gen extraño y que tiene elementos además del gen extraño que permiten la expresión de ese gen en un huésped.

4.2 Métodos de síntesis de variantes de polinucleótidos

[0069] La presente descripción proporciona métodos de generación de variantes de polinucleótido que tienen un conjunto definido de la secuencia de diferencias de una secuencia de polinucleótidos de referencia. En algunas realizaciones, los métodos son aplicables para la generación de polinucleótidos que codifican variantes de polipéptidos que tienen diferencias definidas en la secuencia de aminoácidos en comparación con un polipéptido de referencia. En algunas realizaciones, las variantes de polinucleótidos tienen un conjunto definido de diferencias de nucleótidos en las regiones no codificantes, por ejemplo, mutaciones silenciosas. Los polinucleótidos se generan de manera eficiente mediante el uso de bibliotecas de fragmentos de polinucleótidos, donde los miembros de la biblioteca codifican una o más de las diferencias de aminoácidos en comparación con una secuencia de polipéptido de referencia, y los fragmentos de polinucleótidos están diseñados para tener la superposición de las regiones adyacentes de tal manera que la selección de un conjunto apropiado de fragmentos, con y sin mutaciones, permite su montaje en una variante de polinucleótido, tal como un polinucleótido que codifica una variante del polipéptido deseado.

[0070] En algunas realizaciones, el método para generar un polinucleótido que codifica 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 de restos de aminoácidos se define con respecto a una secuencia de aminoácidos de referencia; (b) la definición de segmentos de superposición de una secuencia de polinucleótido que codifica el polipéptido con la diferente secuencia de aminoácidos, u opcionalmente el polipéptido de referencia, con cada segmento está limitada por un conjunto de secuencias de unión de cebadores delanteros y traseros, donde una diferencia en la secuencia de polinucleótido que codifica cada uno de la pluralidad de diferencias de restos de aminoácidos se engloba en las secuencias de unión de cebadores delanteros y/o traseros que se unen a las secuencias de unión de cebadores; (c) amplificar cada segmento con el conjunto de cebadores delanteros y traseros, donde un cebador delantero y/o trasero contiene las diferencias en la secuencia de polinucleótidos, para generar una biblioteca de amplicones que comprenden miembros que codifican las diferencias de aminoácidos definidas y donde la biblioteca comprende miembros suficientes para el montaje de dos o más diferentes permutaciones de secuencias de aminoácidos de las diferencias de residuos de aminoácidos definidas; (d) realizar el montaje de la biblioteca de un conjunto de amplicones que tienen regiones adyacentes complementarias que juntas codifican el polipéptido con una secuencia de aminoácidos de permutación definida, teniendo una o más diferencias de residuos de aminoácidos definidas; y (e) replicar el conjunto de amplicones ensamblados para sintetizar el polinucleótido que codifica el polipéptido. Una biblioteca de amplicones que contienen todas las diferencias de aminoácidos definidas deberían permitir la síntesis de una pluralidad de polinucleótidos que codifican para todas las posibles permutaciones de secuencias de aminoácidos.

[0071] Como será evidente para el experto en la técnica, la selección de la pluralidad de residuos de aminoácidos

definida las diferencias se pueden obtener de diversas fuentes. En algunas realizaciones, las posiciones de residuo de aminoácido y las mutaciones correspondientes para un polipéptido definido se pueden obtener a partir de estudios de mutagénesis al azar, tales como los descritos en Cramer et al, 1998, "Barajado de ADN de una familia de genes de diversas especies acelera la evolución dirigida". Nature 391: 288-291 ; . Cramer et al, 1997, "La evolución molecular de una vía de arseniato de desintoxicación por barajado de ADN," Nature Biotech 15: 436-438; . Zhang et al, 1997, "Evolución dirigida de una fructosidasa efectiva de una galactosidasa mediante barajado de ADN y la detección," Proc Natl Acad Sci EE.UU. 94:45-4-4509; . Cramer y otros, 1996, "Mejora de la proteína verde fluorescente por la evolución molecular mediante barajado de ADN, Nature Biotech 14:315-319; Stemmer, 1994, "La rápida evolución de una proteína *in vitro* mediante la transposición de ADN", Nature 370:389-391; Stemmer , 1994, "Barajado de ADN mediante fragmentación aleatoria y reensamblaje: En la recombinación *in vitro* para la evolución molecular", Proc Natl Acad Sci EE.UU. 91:10747-10751, WO 95/22625; WO 97/0078; WO 97/35966; WO 98/27230; WO 00/42651; WO 01/75767 y la patente de EE.UU.. 6.537.746.

[0072] Típicamente, una biblioteca mutagenizada de polinucleótidos se expresa y los polipéptidos expresados seleccionados para un rasgo propiedad deseada, y las mutaciones asociadas con los cambios en la propiedad deseada identificada. Un gran número de mutaciones que afectan a la función del polipéptido se puede obtener fácilmente utilizando estas técnicas.

[0073] En algunas formas de realización, las opciones de diferencias de restos de aminoácidos se pueden obtener a partir de la comparación de secuencias de aminoácidos de las proteínas relacionadas, tales como las que se encuentran en una base de datos de secuencia. La comparación de secuencias puede identificar posiciones, por ejemplo, residuos conservados que pueden ser importantes para la función de proteínas, que a su vez pueden ser objeto de cambios de aminoácidos definidas. Véase, por ejemplo, Wankhade et al., 2000, J. Biol. Chem. 275 (38): 29701-29708 y Reddy et al, 2001, Proteínas: Estructura, Función y Genética 42:148-163.

[0074] En algunas realizaciones, la pluralidad de diferencias de restos de aminoácidos definidas pueden basarse en diferencias de secuencia que se encuentran en la naturaleza, como los polimorfismos encontrados para un gen particular. En algunos casos, los polimorfismos están asociados con efectos biológicos particulares y los fenotipos asociados. Véase, por ejemplo, Bidwell et al, 1999, Genes y Inmunidad 1:3-19; Chen et al, 2003, Mol. Biol. Evo. 18: 1771-1788. Las colecciones de polimorfismos pueden formar la base para definir una pluralidad de las diferencias de residuos de aminoácidos para la formación de polipéptidos variantes de una secuencia de aminoácidos de referencia. Las combinaciones de diferentes polimorfismos de aminoácidos se pueden utilizar para examinar la función de una proteína particular.

[0075] Cuando la pluralidad de diferencias de restos de aminoácidos se define con respecto a una secuencia de referencia, el polinucleótido que codifica el polipéptido de referencia o de las variantes de polipéptidos pueden ser utilizadas como una base para la identificación de segmentos para la definición de los amplicones para ser generados para crear el fragmento de polinucleótido (es decir, amplificación) de la biblioteca. En algunas realizaciones, la secuencia de polinucleótido no necesita ser restringido a cualquier secuencia particular siempre que codifica, o se puede utilizar como una base para generar polinucleótidos que codifican, la secuencia de aminoácidos de interés. El polinucleótido se puede basar en la (por ejemplo, de tipo silvestre) secuencia de origen natural o una secuencia optimizada para la expresión en un organismo particular de interés (por ejemplo, codones optimizados). Por ejemplo, si el polipéptido de interés se va a expresar en *E. coli.*, Una secuencia de polinucleótido en el que los codones han sido optimizados para la expresión en *E. coli.* puede ser usada. Técnicas de optimización de codones están bien dentro de la experiencia de expertos en la técnica.

[0076] Como será evidente para el experto en la materia, dividir el polinucleótido en segmentos definidos para la amplificación puede llevarse a cabo utilizando técnicas bien conocidas en la técnica. En algunas realizaciones, ya que los segmentos se definen por las secuencias de unión a cebador, que a su vez se utilizan para introducir mutaciones en el amplicón. La división del polinucleótido en segmentos puede tener inicialmente en cuenta la ubicación de las mutaciones en el polinucleótido. Las divisiones del polinucleótido en segmentos también pueden tener en cuenta la longitud total del polinucleótido, la eficiencia de la replicación (por ejemplo, la amplificación de segmentos), y el número deseado de amplicones para el montaje. Otras consideraciones serán evidentes para el experto en la materia.

[0077] Las reacciones de amplificación pueden ser afectadas por la secuencia, el tipo de polimerasa utilizado, la eficiencia de los cebadores, y reacciones secundarias no deseadas (por ejemplo, dímeros de cebadores). Por lo tanto, en algunas realizaciones, dependiendo de la longitud total del polinucleótido a ensamblar, las longitudes de segmento pueden ser de 2000 bases o menos, 1500 bases o menos, 1200 bases o menos, 1000 bases o menos, 900 bases o menos, 800 bases o menos, 700 bases o menos, 600 bases o menos, 500 bases o menos, 400 bases o menos, 300 bases o menos, 250 bases o menos, o 200 bases o menos a aproximadamente 100 o tan pocos como aproximadamente 50 bases de longitud. En general, la duración de los segmentos es de aproximadamente 50 a aproximadamente 1.000 bases, aproximadamente 200 a 1.000 bases, aproximadamente 300 a 700 bases, o aproximadamente 400 a 600 bases, con alrededor de 500 bases siendo longitud media útil dada la eficiencia de las polimerasas utilizadas en las reacciones de amplificación. En diversas formas de realización, los segmentos se solapan de tal manera que los amplicones producidos a partir también tendrán las regiones adyacentes se

superponen (es decir, la superposición de regiones complementarias) para el montaje del polinucleótido.

[0078] En algunas realizaciones, las regiones superpuestas adyacentes deben ser de longitud y complementariedad suficiente para permitir la formación de amplicones hibridados estables (es decir, hibridadas) durante el montaje del polinucleótido. Por lo tanto, en algunas realizaciones, la longitud del solapamiento puede ser de 4 o más nucleótidos, 5 o más nucleótidos, 6 o más nucleótidos, 8 o más nucleótidos, 10 o más nucleótidos, 15 o más nucleótidos, 20 o más nucleótidos, 25 o más nucleótidos, 30 o más nucleótidos, 40 o más nucleótidos, 50 o más nucleótidos, y 100 o menos, 90 o menos, 80 o menos, 70 o menos, 60 o menos nucleótidos de longitud según lo permitido por la capacidad de formar amplicones hibridados estables. Dado que las regiones de superposición generalmente incluyen la unión de secuencias utilizadas para generar los amplicones de imprimación, la longitud de solapamiento pueden dar cuenta de las diferencias en la secuencia del cebador (por ejemplo, delanteros y/o hacia atrás) se utiliza para generar las diferencias de polinucleótidos que codifican la mutación a Ser presentado.

[0079] En algunas formas de realización, los segmentos están delimitados por las secuencias de unión del cebador a la que los cebadores recocidos delanteros / traseros. En su caso, las secuencias de unión de cebadores que definen los segmentos también pueden abarcar la posición del polinucleótido que codifica una diferencia en la secuencia de aminoácidos. La secuencia de unión a cebador puede ser de cualquier longitud suficiente para hibridarse al cebador (adelante o atrás) durante la reacción de amplificación. Por consiguiente, la secuencia de unión a cebador puede ser de 100 bases o menos, 90 bases o menos, 80 bases o menos, 70 bases o menos, 60 bases o menos, 50 bases o menos, 40 bases o menos, 30 bases o menos, 20 bases o menos 15 bases o menos, a aproximadamente 8 bases o 10 bases. En algunas realizaciones, la longitud de las secuencias de unión de cebadores puede comprender de aproximadamente 8 a 50 bases, de aproximadamente 8 a 40 bases, aproximadamente 10 a 30 bases, o alrededor de 15 a 25 bases.

