

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 813 825**

51 Int. Cl.:

**C12Q 1/6869** (2008.01)

**G16B 30/00** (2009.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **27.04.2015 PCT/GB2015/051215**

87 Fecha y número de publicación internacional: **29.10.2015 WO15162438**

96 Fecha de presentación y número de la solicitud europea: **27.04.2015 E 15719281 (6)**

97 Fecha y número de publicación de la concesión europea: **03.06.2020 EP 3134542**

54 Título: **Métodos de secuenciación**

30 Prioridad:

**25.04.2014 GB 201407334**

**02.03.2015 GB 201503465**

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

**25.03.2021**

73 Titular/es:

**DNAE GROUP HOLDINGS LIMITED (100.0%)**

**Ugli Campus Block C, 56 Wood Lane**

**London W12 7SB, GB**

72 Inventor/es:

**MORLEY, DANIEL**

74 Agente/Representante:

**FLORES DREOSTI, Lucas**

ES 2 813 825 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

## DESCRIPCIÓN

Métodos de secuenciación

## CAMPO DE LA INVENCION

5 La presente invención se refiere a métodos para secuenciar cadenas de polinucleótidos mediante el uso de una secuenciación por método de síntesis; y en particular a métodos que tienen orden de flujo de nucleótidos optimizado o mejorado.

## ANTECEDENTES DE LA INVENCION

10 La secuenciación por métodos de síntesis se usa comúnmente en tecnologías de secuenciación de última generación (NGS, por sus siglas en inglés). Las cadenas de nucleótidos complementarias a un fragmento polinucleótido objetivo se extienden por incorporación de nucleótidos (p. ej. dNTP) por una enzima polimerasa, y se detecta la incorporación; por ejemplo, por fluorescencia o por la detección de iones de hidrógeno liberados durante la polimerización. Esta última técnica se usa en métodos de secuenciación por semiconducción de iones. La incorporación de un dNTP determinado en una cadena significa que el nucleótido complementario está presente en esa posición en la cadena molde.

15 En algunas técnicas, a nucleótidos diferentes se les dan etiquetas detectables diferentes, de modo que se puede determinar el nucleótido específico incorporado. Sin embargo, un enfoque alternativo es simplemente agregar un único tipo de nucleótido a la vez a la reacción de polimerasa; si se detecta la incorporación del nucleótido, entonces se conoce el nucleótido complementario en la cadena molde. Típicamente una reacción de secuenciación pasará a través de todos los cuatro nucleótidos en orden, y lo repetirá mientras dure la secuenciación. Sin embargo, esto impone limitaciones de tiempo al proceso, dado que es necesario repetir el ciclo varias veces para obtener la secuencia, y dependiendo del orden de los nucleótidos en la cadena molde, pueden ser necesarios hasta cuatro flujos de nucleótidos par obtener información de una sola base.

20 Sin embargo, para varias aplicaciones de tecnología de secuenciación por síntesis, se conoce la secuencia esperada del molde, o al menos se conoce parcialmente. Por ejemplo, se puede analizar la muestra de un paciente en busca de la presencia de un patógeno sospechado, mediante la cual se detecta un diagnóstico de secuencia para un patógeno dado. En este ejemplo, la secuencia a detectar ya se conoce. De manera alternativa, por ejemplo, las variantes en determinadas secuencias génicas se pueden detectar para determinar la presencia o ausencia de un determinado polimorfismo o mutación. Aquí, nuevamente, al menos una parte de la secuencia se conoce. En determinadas aplicaciones de la secuenciación por síntesis, se preparan fragmentos polinucleótidos para la secuenciación mediante la ligación o incorporación de otro modo de adaptadores de una secuencia conocida, a la cual se pueden unir cebadores de secuenciación. Al menos esta parte de la reacción de secuenciación se puede beneficiar del conocimiento de la región a secuenciar.

30 US2014/0031238 y WO2011/156707 describen el uso de un flujo de nucleótidos alternativo que no es simplemente una repetición continua de todos los cuatro nucleótidos. Se dice que este ordenamiento alternativo aborda problemas potenciales con pérdida de sincronización física que resulta de una extensión incompleta. No hay nada que sugiera que el orden del flujo de nucleótidos se pueda modificar mediante el aprovechamiento de la existencia de secuencias conocidas.

Ahmadian et al, "Single Nucleotide Polymorphism Analysis by Pyrosequencing", Analytical Biochemistry, vol 280, 1 enero 2000, páginas 103-110, describe la secuenciación mediante síntesis para distinguir SNP en moldes de secuencia. Se usan órdenes de distribución de nucleótidos modificados para optimizar la secuenciación.