[0080] Cuando el cebador contiene una secuencia que codifica una diferencia de aminoácidos definidos, la mutación puede estar situada en una región del cebador que no interfiera con la extensión del cebador. En algunas formas de realización, la mutación se encuentra en alrededor de la mitad del cebador mutagénico, donde el cebador tiene una Tm que es suficiente para hibridarse con el ácido nucleico molde y servir como cebador para la reacción de extensión de la polimerasa mediada. En algunas realizaciones, las diferencias en la secuencia de polinucleótidos puede ser localizada, en función de la longitud del cebador, aproximadamente 5 bases, 6 bases, 8 bases, 10 bases, 12 bases, 15 bases, 20 bases, 25 bases desde el extremo 3' de el cebador. En consecuencia, en algunas realizaciones, la longitud del cebadores delanteros/traseros pueden ser de aproximadamente 8 a 50 nucleótidos, aproximadamente de 8 a 40 nucleótidos, aproximadamente 10 a 30 nucleótidos, o aproximadamente 15 a 25 nucleótidos, y comprenden además nucleótidos diferencia en la secuencia en alrededor de la mitad de la imprimación. Por lo tanto, en algunas realizaciones los cebadores delanteros/traseros son alrededor de 50 nucleótidos de longitud con una diferencia de nucleótidos aproximadamente 25 nucleótidos a partir del extremo 3', alrededor de 40 nucleótidos de longitud con una diferencia de nucleótidos aproximadamente 20 nucleótidos del extremo 3', alrededor de 30 nucleótidos de longitud con una diferencia de nucleótidos aproximadamente 15 nucleótidos del extremo 3', alrededor de 25 nucleótidos de longitud con una diferencia de nucleótidos sobre 12 nucleótidos desde el extremo 3', o alrededor de 20 nucleótidos de longitud con una diferencia de nucleótidos aproximadamente 10 nucleótidos desde el extremo 3',

[0081] La estabilidad de los cebadores de oligonucleótidos, por ejemplo, la temperatura de fusión térmica, es una función de la fuerza iónica, temperatura, contenido de G / C, y la presencia de agentes caotrópicos y se puede calcular usando métodos conocidos para predecir las temperaturas de fusión (véase, por ejemplo, Baldino y otros, Methods Enzymology 168:761-777; Bolton et al, 1962, Proc Natl Acad Sci EE.UU. 48:1390; Bresslauer et al, 1986, Proc Natl Acad Sci EE.UU. 83:8893-8897; Freier et al, 1986, Proc Natl Acad. Sci EE.UU. 83:9373-9377; Kierzek et al, Biochemistry 25:7840-7846; Rychlik et al, 1990, Ácidos Nucleicos Res 18:6409-6412 (errata, 1991, Ácidos Nucleicos Res 19:698); Sambrook et al, 2001, Molecular Cloning: A Laboratory Manual, 3ª Ed, Cold Spring Harbor Laboratory Press, NY; Suggs et al., 1981, En biología de desarrollo utilizando genes purificados (Brown et al., Eds.), Pp. 683-693,

[0082] Para generar la biblioteca de amplicones, delanteros y cebadores que hibridan con la se utilizan en una reacción de amplificación para generar amplicones de secuencias de cada segmento del polinucleótido de unión a cebador trasero. Cuando el amplicón tiene una diferencia de polinucleótido que codifica un aminoácido definido cambio relativo a la secuencia de referencia, la secuencia de cebadores traseros y/o delanteros se diseñan para introducir la secuencia diferente (es decir, mutación) en la reacción de amplificación. Las combinaciones adecuadas de cebadores delanteros y traseros se utilizan para generar una biblioteca de amplicones que comprenden miembros que pueden codificar para cada una de la pluralidad de diferencias de restos de aminoácidos.

[0083] En algunas realizaciones, los conjuntos de cebadores delanteros y traseros pueden ser almacenados en una matriz, por ejemplo una matriz de imprimación, de manera que se puede acceder fácilmente cuando se necesitan amplicones para la síntesis de un polinucleótido que codifica una secuencia de amino definida de ácido de permutación. Como se apreciará en la técnica, los cebadores de oligonucleótidos se pueden usar para introducir cualquier tipo de mutación seleccionada en la pluralidad definida de diferencias de restos de aminoácidos, incluidos, entre otros, inserciones de aminoácidos, deleciones y sustituciones. Las sustituciones pueden ser mutaciones

conservativas o no conservativas, según lo dictado por la pluralidad escogida de las diferencias de residuos de aminoácidos.

[0084] En algunas realizaciones, más de un ácido amino diferencia en la secuencia puede estar presente en la misma posición de residuos de aminoácidos en una secuencia polipeptídica. En estas realizaciones, diferentes amplicones desde el mismo segmento de solapamiento se pueden generar, en donde cada amplicón se prepara con pares de cebadores delanteros y traseros para cada mutación definida en la posición de residuo idéntico. Para preparar un polinucleótido que codifica una permutación de secuencia particular en esa posición de residuo de aminoácido específico, uno de los amplicones que contienen la mutación deseada (es decir, diferencia de nucleótidos definida) se elige y se monta como un miembro del conjunto de amplicons para generar el polinucleótido que codifica un polipéptido que contiene la mutación deseada en la posición de residuo de aminoácido especificado.

[0085] En algunas realizaciones, más de un par de cebadores (por ejemplo, un conjunto de cebadores degenerados) se puede utilizar para generar un conjunto de amplicones (es decir, fragmentos de polinucleótidos) que pueden utilizarse para ensamblar un conjunto de variantes de polinucleótido de polipéptidos que codifican tienen cambios más residuos de aminoácidos (por ejemplo, sustituciones) en una posición definida específica. Las variantes de polinucleótidos ensamblados a partir de los amplicones realizan cebadores degenerados pueden ser secuenciados antes o después de su polipéptido codificado se ensaya con el fin de determinar la secuencia específica a la posición de interés.

[0086] Como será evidente para el experto en la materia, en algunas realizaciones, un segmento de solapamiento definido para una secuencia de polinucleótidos no puede tener ningunas mutaciones asociadas. Además, el mismo segmento puede en una secuencia de aminoácidos de permutación abarcan una mutación especificada, pero en algunas permutaciones de secuencia pueden no tener ninguna mutación asociada con el segmento. Así, en algunas formas de realización, la biblioteca de los amplicones pueden contener miembros que no tienen las diferencias de secuencia de polinucleótido en comparación con la secuencia de referencia para un segmento particular. Estos polinucleótidos de puente, que no tienen cambios asociados en la secuencia en comparación con la secuencia de referencia, pueden utilizarse como un conector para montar el polinucleótido completo.

[0087] Con la elección apropiada de los segmentos, la biblioteca de amplicón comprende miembros que pueden ser utilizados para ensamblar al menos dos o más diferentes permutaciones de secuencias de aminoácidos de las diferencias de aminoácidos definidos en relación con la secuencia de referencia. Por ejemplo, una pluralidad de mutaciones definidas por las diferencias de restos de aminoácidos A y B puede tener las siguientes permutaciones: A solo, B solo, o A y B. Así, la biblioteca de amplicón tiene miembros suficiente para generar una permutación de secuencia de aminoácidos que tiene, independientemente, A mutación o B mutación. En algunas realizaciones, la biblioteca de amplicón tiene miembros suficientes para generar cada secuencia de aminoácidos de permutación de las diferencias de residuos de aminoácidos definidas en relación con la secuencia de referencia. Así, para el ejemplo dado, la biblioteca de amplicón tiene miembros suficiente para generar permutaciones de secuencias de aminoácidos que tienen, independientemente, una mutación A o mutación B, o una mutación A + B. Dado que el tamaño de los amplicones se corresponderá con el tamaño de los segmentos, los amplicones pueden ser de 2000 bases o menos, 1500 bases o menos, 1200 bases o menos, 1000 bases o menos, 900 bases o menos, 800 bases o menos, 700 bases o menos, 600 bases o menos, 500 bases o menos, 400 bases o menos, 300 bases o menos, 250 bases o menos, o 200 bases o menos a aproximadamente 100 o tan pocos como aproximadamente 50 bases de longitud. En general, la longitud de los amplicones es de aproximadamente 50 a aproximadamente 1.000 bases, aproximadamente 200 a 1000 bases, aproximadamente 300 a 700 bases, o aproximadamente 400 a 600 bases, con alrededor de 500 bases o menos una longitud útil, dada la eficiencia de polimerasas utilizadas en las reacciones de amplificación. En algunas realizaciones, los amplicones son alrededor de 400 bases o menos de longitud.

[0088] En general, la reacción de amplificación puede utilizar cualquier enzima utilizada para la polimerasa mediada por reacciones de extensión, tales como la polimerasa Taq, la polimerasa Pfu, polimerasa Pwo, polimerasa Tfl, polimerasa rTth, polimerasa Tli, polimerasas Tma, y el fragmento Klenow. Condiciones para la amplificación de un segmento de polinucleótido usando la reacción en cadena de la polimerasa puede seguir condiciones estándar conocidas en la técnica. Véase, por ejemplo, Sambrook et al, 2001, Molecular Cloning: A Laboratory Manual, 3ª Ed, Cold Spring Harbor Laboratory Press, NY y Ausubel et al, 1989, Current Protocols in Molecular Biology, Greene Publishing Associates y Wiley Interscience, NY (actualizaciones hasta 2008).

[0089] En algunas realizaciones, la amplificación de cada amplicon puede llevarse a cabo en reacciones separadas, minimizando así la necesidad de aislar un producto de amplificación de otro amplicón. Sin embargo, las reacciones de amplificación de dos o más amplicones pueden llevarse a cabo en una sola reacción y los productos aislados, como por electroforesis o cromatografía. En algunas formas de realización, los productos de la reacción de amplificación pueden ser tratados con varias combinaciones de exonucleasas y fosfatasas para eliminar restantes cebadores y nucleótidos libres (por ejemplo, combinación de exonucleasa I y fosfatasa de alcalina).

[0090] Para generar el polinucleótido que codifica el polipéptido con la secuencia de permutación de aminoácidos definida, se selecciona un conjunto de amplicones que tienen regiones solapadas complementarias y montadas en condiciones que permitan la reasociación de las regiones solapadas complementarias entre sí. Por ejemplo, los

amplicones se pueden desnaturalizar y luego se dejó recocido para formar un complejo de amplicones que juntos codifican el polipéptido con una permutación de secuencia de aminoácidos definida que tiene una o más de las diferencias de residuos de aminoácidos con relación a una secuencia de referencia. Generalmente, el montaje de cada conjunto de amplicones puede realizarse por separado de tal manera que el polinucleótido que codifica una secuencia de aminoácidos de permutación se distingue fácilmente de otro polinucleótido que codifica una secuencia de aminoácidos diferente de permutación. En algunas realizaciones, el montaje puede llevarse a cabo en ubicaciones direccionables sobre un sustrato (por ejemplo, una matriz) de tal manera que una pluralidad de polinucleótidos que codifican una pluralidad de secuencias de permutaciones de aminoácidos definidas pueden ser generadas simultáneamente.

[0091] En algunas formas de realización, los conjuntos se pueden preparar de tal manera que múltiples (es decir, 2 o más) amplicones están representados por el mismo fragmento. El producto resultante de esta reacción de ensamblaje contendrá una mezcla de polinucleótidos que contienen diferentes permutaciones de secuencia de diferencias de aminoácido definidas. Esta mezcla se puede clonar directamente y variantes puede ser secuenciadas antes o después de polipéptidos codificados se ensayan.

[0092] Los amplicones reunidos se replican utilizando una polimerasa para sintetizar el polinucleótido que codifica el 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 reunidos actúan como iniciadores de tal manera que una sola ronda de replicación crea un duplicado de los amplicones reunidos. Generalmente, en la etapa de replicación, cebadores que hibridan con secuencias de unión cebador que flanquean el polinucleótido (es decir, el región 5' terminal y el región 3' terminal) puede ser añadido para amplificar el producto de polinucleótido mediante la realización de reacciones de amplificación adicionales. En algunas realizaciones, estos cebadores flanqueantes pueden incorporar secuencias de reconocimiento para enzimas de restricción para facilitar la clonación del producto polinucleótido sintetizado en plásmidos o vectores, tales como vectores de expresión.

[0093] En algunas realizaciones, los cebadores de flanqueo pueden tener secuencias que permiten la expresión directa *in vitro* utilizando un sistema de traducción acopladas de transcripción para la síntesis del producto de proteína sin la necesidad de transformación en un organismo huésped. Por lo tanto, algunos cebadores que flanquean puede incorporar secuencias de control para controlar la expresión de la región codificante del polipéptido. Las reacciones de amplificación utilizando dichos cebadores flanqueantes pueden ligarse operativamente las secuencias de control para el polipéptido de la región codificante de interés.

[0094] En algunas realizaciones, la pluralidad de diferencias de aminoácidos es al menos 2. En algunas realizaciones, la pluralidad de diferencias de aminoácidos es al menos 2,3,4,5,6, 7, 8, 9, 10, 11, 12, 13, 14, 15, o más. De acuerdo con ello, el número de diferencias de nucleótidos definidas puede variar de 2 a 45, o más. El número de permutaciones para pluralidad "n" diferencias de residuo de aminoácidos definidas es dado por la fórmula $n!/(k!(n-k)!)$, donde n es el número de no mutaciones mutuamente excluyentes y k es el número de diferencias de aminoácidos y n! denota el operador factorial. En algunas realizaciones, el tamaño de la biblioteca de amplificación, por ejemplo durante un mínimo de 2 diferencias de residuos de aminoácidos es un tamaño de la biblioteca que contiene al menos 3 diferentes amplicones. En algunas realizaciones, el tamaño de la biblioteca es de al menos 5, 6, 7, 8, 9, 10, 11, o incluso más amplicones diferentes. Por ejemplo, para una pluralidad de variantes que comprende al menos 10 diferencias definidas, asumiendo ninguna de las diferencias se encuentran madamente tal que más de lo que puede ser incluido por imprimación, el montaje de las variantes con 10 diferencias definidas utiliza hasta 11 amplicones por reacción de ensamblaje. Suponiendo una pluralidad de diferentes mutaciones que se desean en cualquiera de la pluralidad de posiciones que tienen diferencias definidas, las bibliotecas mucho más grandes de los amplicones se pueden utilizar. Por consiguiente, en algunas formas de realización, la biblioteca de los amplicones pueden comprender al menos 5, 10, 20, 30, 40, 50, 75, 100, o más amplicones diferentes.

[0095] Una vez que la biblioteca de los amplicones han sido sintetizados, cualquier polinucleótido que codifica una secuencia de permutación de aminoácidos especificada en base a una pluralidad de diferencia de residuo de aminoácidos se puede hacer uso de la biblioteca de los amplicones. En algunas realizaciones, el método de generación de un polinucleótido que codifica un polipéptido que tiene una secuencia de aminoácidos con una o más diferencias definidas en los residuos de aminoácidos en comparación a una secuencia de polipéptido de referencia puede comprender las etapas de: (a) el montaje de un conjunto de amplicones que tienen regiones adyacentes solapadas complementarias, donde el conjunto montado de los amplicones comprende una secuencia de polinucleótido que codifica una secuencia de aminoácidos con una o más diferencias definidas de residuo de aminoácido en comparación con una secuencia de referencia, en donde se seleccionan los amplicones de una biblioteca de amplicones que tienen miembros que codifican una pluralidad de diferencias de aminoácidos, y (b) reproducir el conjunto de fragmentos de polinucleótidos superpuestos ensamblados a sintetizar el polinucleótido de interés.

[0096] En algunas realizaciones, la biblioteca de amplicón se puede utilizar para generar polinucleótidos que codifican cualquier permutación de una pluralidad definida de diferencias de aminoácidos definidas, comprendiendo el método: (a) la generación de permutaciones de secuencias de aminoácidos que difieren de una secuencia de aminoácidos de referencia basado en una pluralidad de diferencias de restos de aminoácidos definidas en

comparación con una secuencia de aminoácidos de referencia, (b) seleccionar una secuencia de permutación de aminoácidos definidas y determinar una secuencia correspondiente de polinucleótidos basada en una secuencia de referencia, (c) seleccionar un conjunto de la superposición de fragmentos de polinucleótido que codifica las permutaciones de secuencias de aminoácidos definidas, donde al menos cada fragmento de polinucleótido de superposición que codifica una diferencia de un aminoácido es de una pluralidad de fragmentos de codificación de polinucleótido diferente de aminoácidos que conocida las diferencias de residuos, en el que la pluralidad de fragmentos tiene miembros suficientes para ensamblar polinucleótidos que codifican al menos dos diferentes permutaciones de secuencias de aminoácidos, (d) montar el conjunto de fragmentos de polinucleótidos que tiene regiones adyacentes solapadas complementarias, y (e) que replican el conjunto de fragmentos superpuestos montados para sintetizar el polinucleótido que codifica el polipéptido. Para cada secuencia de permutación de aminoácidos deseada, las etapas de (b) a (e) pueden repetirse.

[0097] un proceso ejemplar para la generación de los amplicones para el número "n" de variantes se muestra en la **FIG. 4**. En la realización ilustrada, el procedimiento comprende: (a) la importación de una secuencia de referencia y una lista de mutaciones asociadas con la secuencia, (b) la creación de una lista de variaciones basada en la lista de mutaciones, (c) seleccionar una permutación definida de la secuencia de aminoácidos (es decir, la variante 1), (d) la identificación de la superposición de fragmentos de polinucleótidos de una biblioteca de amplicones (por ejemplo, como la preparada en la **FIG. 5**), (e) determinar el número de variantes y si el número de variantes es menor que el número total de variantes deseadas, reiterando las etapas (a) a (d).

[0098] Para la síntesis eficiente de las bibliotecas de amplificación, cebadores diseñados apropiadamente oligonucleotídicos se utilizan en una reacción de amplificación. En algunas realizaciones, el método de generación de una biblioteca de la superposición de fragmentos de polinucleótidos puede comprender: (a) generar una pluralidad de permutaciones de secuencias de aminoácidos que difieren de una secuencia de aminoácidos de referencia basado en una pluralidad de diferencias de restos de aminoácidos definida a partir de una secuencia de aminoácidos de referencia, y para cada permutación (i) la determinación de una secuencia de polinucleótido que codifica la secuencia de aminoácidos basadas en un polinucleótido de referencia secuencia; (ii) exploración de una secuencia de polinucleótidos y la identificación de un cambio en la secuencia de polinucleótido que codifica una diferencia residuo de aminoácido, y determinar opcionalmente la proximidad de un próximo cambio en la secuencia de polinucleótidos que codifica un siguiente diferencia residuo de aminoácido en la secuencia de aminoácidos de permutación; (iii) seleccionar un oligonucleótido cebador delanteros que tiene una secuencia que codifica la diferencia de un aminoácido, y opcionalmente incluyendo el siguiente cambio en la secuencia de polinucleótido en el mismo cebador directo si próximo a el cambio en la secuencia de polinucleótido; (iv) la exploración de una secuencia de polinucleótidos de la ubicación del cebador delantero hasta que se identifica el siguiente cambio en la secuencia de polinucleótido o hasta que el extremo del polinucleótido, y la selección de un oligonucleótido de cebador trasero para amplificar un fragmento de polinucleótido con el cebador de oligonucleótido directo, en el que el cebador trasero tiene una secuencia que codifica opcionalmente el siguiente cambio en la diferencia de residuo de aminoácido; (v) reiterar las etapas (ii) a (iv) para cada cambio en la secuencia de polinucleótido que codifica una diferencia de residuo de aminoácido hasta que todos los cambios en la secuencia de polinucleótidos que están presentes en los cebadores de oligonucleótidos y los fines de que se alcance la secuencia de polinucleótido; y (g) amplificar con cada conjunto de cebadores delanteros y traseros de oligonucleótidos para generar la biblioteca de amplicones que tienen miembros que codifican las diferencias de aminoácidos superpuestas. En estas realizaciones, cuando la exploración de la secuencia de polinucleótidos se encuentra con el extremo del polinucleótido, cebadores flanqueantes se pueden utilizar en combinación con los cebadores internos para completar la generación de los amplicones.

[0099] Un proceso ejemplar para la selección de cebadores adecuados delanteros y traseros se ilustra en la figura. 5. En la fig. 5, el proceso de selección de los cebadores de oligonucleótidos comprende: (a) seleccionar una variante (un aminoácido de secuencia de permutación) y la generación de la secuencia de polinucleótido correspondiente sobre la base de una secuencia de referencia, (b) la creación de un cebador de oligonucleótidos delanteros para un fragmento con una primera mutación, (c) exploración de la secuencia a partir de la primera mutación a la siguiente mutación o hasta el final del gen y la creación de un cebador de oligonucleótido trasero para la próxima mutación, (d) y si el siguiente mutación es próxima a la primera mutación, colocando la siguiente mutación en el mismo cebador de oligonucleótidos delanteros, (e) reiterar las etapas (b) a (d) hasta que los extremos de la variante de polinucleótido n se alcance.

[0100] Como se ha señalado anteriormente, en algunas realizaciones en las que el polinucleótido ha separado en segmentos superpuestos definidos por un conjunto de cebadores delanteros y traseros, cebadores delanteros y traseros pueden no tener mutaciones asociadas. Un contexto en el que esto puede ocurrir es que si los segmentos de polinucleótidos deben ser restringidos en tamaño, por ejemplo, aproximadamente menos de 1000 bases, debido a la necesidad para la síntesis eficiente de un amplicón, tales cambios que no todos los segmentos han definido en la secuencia de polinucleótido. En algunas formas de realización, en la preparación de los oligonucleótidos basados en el método anterior, la búsqueda de la secuencia se puede limitar a un tamaño particular "l", por ejemplo en alrededor de 1.200 bases en la etapa (iv) para la selección de un cebador trasero. En otras palabras, tras la identificación de un cebador delantero basado en una diferencia en la secuencia, una exploración se realiza en uno o el otro sentido de la secuencia de polinucleótidos para determinar la distancia de nucleótidos a la siguiente

mutación. Si la distancia excede el límite establecido, un segmento que no abarca ningunas mutaciones puede ser creadas para tender un puente sobre dos segmentos que contienen las dos mutaciones distantes. El proceso de digitalización puede ser reiterado en el punto de la siguiente mutación.

[0101] Como se señaló anteriormente, los cebadores de oligonucleótidos, ya sea solos o en conjunto (por ejemplo, delanteros y traseros) oligonucleótidos, así como los amplicones correspondientes se pueden colocar sobre sustratos direccionables para la automatización y/o almacenamiento. Los cebadores de oligonucleótidos en los sustratos direccionables, también descritos en el presente documento como un ensayo de cebador, pueden ser accedidos robóticamente para sintetizar bibliotecas de amplicones para una pluralidad definida de diferencias de aminoácidos. Del mismo modo, los amplicones en los sustratos direccionables, también aquí se describen como matrices de amplicón, se puede acceder para generar una secuencia de polinucleótidos que codifica un aminoácido deseado de secuencia de permutación de ácido basado en la pluralidad definida de diferencias de residuos de aminoácidos. Un sustrato o soporte sólido para la matriz pueden estar compuestas de polímeros orgánicos tales como poliestireno, polietileno, polipropileno, polifluoroetileno, polietileno y poliacrilamida, así como copolímeros e injertos de los mismos. Un soporte sólido también puede ser inorgánico, tal como vidrio, sílice, vidrio de poro controlado (CPG), sílice de fase inversa o de metal, como el oro o el platino. La configuración de un sustrato puede estar en forma de perlas, 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 hinchazón o inflamación no características. Un soporte sólido puede estar configurado en forma de un pozo, depresión, u otro recipiente, recipiente, característica, o la ubicación. Una pluralidad de soportes pueden configurarse en una matriz en varios lugares, direccionable para la entrega robótica de reactivos, o por métodos y/o instrumentos de detección, en algunas realizaciones, el sustrato es una cámara de reacción. Comercialmente recipientes de reacción disponibles contienen al menos una cámara de reacción, pero pueden contener cámaras de reacción 8, 24, 96 o 384. Un ejemplo de una cámara de reacción es uno de los 96 pocillos de microtitulación en una placa de microtitulación de 96 pocillos.

[0102] En algunas realizaciones, un sistema robótico y un sistema de ordenador asociado capaz de cebadores de muestreo o pares de cebadores a partir de las matrices se pueden utilizar para entregarlos a una cámara de reacción. Los reactivos para la amplificación de la polimerasa mediada también pueden ser entregada a cada conjunto de cebadores en la cámara de reacción, seguida de la aplicación de una rutina de amplificación (por ejemplo, en un termociclador automatizado). Esto permite la formación de un sustrato que contiene amplicones direccionables definidos basados en segmentos de una secuencia de polinucleótidos que se superponen. El sistema robótico puede elegir el conjunto apropiado de los amplicones en base a la permutación deseada de la secuencia de aminoácidos, los cebadores flanqueantes para la amplificación del producto de polinucleótido final, y entregar los reactivos para la reacción de ensamblaje y amplificación. Un sistema robótico ejemplar se proporciona en la **FIG. 6**. El sistema robótico en la **FIG. 6** comprende instrucciones para (a) seleccionar un segmento y amplicón asociado para la amplificación, (b) la identificación de oligonucleótidos delanteros y traseros para el fragmento seleccionado (es decir, amplificación), el almacenamiento de la información de datos en los oligonucleótidos en la lista de oligonucleótidos únicos (por ejemplo, 96 placa de microtitulación de pozo), y la colocación de los oligonucleótidos en un primer sustrato direccionable (c) almacenar la información de datos en fragmento sintetizado (por ejemplo, la posición en la matriz, la secuencia, los oligonucleótidos utilizados, etc) a la lista de fragmentos únicos, y colocando el oligonucleótido en un sustrato segundo direccionable, (d) determinar el número de fragmentos seleccionados contra el número total de fragmentos necesarios para el montaje, y reiterando las etapas (a) a (d) hasta que todos los fragmentos se han seleccionado, (e) colocar el gen de montaje en un tercer sustrato direccionable y pasos reiterando (a) a (d) hasta que todas las variantes deseadas que se han generado.

[0103] En algunas formas de realización, la presente descripción también proporciona bibliotecas de fragmentos de polinucleótidos (es decir, amplicones) para el montaje de una pluralidad de polinucleótidos que codifican diferentes secuencia de permutaciones de aminoácidos. En algunas realizaciones, la pluralidad de polinucleótidos comprende: fragmentos de polinucleótidos con la superposición de las regiones adyacentes, cada fragmento de polinucleótido estando limitada por las secuencias de unión del cebador para los cebadores delanteros y traseros, en el que la pluralidad de polinucleótidos tienen miembros que codifican en las secuencias de una diferencia de residuo de aminoácido específica de unión a partir de una pluralidad definida de las diferencias de residuos de aminoácidos con respecto a una secuencia de aminoácidos de referencia de cebador de tal modo que la pluralidad de fragmentos de polinucleótido codifica toda una pluralidad seleccionada de restos de aminoácidos de diferencias de pluralidad de diferencias definidas de residuos de aminoácidos; y en el que la pluralidad de fragmento de polinucleótido comprende miembros para el montaje de dos o más diferentes permutaciones de secuencias de aminoácidos de las diferencias de aminoácidos definidas. En algunas realizaciones, la pluralidad de fragmentos de polinucleótido comprende miembros suficientes para el montaje de todas las posibles permutaciones de la secuencia de aminoácidos de la pluralidad seleccionada de las diferencias de residuos de aminoácidos. En algunas realizaciones, los miembros de la pluralidad son amplicones formadas utilizando los cebadores delanteros y traseros.