40 Ramon et al, "Pyrosequencing: A one-step method for high resolution HLA typing", Journal of Translational Medicine, vol 1, n.º 1, 26 noviembre 2003, página 9, describe el diseño de órdenes de distribución de nucleótidos para distinguir entre moldes de secuencia HLA heterocigota de otro modo ambiguos.

45 Binladen et al, "The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing", PLOS One, vol 2, n.º 2, 1 enero 2007, páginas e197.1-e197.9, describe el uso de etiquetas de cebadores de PCR para generar amplicones de múltiples moldes, que luego se secuencian; la parte de la secuencia que deriva de las etiquetas permite la asignación de una secuencia particular a un origen determinado.

US2013/178374 describe un método para variar la distribución de nucleótidos en una reacción de pirosecuenciación en la cual una secuencia de novo se compara con una base de datos de posibles secuencias y el siguiente nucleótido se selecciona en función de la base más probable para esa posición.

5 Sería ventajoso proporcionar una secuenciación mediante un método de síntesis mediante el cual el orden del flujo de nucleótidos se puede mejorar u optimizar. En determinadas realizaciones esto se logra mediante el uso de conocimiento a priori de la secuencia a detectar. En otras realizaciones, se pueden seleccionar secuencias candidatas probables, y determinar el flujo de nucleótidos en función de la probabilidad de que determinadas secuencias estén presentes. En incluso otras realizaciones, se puede utilizar un mecanismo de retroalimentación para modificar el flujo de nucleótidos durante la secuenciación por síntesis.

#### BREVE COMPENDIO DE LA INVENCION

Según un primer aspecto de la presente invención, se proporciona un método para secuenciar una cadena de polinucleótidos mediante la secuenciación por síntesis, donde el método comprende las etapas de:

- 10 a) proporcionar en una reacción una cadena de polinucleótidos a secuenciar, un cebador, y una enzima polimerasa;
- b) proporcionar a la reacción nucleótidos en un orden predeterminado, en donde el orden predeterminado se selecciona para correlacionarse con una secuencia prevista para la cadena de polinucleótidos;
- c) controlar la reacción para detectar la incorporación de un nucleótido en una cadena de polinucleótidos sintetizados;

15 en donde, en el caso de que se detecte la incorporación de nucleótidos, se procede a proporcionar el siguiente nucleótido en el orden predeterminado, y en donde en el caso de que no se detecte la incorporación de nucleótidos, se revisa la secuencia prevista para la cadena de polinucleótidos y se selecciona un nuevo orden predeterminado de nucleótidos, en donde el nuevo orden predeterminado se selecciona para correlacionarse con la secuencia prevista revisada.

20 Como alternativa a revisar la secuencia prevista y continuar con la síntesis, simplemente se puede detener la reacción. Esto puede ser deseable cuando se usa la secuenciación para determinar la presencia de una secuencia prevista específica (por lo tanto, si no está presente, la reacción puede detenerse), o donde se han probado varias secuencias previstas candidatas probables. Incluso una alternativa adicional para detener la reacción sería seleccionar un nuevo orden predeterminado de nucleótidos que represente ciclos múltiples de los cuatro nucleótidos (por ejemplo, A, C, G, T, repetidos). Esto puede ser ventajoso cuando se han probado varias secuencias previstas candidatas probables y la secuencia real todavía se desconoce; por tanto, la secuenciación puede volver a un ciclo repetitivo simple de los cuatro

25 nucleótidos.

30 Por lo tanto, en el presente método, el orden en el que se proporcionan los nucleótidos a la reacción se selecciona basándose en el conocimiento del polinucleótido a secuenciar. Si la secuencia prevista es correcta, entonces se producirá la síntesis de nucleótidos cada vez que se proporcione un nucleótido (permitiendo errores de síntesis y asumiendo que el orden predeterminado es perfectamente complementario a la secuencia prevista). La detección de la incorporación de nucleótidos confirma que la secuencia prevista es correcta y se proporciona el siguiente nucleótido. Si no se detecta la incorporación de nucleótidos, entonces (de nuevo, permitiendo errores de síntesis) la secuencia de nucleótidos prevista es incorrecta. La secuencia prevista puede entonces revisarse para tener en cuenta el error, y el orden de los nucleótidos puede modificarse por consiguiente.

35 Por lo tanto, el presente método es adaptable, ya que utiliza la retroalimentación de la reacción de síntesis para informar y modificar el orden en el que se proporcionan los nucleótidos. Esto permite que la tasa esperada de incorporación de nucleótidos aumente de 1 en 4 (donde, de cada 4 flujos de nucleótidos, solo uno es correcto) a más; tanto como 1 en 1 donde la secuencia se predice perfectamente, aunque en la práctica menos. Esto no solo puede reducir el tiempo necesario para realizar una operación de secuenciación, sino que también puede reducir la cantidad de reactivos necesarios.