[0104] Como será evidente para el experto artesano, los métodos descritos en el presente documento se puede practicar usando técnicas estándar disponibles para el experto en la técnica, tales como los descritos en Sambrook et al, 2001, Molecular Cloning: A Laboratory Manual, 3ª Ed. , Cold Spring Harbor Laboratory Press, Nueva York y Ausubel et al., 1989, Current Protocols in Molecular Biology, Greene Publishing Associates y Wiley Interscience,

Nueva York (actualizaciones hasta 2008). Los oligonucleótidos se pueden sintetizar utilizando metodologías químicas conocidas, tales como los basados en metodologías de síntesis en fase sólida de fosoramidita (Véase, por ejemplo, Wright, et al, 1993, Tetrahedron Letters 34, 3373-3376; Caruthers, 1991, Acc. Chem. Res. 24, 278-284; y las referencias citadas en este documento).

[0105] También se describe aquí sistemas informáticos implementados en forma de software de ordenador para llevar a cabo los métodos descritos anteriormente. En algunas realizaciones, el producto de programa de ordenador comprende un medio de almacenamiento legible por máquina que tiene instrucciones de programa que comprenden códigos para cada una de las etapas de: (a) la importación de una secuencia de referencia y una lista de mutaciones asociadas con la secuencia, (b) lista de permutaciones creadas basadas en la lista de mutaciones, (c) seleccionar una permutación definida de la secuencia de aminoácidos, (d) la identificación de la superposición de fragmentos de polinucleótidos de una biblioteca de amplicones (por ejemplo, como se preparó en la **FIG. 5**), (e) determinar el número de variantes y si el número de variantes es menor que el número total de variantes deseadas, reiterando las etapas (a) a (d).

[0106] En algunas realizaciones, el producto de programa de ordenador comprende un medio de almacenamiento legible por máquina que tiene instrucciones de programa que comprenden códigos para cada una de las etapas de: (a) seleccionar una variante (un aminoácido de secuencia de permutación) y la generación de la secuencia de polinucleótido correspondiente basado en una referencia secuencia, (b) la creación de un cebador de oligonucleótido delantero para un fragmento con una primera mutación, (c) exploración de la secuencia a partir de la primera mutación a la siguiente mutación o hasta el final del gen y la creación de un cebador trasero de oligonucleótidos para la próxima mutación, (d) y si el siguiente mutación es próxima a la primera mutación, la colocación de la siguiente mutación en el mismo oligonucleótido delantero, (e) las medidas que reiteran (b) a (d) hasta que extremos de la variante de polinucleótido n se alcanza.

[0107] Como se muestra en las ilustraciones de la **FIG. 4**, **FIG. 5**, y **FIG. 6**, el equipo llevó a cabo programas para la selección de los amplicones, la selección de cebadores de oligonucleótidos, y el almacenamiento en formatos direccionables pueden ser integrados para permitir la automatización de las diversas etapas de los métodos de la descripción.

[0108] Como se describe en el presente documento, en algunas realizaciones, el método se puede utilizar para sintetizar polinucleótidos que codifican polipéptidos que tienen un conjunto definido de mutaciones seleccionadas de entre una pluralidad de diferencias definidas en los residuos de aminoácidos de una secuencia de referencia. Los métodos de la presente memoria permiten la síntesis eficiente de varias permutaciones de secuencias de aminoácidos sobre la base de las diferencias de residuos de aminoácidos. Síntesis eficiente de los polinucleótidos que codifican diferentes secuencias de permutaciones de aminoácidos es útil para una variedad de aplicaciones de ingeniería de proteínas. Véase, por ejemplo, la publicación de solicitud de EE.UU. US20060195947; La publicación de solicitud de Estados Unidos. US20050153417; y la patente de Estados Unidos N° 7.220.566. En algunas realizaciones, los métodos se pueden utilizar para sintetizar polinucleótidos que codifican enzimas variantes que tienen propiedades mejoradas sobre la base de un conjunto de mutaciones que se sabe afectan diferentes propiedades de la enzima. Por ejemplo, algunas mutaciones pueden afectar, entre otros, la actividad enzimática, estabilidad térmica, especificidad de sustrato, estereoselectividad, estereospecificidad, y refractariedad a la inhibición del producto. Mientras que las técnicas tradicionales de mutagénesis aleatoria y técnicas de evolución de proteínas pueden conducir a la identificación de mutaciones que afectan a estas diversas propiedades de las enzimas, muchas de estas mutaciones pueden ocurrir independientemente de los otros. Utilizando los métodos en el presente documento, varias permutaciones de mutaciones que afectan a diferentes rutas, tales como la actividad enzimática, especificidad de sustrato, estabilidad térmica y se pueden hacer y se cribaron para identificar enzimas de ingeniería que tienen múltiples alterado rasgos deseados.

[0109] Los métodos proporcionados en este documento proporcionan eficiencia y precisión sorprendente en la generación de grandes bibliotecas de variantes de polinucleótido que comprende varias permutaciones de la secuencia de cambios. Por ejemplo, la secuencia de la proteína para la gulonolactona (L-) oxidasa (GLO) de *Railus norvegicus* (adhesión: gi-92090602-sp-P10867,3-GGLO_RAT) se puede traducir al revés para proporcionar una secuencia de ADN de 1,3 kb que se puede utilizar como una plantilla para diseñar 90 variantes de polinucleótido, cada uno codifica una variante de polipéptido que tiene una combinación diferente de tres a cinco sustituciones de aminoácidos. Por ejemplo, una lista de 90 permutaciones de sustituciones de 3-5 aminoácidos se puede seleccionar de la siguiente lista de 10 posibles sustituciones: T28S, D95A, S156N, G175S, R212D, I251E, F302S, H330I, Y370G y K423N. Las 90 diferentes permutaciones de sustituciones de aminoácidos codificadas por las variantes de polinucleótidos fueron las siguientes: D95A/F302S/H330I/K423N; D95A/F302S/Y370G; 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/F302; 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/F302/Y370G/K423N; G175S/H330I/K423N; G175S/I251E/F302; G175S/R212D/H330I; G175S/R212D/K125E/H330I; G175S/R212D/K423N; S175S/R212D/Y370G; G175S/R212D/Y370G/K423N;

H3301/Y370G/K423N; 1251E/H3301/Y3700; 1251E/H3301/Y370G; 1251E/Y370G/K423N;
R212D/F302S/Y370G/K423N; R212D/H3301/K423N; R212D/1251E/F302S; R212D/1251E/F302S/H3301;
R212D/I251E/Y370G; R212D/I251E/Y370G; S156N/F302S/H3301; S156N/F302S/K423N; S156N/F302S/Y370G;
S156N/G175S/F302S/Y370G; S156N/G175S/I251E/F302S; S156N/G175S/K423N; S156N/G175S/K423N;
5 S156N/G175S/R212D/F302S/H3301; S156N/1251E/F302S/H3301; S156N/I251E/H3301/Y370G;
S156N/I251E/H3301/Y3700/K423N; S156N/I251E/Y370G; S156N/R212D/F302S/H3301/Y370G;
S156N/R212D/K423N; T28S/D95A/G175S/F302; T28S/D95A/G175S/F302S/Y3700; T28S/D95A/H3301;
T28S/D95A/I251E; T28S/D95A/I251E/F302/K423N; T28S/D95A/R212D; T28S/D95A/S156N/H3301/Y370G;
T28S/D95A/S156N/R212D; T28S/D95A/S156N/R212D; T28S/D95A/S156N/R212D/Y370G; T28S/D95A/Y370G;
10 T28S/D95A/Y3700/K423N; T28S/F302S/K423N; T28S/G175S/H3301; T28S/G175S/H3301/Y3700;
T28S/G175S/I251E/F302; T28S/G175S/I251E/F302S/Y370G; T28S/G175S/I251E/H3301;
T28S/G175S/I251E/K423N; T28S/H3301/K423N; T28S/I251E/F302S/H3301/K423N; T28S/R212D/F302S/H3301;
T28S/R212D/H3301; T28S/R212D/I251E/F302; T28S/R212D/I251E/Y370G/K423N; T28S/R212D/Y370G;
T28S/S156N/F302S/H3301/Y370G; T28S/S156N/F302S/Y3700; T28S/S156N/F302S/Y370G; T28S/S156N/G175S;
15 T28S/S156N/G175S; T28S/S156N/G175S/I251E; T28S/S156N/G175S/I251E/K423N;
T28S/S156N/R212D/1251E/H3301; T28S/S156N/R212D/I251E/K423N; y T28S/S156N/R212D/K423N.

[0110] Software (por ejemplo, como se describe en las figuras 5-7) se puede utilizar para determinar un total de sólo 55 amplicones, que corresponden a fragmentos de variantes de polinucleótido con regiones de secuencia de la superposición, pueden ser utilizados para montar las 90 variantes de polinucleótidos en una reacción de Ronda 2 SOE-PCR. El software también se puede utilizar para determinar que se necesita un total de sólo 22 cebadores de oligonucleótidos en 55 reacciones separadas Ronda 1 de PCR con el polinucleótido de referencia 1,3 kb como molde para generar los necesarios 55 amplicones. Los cebadores de oligonucleótidos 22 son sólo 30 o 33 nucleótidos de longitud, e incluyen cebadores mutagénicos que comprenden cambios de nucleótidos en el medio de la secuencia (por ejemplo, en los nucleótidos 15-17).

[0111] Por lo tanto, de acuerdo con los métodos descritos en este documento, la construcción de las 90 diferentes variantes de polinucleótidos requiere síntesis de sólo 22 oligonucleótidos relativamente cortos (30-mer a 33-mer), una primera reacción de Redonda 1 PCR para crear los 55 amplicones (es decir, la variante de fragmentos de polinucleótido), y una segunda reacción de la Ronda 2 SOE-PCR en la que los 55 amplicones se agrupan en diversas combinaciones (con cebadores flanqueantes delanteros y traseros) para permitir el montaje de SOE-PCR de las 90 variantes de polinucleótidos. En la preparación de las reacciones de Ronda 2 de PCR-SOE, cada uno de los 55 amplicones pueden ser reutilizados en promedio 7,8 veces, con ciertos fragmentos utilizados sólo una vez o dos veces, y otros utilizados hasta 36 veces.

[0112] El flujo de trabajo de las reacciones Ronda 1 y Ronda 2 y pueden ser controladas por listas de trabajo generadas por software (por ejemplo, como en Figs. 4 y 6) que se utilizan para ejecutar la robótica Tecan para el manejo de líquidos. Las listas de trabajo de esta construcción de biblioteca ilustrativa llamada 90 variantes para sólo 110 operaciones de manejo de líquidos para las reacciones de Ronda 1 de PCR para generar los 55 amplicones utilizando 22 cebadores, y sólo 430 operaciones de manejo de líquidos para las reacciones de Ronda 2 de ensamblaje SOE-PCR para preparar 90 variantes completas de longitud de polinucleótidos de 55 amplicones.

[0113] La precisión de las secuencias de variantes de polinucleótidos proporcionadas por los métodos descritos en el presente documento se puede determinar mediante etapas adicionales de clonación y secuenciación de cada una de la pluralidad de construcciones de las reacciones de Ronda 2. Como se ilustra por los Ejemplos (más abajo), los métodos descritos en el presente documento como resultado sorprendentemente de alto nivel de las secuencias correctas (secuencias perfectas de larga duración de (FLP))... es decir secuencias que tienen el nucleótido deseadas relativas al polinucleótido de referencia.

[0114] Al menos algunas de las sorprendentes ventajas de los métodos descritos en este documento están en la mayor precisión de las grandes bibliotecas de variantes de polinucleótidos producidos. En algunas realizaciones, los métodos se pueden utilizar para preparar una biblioteca direccionable de al menos 10 variantes de polinucleótidos diferentes, comprendiendo cada una al menos una diferencia en la secuencia definida con respecto a una secuencia de polinucleótidos de referencia, en el que al menos una media de 75% de las secuencias variantes de polinucleótidos son secuencias correctas (por ejemplo, las secuencias que comprenden la secuencia de referencia de longitud completa con las diferencias de nucleótidos definidas introducidas por los cebadores utilizados en el método). En algunas realizaciones, los métodos proporcionan una biblioteca direccionable de al menos 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200 o más diferentes variantes de polinucleótidos, que comprende cada uno al menos una diferencia secuencia definida relativa a una secuencia de polinucleótidos de referencia, donde al menos 75%, 80%, 85%, 90%, 95% o más de las secuencias variantes de polinucleótidos son correctas, por ejemplo, FLP por análisis de secuencia.

[0115] En ciertas realizaciones, los métodos descritos en el presente documento comprenden una pluralidad de reacciones de Ronda 1 de PCR utilizando una plantilla de polinucleótido de referencia y una pluralidad de Ronda 2 SOE-PCR reacciones de montaje amplicón se puede utilizar para preparar una biblioteca direccionable de 10 o más variantes de polinucleótidos de referencia de polinucleótido de al menos 500 pb, 750 pb, 1000 pb, 1250 pb, 1500 pb

o más, cada variante que comprende de aproximadamente 1-30, 1-25, 1-20, 1-15, 3-30, 3-20, o 3-15 nucleótidos cambios relativos con el polinucleótido de referencia, en el que las reacciones Ronda 1 de PCR comprenden aproximadamente 6-300, 6-200, 6-100, 6-50, 6-40, 6-30, 6-25, 6-20, 6-15, o tan sólo 6-10 cebadores diferentes de oligonucleótidos, y al menos 75%, 80%, 85%, 90%, 95% o más de las secuencias de variantes de polinucleótido son de longitud perfecta completa.

[0116] En algunas realizaciones, varias permutaciones de fragmentos de polinucleótidos (por ejemplo, seleccionado de una biblioteca direccionable) se puede montar en una biblioteca direccionable de variantes de polinucleótido cada codifica una variante diferente de polipéptido que tiene una diferencia definida de residuos de aminoácido. Cada una de estas variantes de polinucleótidos a continuación se puede clonar en un sistema de expresión para generar una biblioteca de clones direccionable, cada uno capaz de generar una variante de polipéptido diferente. Esta biblioteca de clones direccionable se puede transformar en células (*por ejemplo, E. coli*), para la traducción, y de las placas automatizadas y recogiendo de colonias (es decir, transformantes viables). La secuenciación entonces se puede ganar a cabo para confirmar la combinación de mutaciones en cada secuencia de variante de polipéptido así generada. Ensayo (por ejemplo, a través del examen de alto rendimiento) de las variantes de polipéptidos para los rasgos alterados deseados puede llevarse a cabo en todas las variantes de polipéptidos, u opcionalmente sólo en aquellas variantes de polipéptidos confirmadas por secuenciación que tienen la combinación deseada de mutaciones.

[0117] Alternativamente, la biblioteca direccionable de la variante de polinucleótido que codifican cada uno un polipéptido de variante diferente que tiene una diferencia de residuos de aminoácidos definida se puede combinar (por ejemplo, agrupado) y se clonó en un sistema de expresión, creando de este modo una biblioteca agrupada de clones. Esta biblioteca combinada de los clones puede ser transformada (por ejemplo, en una única etapa de transformación) en células para la traducción, enchapada y recogida de colonias (es decir, transformantes viables). Ensayo de las colonias de esta biblioteca agrupada de clones puede llevarse a cabo (por ejemplo, mediante examen de alto rendimiento) antes de la secuenciación para identificar polinucleótidos que codifican variantes de polipéptidos que tienen los rasgos alterados deseados. Una vez que tal "éxito" se identifica por un rasgo alterado, se puede secuenciar para determinar la combinación específica de las mutaciones presentes en la secuencia variante de polinucleótido. Opcionalmente, las variantes que codifican polipéptidos que tienen los rasgos no alterados deseados buscados en el ensayo no necesitan ser secuenciados. De acuerdo con ello, la biblioteca combinada de método de clones puede dar más eficiencia requiriendo sólo una sola transformación en lugar de un conjunto de reacciones de transformación paralelas.

[0118] Análogamente, el método puede también ser usado para generar varias permutaciones de combinaciones de mutación para examinar las características estructurales de las proteínas biológicamente importantes. Por ejemplo, receptores implicados en la transducción de señales de moléculas extracelulares actúan a través de la interacción con otros receptores, así como diversas proteínas intracelulares. Estas interacciones complejas pueden afectar a diferentes procesos de señalización celulares del mismo tipo de molécula de receptor. Tanto la señalización negativa y positiva puede ser iniciada por el mismo receptor. Un ejemplo específico es receptores de proteína de acoplamiento G, que interactúan con proteínas, $\beta\gamma$ Gs α , y Gi α . Véase, por ejemplo, las madres et al., 1999, Physiol. Rev. 79:1373-1430. Debido a que las mutaciones en diferentes dominios del receptor pueden tener diferentes efectos, los métodos de la presente memoria proporcionan un procedimiento eficaz para la generación de diferentes permutaciones de combinaciones de mutaciones que se sabe afectan diferentes aspectos de la función del receptor, lo que permite estudios sobre las funciones biológicas y estructurales asociadas de la proteína de interés.

[0119] Si bien los métodos para la generación de diferentes permutaciones de una secuencia de polinucleótidos se han ilustrado para la generación de polinucleótidos que codifican varias permutaciones de un polipéptido a partir de un conjunto definido de diferencias de residuos de aminoácidos, es de entenderse que los métodos pueden adaptarse generalmente para generar permutaciones de secuencias de polinucleótidos. Por ejemplo, los métodos de este documento pueden usarse para generar diferentes permutaciones de polinucleótidos funcionales, tales como los genes de ARNs ribosomales. Varios ARNs forman complejos de nucleoproteína que participan en la síntesis de proteínas en procariotas y eucariotas. Muchos antibióticos funcionan mediante la interrupción de la función de ribosoma y se sabe que interactúan con regiones definidas de ARNs determinadas. Varias mutaciones se han identificado que afectan a la síntesis de proteínas y estas regiones se correlaciona con los sitios de interacción con antibióticos. Véase, por ejemplo, Yassin et al., 2005, Proc Natl Acad Sci. EE.UU. 102(46):16620-16625.

[0120] Utilizando los métodos descritos en el presente documento, varias permutaciones de mutaciones conocidas que afectan a la función del ARN ribosomal pueden ser sintetizados y los efectos de cierta combinación de mutación examinada. Otras aplicaciones serán evidentes para el experto en la materia.

5. EJEMPLOS

Ejemplo 1: Preparación de amplicones

[0121] Preparación de oligonucleótidos. Cebadores de oligonucleótidos a una concentración 200uM se diluyen a 4uM con agua estéril en una placa Axygen HalfDeep 96 (1,1 ml). Para la mayoría de las posiciones en la placa de

microtitulación, la adición de 10 µl de oligo a 490 µl dH₂O será suficiente. Para los puestos de A01 y D01 en la placa de oligonucleótido, los cebadores comunes delanteros y traseros, un mayor volumen puede ser necesario. Los volúmenes máximos de aspiración en el informe de salida en la sección de volúmenes aspirados y dispensados se verifica antes de la siguiente etapa.

[0122] Ronda 1 - Formación de Amplicones por PCR. Tecan robótico se utilizan para alícuota de 5 µl de cada cabador delantero y trasero de oligonucleótido en placas BioRad HardShellPCR96 (salida del script Tecan) y añadir 40 µl de la mezcla maestra. Realizar la Ronda 1 PCR y verificar la amplificación utilizando un 2% de 96 pocillos e gel. Reactivo para la PCR es el siguiente: 5 µl 10x tampón Herculanase, 1 µl 40 mM dNTPs, 1 µl AL de 100 ng/µl de plantilla de SOE, 2,5 unidades de la polimerasa de Herculanase (Stratagene, LaJolla, CA, EE.UU.). PCR se lleva a cabo de la siguiente manera: 2 min de desnaturalización a 95 ° C seguido de un ciclo a 95 ° C durante 30 s, 56 ° C durante 30 s, 72 ° C durante 1 min/Kb. El número de ciclos es 17.

[0123] Tratamiento con ExoSAP-it. Para la Ronda 1 de PCR, 25 µl del producto de reacción se transfiere a una placa fresca 96, y 2 µl ExoSAP-It (USB Corp., Ohio, EE.UU.) más 0,5 µl DpnI se añade y el ciclaje se lleva a cabo (transferencia manual de 37 ° C 1 hora 80 ° C 15 mm). Las muestras se diluyeron a un volumen final de 100 µl por la adición de 73 µl dH₂O y se agruparon en otra placa BioRad HardShellPCR96 PCR utilizando una secuencia de comandos Tecan.

Ejemplo 2: Montaje de amplicones y análisis de los productos

[0124] Ronda 2 - Asamblea y SOE-PCR La robótica Tecan se utiliza para alícuotar un conjunto de fragmentos de 15 µl (es decir, amplicones de Ronda 1) en placas BioRad HardShellPCR96 (secuencia de comandos Tecan) y añadir 35 µl de mezcla maestra (5 µl de 10X Tampon de Herculanase, 1 µl de 40 mM dNTPs, 0,2 µl de cebadores directos, 0,2 µl de cebadores delanteros, 2,5 unidades de enzima de Herculanase y 28.1 µl dH₂O). La realización de PCR y verificar la amplificación usando un e-gel de 2% 96. PCR se lleva a cabo como sigue: 2 min de desnaturalización a 95 ° C después de un ciclo a 95 ° C durante 30 s, 56 ° C durante 30 s, 72 ° C durante 1 min / Kb El número de ciclos es 17.

[0125] Purificación de placa de 96 pocillos. La purificación de todas las muestras usando Zymo ZR-96 PCR limpiar placas de 96 pocillos (protocolo del fabricante con las siguientes modificaciones) (Investigación Zymo, CA, EE.UU.). La realización de todos los pasos de centrifugación a 2800 rpm durante 10 minutos. Para eluir el ADN, se aplican 25 µl dH₂O de temperatura de 55 ° C directamente a la membrana de sílice, centrifugado durante 10 minutos y derogación. Usando este método 48-50 µl de producto se recupera.

[0126] Enzima de digestión de restricción BglI. Transferencia 30 µl. de cada inserción purificada es una placa de PCR medio skirt fresca, añadir 20 µl de BglI maestro de mezcla de digestión (5 µl 10X Tampon NEB 3, 20 unidades de BglI (New England Biolabs, MA, EE.UU.), que 13uL dH₂O) a todas las muestras e incubar a 37 ° C durante 4 horas. La digestión Bgl es para la clonación en BglI sitio de un vector de expresión.

[0127] Purificación de 96 pocillos. Purificar todas las muestras usando Zymo ZR-96 PCR limpiar placas de 96 pocillos (protocolo del fabricante con las siguientes modificaciones). Realizar todos los pasos de centrifugación a 2800 rpm durante 10 minutos. Para eluir el ADN, aplique 25 µl H₂O destilada a la temperatura de 55 ° C directamente a la membrana de sílice, giro de 10 mm, y repetir. Usando este método, se recupera 48 50 µl de producto. Verificar la recuperación de producto usando un gel e de 2% de 96.

[0128] Ligaduras a vector de expresión. Transferencia 3 µl de cada inserción purificada a una placa fresca y añadir 27 µl de mezcla maestra de la ligadura de las muestras. Se incubaron durante 14 horas a 16 ° C seguido de 15 minutos en 65 ° C, seguido de retención 8 ° C. Mezcla de ligadura: 3 µl 10X Tampón de Ligasa (New England Biolabs, MA, EE.UU.), 1 µl BglI vector digerido (50 ng/µl), 400 unidades de ligasa T4 (New England Biolabs, MA, EE.UU.), 22 µl dH₂O.

[0129] Transformación-HTP. Transferencia 2uL de cada reacción de ligación a 20uL de células químicamente competentes TSS e incubar en hielo en un bloque de metal durante al menos 15 mm. choque térmico a 42 ° C durante 35 segundos y volver al bloque de metal de 2 min. Añadir 80 µl de 37 ° C medios SOC a cada muestra. Incubar a 37 ° C durante 1 hora antes del enchapado.

[0130] Enchapado: Enchapar la mezcla de transformación de 48 bandejas Q bien divididas utilizando el Tecan. Dispensar tres 5 mm grano a cada pocillo de la bandeja Q usando un dispensador de cuentas. Utilice la dispensa Tecan para 40 µl por pocillo de mezcla de transformación. Cultiva los transformantes durante la noche a 37 ° C.

[0131] Recolección y el cultivo para placas de secuencia verificada, solicitar dos colonias por bandeja Q bien recogidas a dos placas de fondo plano Nunc separadas que contienen LB, CAM, y 1% de glucosa. Para placas variantes no secuencia verificadas, solicitan que tres colonias por bandeja Q se recogerán por separado a tres placas de fondo plano Nunc que contienen LB, CAM, y 1% de glucosa.