40 Los nucleótidos son preferiblemente dNTP; más preferiblemente se proporcionan cuatro dNTP diferentes (A, C, G, T).

El término "seleccionado para correlacionarse" significa preferiblemente que el orden de los nucleótidos coincide con el de una secuencia complementaria al de la secuencia de polinucleótidos prevista. Por "complementario" se entiende preferiblemente "perfectamente complementario", aunque puede haber circunstancias en las que se utilice menos del 100 % de complementariedad.

45 La secuencia de polinucleótidos prevista puede predecirse en función de información a priori. Por ejemplo, la secuencia puede predecirse en función del conocimiento de las secuencias que es probable que se detecten. Por ejemplo, en un entorno clínico en el que se sospecha que un paciente tiene una infección bacteriana o de otro tipo, el conocimiento de la situación clínica, como los síntomas, el historial médico reciente o cualquier otra información del paciente, puede hacer que ciertas infecciones en particular sean más probables que otras. La secuencia prevista puede determinarse en función

de esta información. Por ejemplo, si se sospecha que un paciente tiene una infección por MRSA, entonces la secuencia prevista puede seleccionarse para que sea diagnóstica de infección por MRSA. Se pueden considerar varias infecciones diferentes posibles, pero con diferentes probabilidades. Por lo tanto, el método puede comprender la etapa de determinar la probabilidad relativa inicial de una pluralidad de secuencias previstas y seleccionar la secuencia prevista inicial para que sea la más probable de la pluralidad de secuencias previstas. El método puede comprender además la etapa de, cuando se revisa la secuencia prevista, seleccionar la siguiente más probable de la pluralidad de secuencias previstas como la secuencia prevista revisada. De esta forma, se pueden considerar y confirmar o rechazar varias posibilidades.

Alternativamente, o además, la secuencia prevista puede basarse en una secuencia conocida de interés. Por ejemplo, si se está llevando a cabo un ensayo para detectar la presencia o ausencia de un organismo específico, entonces puede usarse una secuencia de polinucleótidos de diagnóstico de ese organismo como la secuencia prevista. Esto también puede extenderse para cubrir una variedad de organismos diferentes; por ejemplo, los genes ribosomales 16S y 23S pueden ser diagnósticos de una cantidad de organismos diferentes, y uno u otro o ambos pueden usarse como una secuencia prevista inicial. Esto se puede combinar con la realización anterior en la que se selecciona inicialmente una secuencia más probable, y luego se revisa a la siguiente secuencia prevista más probable en el caso de que no se detecte la incorporación de nucleótidos. Pueden usarse muchas otras secuencias de diagnóstico; por ejemplo, oncogenes, toxinas bacterianas.

En determinadas realizaciones, la etapa de controlar la reacción para detectar la incorporación de un nucleótido también puede incluir, en el caso de que se detecte la incorporación, agregar datos que representen ese nucleótido a los datos de secuencia registrados que representan el polinucleótido que se va a secuenciar. Los datos que representan el nucleótido pueden representar el complemento de ese nucleótido. Esto permite que la información de secuencia se registre a medida que avanza la secuenciación. En el caso de que no se detecte la incorporación de nucleótidos, los datos que representan la ausencia de ese nucleótido (o su complemento) de la secuencia también pueden añadirse a los datos de secuencia registrados.

La etapa de revisar los datos de la secuencia prevista puede incluir comparar los datos de la secuencia registrada (incluso, cuando sea relevante, la información de la ausencia de un nucleótido determinado) con los datos de la secuencia de polinucleótidos almacenados en una base de datos, y seleccionar la secuencia candidata más probable de la base de datos que coincida con los datos de secuencia registrados como una secuencia prevista revisada. La base de datos puede ser una base de datos remota; por ejemplo, accesible a través de una red informática como internet; puede estar disponible públicamente (por ejemplo, la base de datos de secuencias de GenBank). Alternativamente, la base de datos puede ser una base de datos local; por ejemplo, almacenados en la memoria de una computadora local o en un dispositivo de almacenamiento de datos local. La base de datos puede representar un subconjunto restringido de información de secuencia, tal como solo aquellas secuencias que se consideran probables de ser de interés. Por ejemplo, una base de datos puede incluir solo datos de las secuencias 16S y 23S de patógenos comunes; el uso de una base de datos restringida puede reducir el tiempo necesario para comparar los datos de secuencia registrados con los datos de secuencia en la base de datos.