[0132] PCR de colonias Para llevar a cabo la PCR de colonias en una de las dos placas maestras replicadas, añadir 2uL de cultivo a la colonia estándar de PCR Master Mix y llevar a cabo la PCR de colonias. ExoSAP se utiliza para limpiar el producto del PCT de la siguiente manera: Transferencia 5 µl de la muestra de PCR a una nueva placa de PCR que contiene 2 µl ExoSap-it. Incubar a 37 ° C durante 15 min y 80 ° C por 15 min. Diluir las muestras a un volumen final de 40 µl mediante la adición de 33 µl de dH₂O.

[0133] Secuenciación de productos PCR Añadir 4 µl del cebador de secuenciación 1mM a la placa de secuenciación. Añadir 4 µl de limpiado de muestra de PCR.

Ejemplo 3: Generación de un conjunto de 190 variantes de polinucleótido diferentes cada una codificando de un polipéptido que tiene un solo cambio de aminoácido con relación a un polipéptido de referencia

[0134] Diseño experimental: Se seleccionó un polinucleótido de referencia de 1.359 pb (que codifica una enzima de 453 aminoácidos). Un total de 190 diferencias de aminoácidos de restos de ácido a partir de la secuencia de referencia fueron seleccionados en base a la secuencia de cambios observados en las enzimas homólogas. Las 190 variantes se hicieron como polinucleótidos individuales que codifican las 190 proteínas diferentes que se expresan y se ensayaron. Cada una de las 190 variantes de polinucleótidos se ensambló mediante la combinación de dos amplicones que comprenden el cambio de codón único deseado en su zona de solapamiento (como se preparó en Ronda 1 a continuación) en una reacción de SOE (Ronda 2 a continuación).

[0135] Preparación de oligonucleótidos: Un total de 382 cebadores de oligonucleótidos para la PCR se diseñaron y se sintetiza de acuerdo con métodos estándar. Los oligonucleótidos fueron generalmente 31 nucleótidos (nt) de longitud con el cambio deseado para el codón de interés situado en el centro (- base 15) del cebador de oligonucleótido. Todos los oligonucleótidos se diluyeron a 4 pm con agua estéril en una placa Axygen HalfDeep 96 (1,1 ml).

[0136] Ronda 1 - Formación de Amplicones por PCR: Cada amplicón, correspondiente a un fragmento de variante de polinucleótidos, se generó en una reacción PCR utilizando un vector que comprende 1359 nt de polinucleótidos de referencia como molde y utilizando un cebador mutagénico en combinación con un común cebador flanqueante (recocido al vector de aguas arriba o aguas abajo del gen que no contiene la mutación). Reacción de flujo de trabajo, las condiciones y la limpieza eran como se describe en el Ejemplo 1.

[0137] Ronda 2 - Amplicones de Ensamblaje. Amplicones purificados a partir de Ronda 1 se agruparon de tal manera que 2 amplicones con regiones superpuestas que comprende el cambio de la secuencia de codones deseada para cada una de las 190 variantes de polinucleótidos y cada grupo se dividió en alícuotas en el pocillo de una placa de 96 pocillos como se describe en los Ejemplos 1 y 2. Cebadores flanqueantes comunes delanteros y traseros se añadieron a cada conjunto y PCR se llevó a cabo como se describe en el Ejemplo 2 que resulta en el montaje de los amplicones (i.e., fragmentos de polinucleótidos) para formar la variante de polinucleótido de longitud completa. Inspección de las reacciones de ensamblaje se realizó en un gel de agarosa y se encontró que contienen el producto del tamaño esperado. (Ver por ejemplo, Fig. 3.)

[0138] Análisis de Secuencia de Variantes de Polinucleótidos. Después de la purificación mediante Xymo ZR-96PCR limpiar, los productos se clonaron en un vector de expresión, y cada ligación se transformó en células huésped de E. coli. Dos colonias de cada transformación se recogieron y se preparó ADN plasmídica para la secuenciación de ADN. Una muestra de cada transformación se secuenció usando cebadores de secuenciación internos para y flanqueando el gen. De las 190 variantes de polinucleótidos, 160 (84%) mostraron tener secuencia de longitud perfecta completa (FLP) con los cambios de secuencia de codones deseados. Al secuenciar la segunda preparación de plásmido de las 30 variantes que eran incorrectas, se identificaron 25 secuencias correctas adicionales. Con ello, el porcentaje de éxito global de las secuencias correctas 97%. (185 de los 190 polinucleótidos deseados se identificaron). Los polinucleótidos se expresaron y se ensayaron los variantes de polipéptidos.

Ejemplo 4: Generación de un conjunto de 96 diferentes variantes de polinucleótidos que codifican cada uno un polipéptido, teniendo cambio de uno de tres aminoácidos en relación a un polipéptido de referencia.

[0139] Diseño Experimental: Se seleccionó un polinucleótido de referencia de 1359 nt. 96 variantes se diseñaron, que contienen cada uno tres mutaciones relativas con la secuencia de referencia. Las 96 variantes se hicieron como polinucleótidos individuales de cada uno de codificación de 96 proteínas diferentes que se expresaron y se ensayaron. Cada una de las variantes de polinucleótidos 96 se ensambló mediante la combinación de cuatro amplicones que comprende los cambios de codón deseados en su zona de solapamiento (como se preparó en Ronda 1 a continuación) en una reacción de SOE (Ronda 2 a continuación).

[0140] Preparación de Oligonucleótido: Un total de 130 cebadores de oligonucleótidos fueron diseñados y sintetizados de acuerdo con métodos estándar. Los oligonucleótidos fueron generalmente 31 nt de longitud con el cambio deseado para el codón de interés en el medio (- base 15) del oligonucleótido. Todos los oligonucleótidos se diluyeron a 4µM con agua estéril en una placa Axygen HalfDeep 96 (1.1ml).

[0141] Ronda 1 - Formación de los Amplicones por PCR: Cada amplicón, correspondiente a un fragmento de variante de polinucleótido, se generó en una reacción PCR utilizando un vector que comprende 1359 nt referencian polinucleótido como molde y utilizando un cebador mutagénico en combinación con otro cebador mutagénico o un cebador flanqueante común (recocido al vector de aguas arriba o aguas abajo del gen que no contiene la mutación). Reacción de flujo de trabajo, las condiciones y la limpieza eran como se describe en Ejemplo 1.

[0142] Ronda 2 - Amplicones de Ensamblaje. Amplicones purificados de Ronda 1 se agruparon de tal manera que 4 amplicones con regiones superpuestas que comprende la secuencia de codones deseada cambia para cada una de las variantes de polinucleótidos 96 y cada grupo se dividió se alícuotaron en el pocillo de una placa de 96 pocillos como se describe en Ejemplos 1 y 2. Cebadores comunes de flanqueo delanteros y traseros se añadieron a cada conjunto y PCR se llevó a cabo como se describe en el Ejemplo 2 que resulta en el montaje de los amplicones (es decir, fragmentos de polinucleótidos) para formar la variante de polinucleótido de longitud completa. Inspección de las reacciones de ensamblaje se realizó en un gel de agarosa y se encontró que contienen el producto del tamaño esperado. (Véase, por ejemplo, Fig. 3.)

[0143] Análisis de secuencias de producto de Polinucleótido. Después de la purificación mediante PCR Xymo ZR-96 limpiar, los productos se clonaron en un vector de expresión, y cada ligación se transformó en células huésped de E. coli. Dos colonias de cada transformación se recogieron y se preparó ADN plasmídica para la secuenciación de ADN. Una muestra de cada transformación se secuenció usando cebadores de secuenciación interna para y flanquear el gen. De las variantes de polinucleótidos 96, 82 (85%) fueron determinadas a tener sólo la secuencia correcta de FLP con los cambios deseados.

Ejemplo 5: Generación de un conjunto de 96 variantes diferentes de polinucleótidos cada una codificando de un polipéptido que tiene de uno a seis cambios de aminoácidos con relación a un polipéptido de referencia.

[0144] Diseño experimental: Se seleccionó una referencia de polinucleótidos de 1,056 nt. 96 variantes se diseñaron, cada uno que contiene de uno (1) a seis (6) mutaciones de la secuencia de referencia. Las 96 variantes se hicieron como polinucleótidos individuales cada una codificando una de las 96 proteínas diferentes que se expresa y se ensayaron. Cada una de las variantes de polinucleótidos 96 se ensambló mediante la combinación de dos (por ejemplo, para la variante que codifica un único cambio de aminoácido) a siete (por ejemplo, para la variante de codificación de seis cambios de aminoácido) amplicones que comprenden los cambios de codón deseados en su zona de solapamiento (como se preparó en Ronda 1 a continuación) en una reacción de SOE (Ronda 2 a continuación).

[0145] Preparación de olionucleótido: Un total de 108 cebadores de oligonucleótidos fueron diseñados y sintetiza de acuerdo con métodos estándar. Los oligonucleótidos fueron generalmente 31 nt de longitud con el cambio deseado para el codón de interés en el medio (- base 15) del oligonucleótido. Si dos cambios de aminoácidos estaban muy juntos, un oligo más largo se diseñó para codificar para que los cambios se introduzcan. Todos los oligonucleótidos se diluyeron a 4µM con agua estéril en una placa de Axygen HalfDeep 96 (1,1 ml).

[0146] Ronda 1 - Formación de los amplicones por PCR: Cada amplicón, correspondiente a un fragmento de variante de polinucleótido, se generó en una reacción PCR utilizando un vector que comprende 1056 nt de referencia de polinucleótido como molde y utilizando un cebador mutagénico en combinación con otro cebador mutagénico o un flanqueo común cebador (hibridación al vector de aguas arriba o aguas abajo del gen que no contiene la mutación). flujo de trabajo de reacción, condiciones y limpieza fueron como se describe en el Ejemplo 1.

[0147] Ronda 2 - Asamblea de Amplicones. Amplicones purificados agrupados (a partir de 2 a 7 amplicones/polinucleótidos) se dividen en partes alícuotas en placas como se describe en el Ejemplo 1. Cebadores delanteros y traseros comunes flanqueados se añadieron y la PCR se llevó a cabo como se describe en el Ejemplo 1.

[0148] Análisis de Secuencia de Producto de Polinucleótido. Después de la purificación mediante Xymo ZR-96 PCR limpiar, los productos se clonaron en un vector de expresión, y cada ligación se transformó en células huésped de E. coli. Dos colonias de cada transformación se recogieron y se preparó ADN plasmídica para la secuenciación de ADN. Una muestra de cada transformación se secuenció usando cebadores de secuenciación internos para y que flanquean el gen. Como se muestra en la Tabla 1 (a continuación), de las variantes de polinucleótidos 96, 72 (75%) mostraron tener sólo la secuencia correcta FLP con cambios deseados de 2-7 codones.

Tabla 1: 96 Variantes Construidas - 88 variantes requeridas para hacer la placa final

	<u>Conjunto de secuenciación</u>	<u>1(n = 96)</u>	<u>2(n = 96)</u>	<u>3(n = 16)</u>
5	Total correcto	72	84	92
	Secuencias fallidas	5	3	2
10	Contaminación cruzada	5	2	0
	Mutación de codificación	8	2	0
15	Indeles	6	4	2

Reivindicaciones

1. Un procedimiento de síntesis de una pluralidad de variantes de polinucleótido teniendo cada uno al menos una diferencia de un nucleótido definido con respecto a una secuencia de polinucleótido de referencia, comprendiendo el método:

(a) amplificar por separado una plantilla de polinucleótido de referencia con cada uno de una pluralidad de pares de cebadores delanteros y traseros, en el que la pluralidad de pares de cebadores delanteros y traseros comprende la pluralidad de diferencias de nucleótidos definidas y en las que cada par genera un amplicón que comprende una secuencia capaz de unirse a una secuencia de solapamiento adyacente de al menos otro amplicón, generando de ese modo una biblioteca direccionable de amplicones en los que:

(i) la biblioteca de amplicones comprende miembros para el montaje de dos o más diferentes variantes de polinucleótido que comprende diferencias definidas de nucleótidos con respecto a una secuencia de polinucleótido de referencia; y

(ii) los miembros de la biblioteca están dispuestos en lugares posicionalmente distintos sobre un sustrato;

(b) ensamblar por separado una pluralidad de conjuntos de amplicones, en los que cada conjunto comprende amplicones que tienen secuencias adyacentes superpuestas capaces de unirse para formar la longitud completa de la secuencia de polinucleótidos de referencia;

(c) la replicación de la pluralidad de conjuntos de amplicones reunidos, generando de ese modo una biblioteca de variantes de polinucleótidos direccionables en los que los miembros de la biblioteca están dispuestos en ubicaciones posicionalmente distintas en un sustrato.

2. El método de la reclamación 1, en la que el polinucleótido de referencia codifica un polipéptido de referencia y cada uno de la pluralidad de variantes de polinucleótido codifica un polipéptido que tiene al menos una diferencia de secuencia de aminoácido.

3. El método de la reclamación 2, en la que el método comprende además la etapa de clonación de cada una de la pluralidad de variantes de polinucleótido en un vector de expresión.

4. El método de la reclamación 3, en la que el método comprende además la transformación de células con los vectores de expresión.

5. El método de la reclamación 4, en la que el método comprende además la detección de las células transformadas para la actividad de los polipéptidos codificados por las variantes de polinucleótidos.

6. El método de la reclamación 5, en la que el método comprende además el aislamiento de al menos un polipéptido codificado por las variantes de polinucleótidos.

7. El método de la reclamación 1, en la que:

(i) la pluralidad de variantes de polinucleótidos comprende al menos 10 variantes de polinucleótidos diferentes, o la pluralidad de variantes de polinucleótidos comprende al menos 40 variantes de polinucleótidos diferentes;

(ii) cada una de la pluralidad de variantes de polinucleótidos comprende al menos 3 diferencias de nucleótidos definidos en relación con la secuencia de referencia de polinucleótido;

(iii) al menos una de la pluralidad de variantes de polinucleótidos comprende al menos 5, 6, 9, 12, 15, 18, 21, 24, 27 o 30 diferencias de nucleótidos definidos en relación con la secuencia de polinucleótidos de referencia;

(iv) al menos una de la pluralidad de conjuntos de amplicones comprende al menos 3 amplicones diferentes, o al menos uno de la pluralidad de conjuntos de amplicones comprende al menos 5 diferentes amplicones;

(v) al menos uno de la pluralidad de conjuntos de amplicones comprende un polinucleótido puente que comprende ninguna diferencia de nucleótidos con respecto a la secuencia de polinucleótidos de referencia;

(vi) las longitudes de los cebadores delanteros y traseros son de aproximadamente 20 a 50 nucleótidos, o las longitudes de los cebadores delanteros y traseros son de aproximadamente 25 a 35 nucleótidos;

(vii) la longitud de la secuencia de polinucleótidos de referencia es de al menos 500 pb, 750 pb, 1000 pb, 1250 bp, o la longitud de la secuencia de polinucleótidos de referencia es de al menos 1500 pb; y/o

(viii) la pluralidad de pares de cebadores delanteros y traseros comprende de aproximadamente 6 a 50 oligonucleótidos diferentes, o la pluralidad de pares de cebadores delanteros y traseros comprende de aproximadamente 6 a 25 oligonucleótidos diferentes.

8. El método de la reclamación 1, en la que las secuencias de la pluralidad de secuencias de cebadores delanteros y traseros son generadas por:

(i) la identificación de una primera diferencia definida en la secuencia de variante de polinucleótido en comparación con la secuencia de referencia y la determinación de la proximidad de un vecino más cercano define diferencia en la secuencia de polinucleótido;

(ii) la selección de un cebador delantero que tiene una secuencia que comprende la primera diferencia de nucleótidos definida, y, opcionalmente, incluyendo cualquier diferencia definida de vecino más cercano en el mismo cebador delantero si la proximidad de la primera diferencia definida de nucleótidos;

(iii) la identificación de una próxima diferencia definida en la secuencia de variante de polinucleótido en comparación con la secuencia de referencia, y la determinación de la proximidad de una diferencia definida de vecino más próxima en la secuencia de polinucleótidos, o la identificación de que el final de la variante de polinucleótido se ha alcanzado;

(iv) la selección de un cebador trasero que tiene una secuencia que comprende la siguiente diferencia de nucleótidos definida, y, opcionalmente, incluyendo cualquier diferencia de vecino más cercana definida en el mismo cebador trasero si la proximidad de la siguiente diferencia de nucleótidos definida; y

(v) repetir las etapas (iii) a (iv) para cada diferencia definida en la secuencia de variante de polinucleótido tal que toda diferencia definida está presente en los cebadores.

9. El método de la reclamación 8, en la que el método comprende además la selección de cebadores de oligonucleótidos delanteros y traseros no mutagénicos para segmentos de polinucleótido no definidos por cebadores delanteros y traseros de (ii) y (iv).

10. El método de la reclamación 1, en la que la amplificación y replicación es por reacción en cadena de polimerasa.

11. Una biblioteca direccionable de amplicones que comprenden una pluralidad de amplicones, en el que cada miembro de la biblioteca de los amplicones comprende al menos una diferencia de nucleótidos definida en relación con una secuencia de polinucleótidos de referencia y una región adyacente superpuesta capaz de unirse a la región adyacente superpuestas de al menos otro amplicón en la biblioteca, en el que la pluralidad de amplicones comprende una pluralidad de conjuntos de amplicones capaz de unirse para formar la longitud completa de la secuencia de polinucleótidos de referencia, en la que la pluralidad de amplicones comprende miembros para el montaje de dos o más diferentes variantes de polinucleótidos de longitud completa comprende diferencias de nucleótidos definidas en relación con la secuencia de polinucleótidos de referencia, en donde:

(i) al menos uno de la pluralidad de conjuntos de amplicones comprende miembros para el montaje de una variante de polinucleótido que comprende al menos 5 diferencias de nucleótidos definidas relativas a la secuencia de polinucleótidos de referencia;

(ii) al menos uno de la pluralidad de conjuntos de amplicones comprende al menos 3 diferentes amplicones; o

(iii) la longitud de la secuencia de polinucleótidos de referencia es de al menos 500bp, y en el que los miembros de la biblioteca están dispuestos en ubicaciones posicionalmente distintas en un sustrato.

12. La biblioteca de la reclamación 11 direccionable, en la que la secuencia de polinucleótido de referencia codifica un polipéptido de referencia y la pluralidad de amplicones comprende miembros suficientes para el montaje de todas las diferencias posibles de nucleótidos que codifica una pluralidad seleccionada de las diferencias de residuo de aminoácidos.

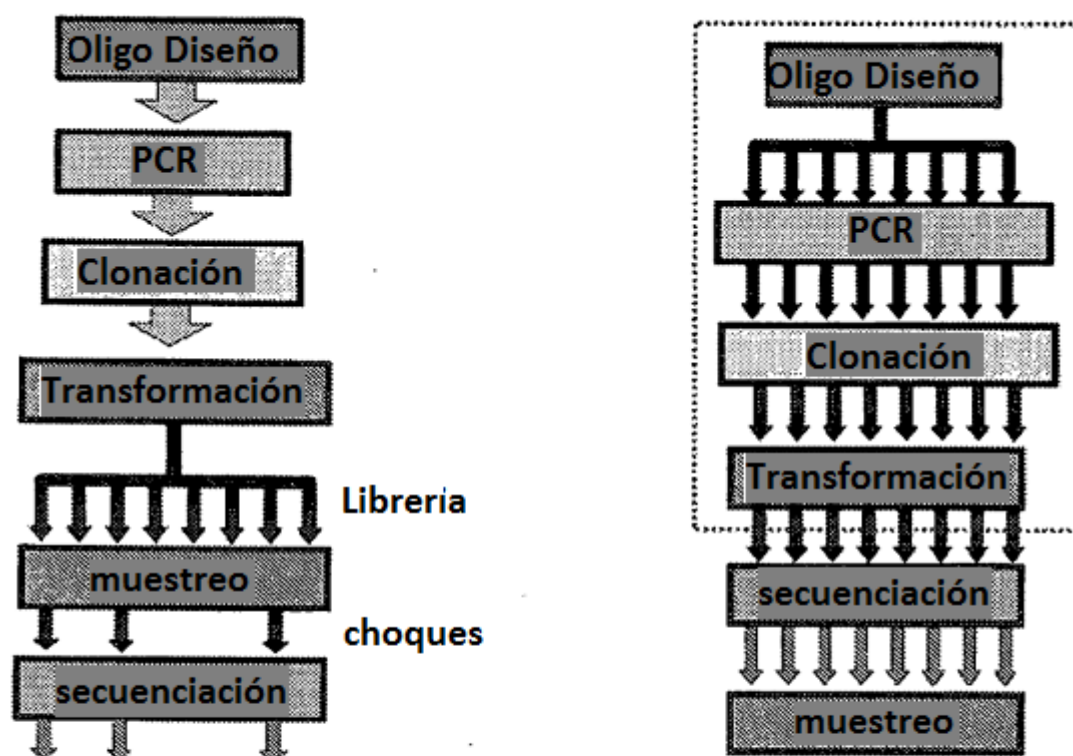


FIG. 1

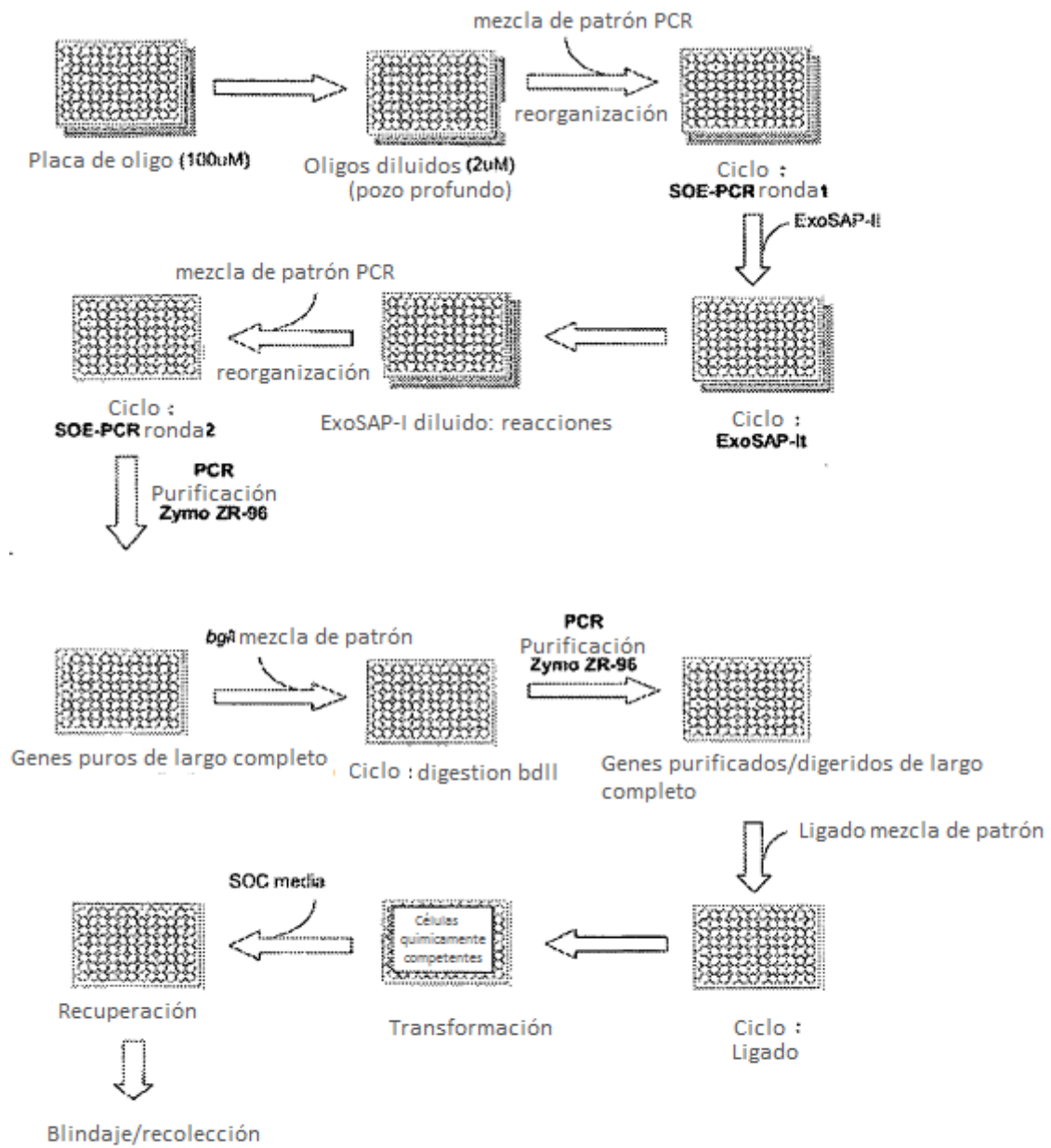


FIG. 2

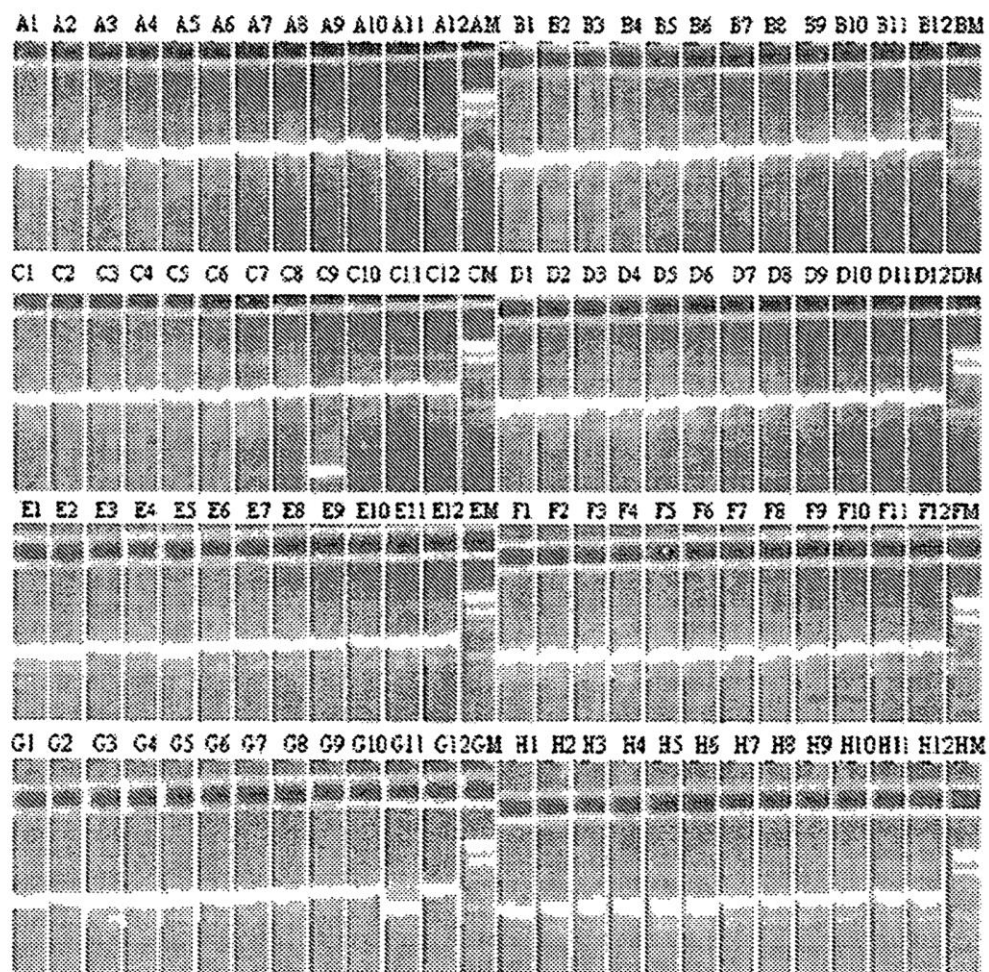


FIG. 3

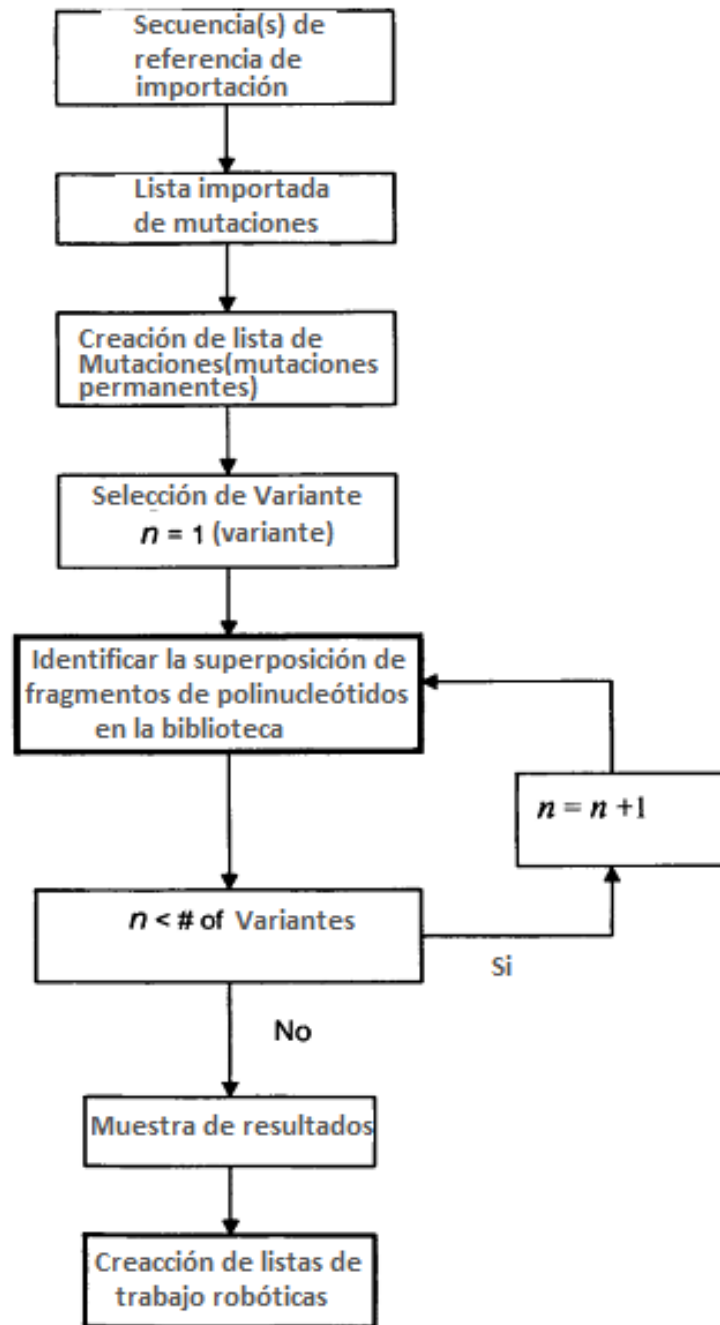


FIG. 4

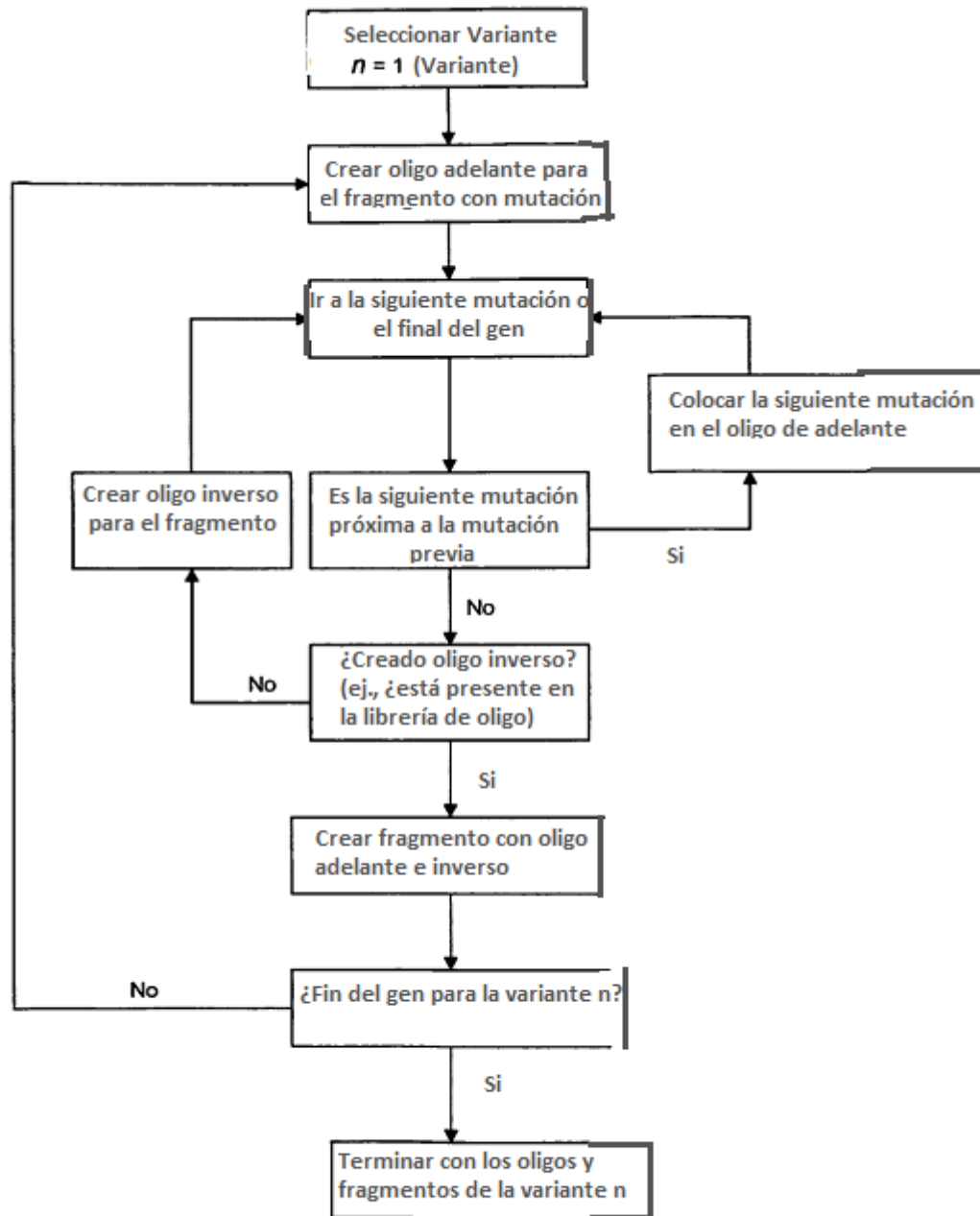


FIG. 5

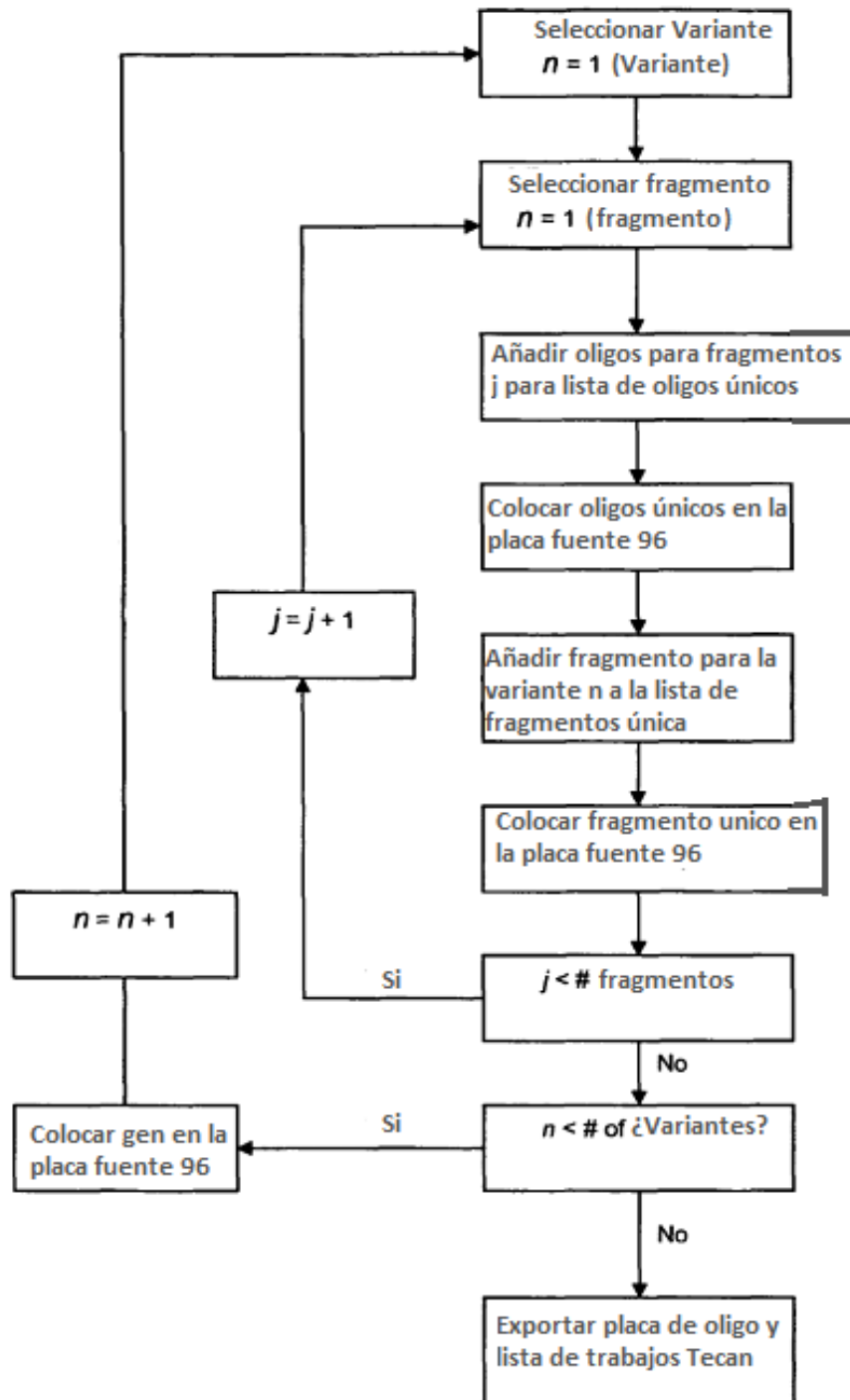


FIG. 6