El método puede comprender además, antes de la etapa (b), proporcionar múltiples nucleótidos diferentes a la reacción simultáneamente. Los múltiples nucleótidos diferentes pueden carecer de uno, o dos, de los cuatro nucleótidos. Esto permite la extensión rápida de una región de la secuencia de polinucleótidos, hasta que se encuentra un nucleótido faltante. Este aspecto del método puede ser de utilidad cuando se conoce una región de secuencia, pero no se considera informativo o de interés para el propósito de la secuenciación; el o los nucleótidos faltantes pueden seleccionarse para permitir la extensión de la secuencia hasta, o cerca de, la región de interés. De este modo, las regiones que no son de interés pueden evitarse rápidamente. Una región de secuencia conocida que no es de interés puede incluir, por ejemplo, secuencias adaptadoras o regiones de genes altamente conservadas. Los múltiples nucleótidos diferentes que se van a proporcionar pueden seleccionarse basándose en la región conocida que no es de interés, para asegurar que la síntesis rápida termine antes de que se alcance la región de interés. Se pueden proporcionar múltiples nucleótidos diferentes en múltiples rondas.

El método se puede utilizar en circunstancias en las que se va a secuenciar una única secuencia de polinucleótidos; o donde se van a secuenciar múltiples secuencias de polinucleótidos diferentes. En la última circunstancia, el orden predeterminado de nucleótidos puede seleccionarse para correlacionarse con las secuencias previstas para dos o más, preferiblemente todas, de las múltiples secuencias de polinucleótidos diferentes. Preferiblemente, el orden de los nucleótidos se selecciona para permitir una secuenciación eficiente de cada una de las dos o más secuencias de polinucleótidos diferentes. "Eficiente" puede definirse como un orden de nucleótidos que proporciona una proporción mejorada de bases incorporadas a nucleótidos proporcionados en todas las secuencias en consideración, en comparación con un orden en el que las cuatro bases se ciclan de forma repetitiva (p. ej., A, C, G, T repetidas). Por ejemplo, las diferentes secuencias pueden compararse y determinarse una secuencia de consenso; el orden de los nucleótidos puede corresponder con la secuencia de consenso (o su complemento). Se pueden usar otros métodos para determinar el orden de nucleótidos. No es necesario que el orden proporcione la secuenciación más eficiente, ya que el orden más eficiente puede no ser posible por otras razones prácticas. El término "optimizado" también se usa en la presente memoria para

referirse a una secuencia que proporciona una secuenciación eficiente; de nuevo, "optimizado" no implica que la secuencia proporcione la secuenciación más eficiente. Por lo tanto, algunas secuencias pueden estar más optimizadas que otras.

En la presente memoria se describe además un método para optimizar el orden de los nucleótidos proporcionados en una secuenciación por reacción de síntesis, donde el método comprende las etapas de:

- 5 a) determinar una secuencia prevista para un polinucleótido a secuenciar;
- b) determinar un orden predeterminado para que los nucleótidos se proporcionen a una reacción de secuenciación, en donde el orden predeterminado se selecciona para correlacionarse con una secuencia prevista para la cadena de polinucleótidos;
- 10 c) determinar una secuencia real para la cadena de polinucleótidos, proporcionando nucleótidos a una reacción de secuenciación que incluye el polinucleótido que se va a secuenciar;
- d) comparar la secuencia prevista con la secuencia real; y
- e) en el caso de que la secuencia prevista difiera de la secuencia real, revisar la secuencia prevista para la cadena de polinucleótidos en función de la secuencia real y seleccionar un nuevo orden predeterminado de nucleótidos, en donde el nuevo orden predeterminado se selecciona para correlacionarse con la secuencia prevista revisada.
- 15 Otras características de este aspecto de la descripción son las del primer aspecto, descrito anteriormente.

En la presente memoria se describe además un método para secuenciar una cadena de polinucleótidos mediante secuenciación por síntesis, donde el método comprende las etapas de:

- a) proporcionar en una reacción una cadena de polinucleótidos a secuenciar, un cebador, y una enzima polimerasa;
- 20 b) proporcionar a la reacción múltiples nucleótidos diferentes simultáneamente, en donde los múltiples nucleótidos diferentes carecen de al menos uno de los cuatro nucleótidos A, C, G T;
- c) posteriormente proporcionar a la reacción nucleótidos en un orden predeterminado, en donde el orden predeterminado se selecciona para correlacionarse con una secuencia prevista para la cadena de polinucleótidos; y
- d) controlar la reacción para detectar la incorporación de los nucleótidos del orden predeterminado en una cadena de polinucleótidos sintetizados;

25 Adicionalmente, el método puede comprender las etapas de:

- e) en el caso de que se detecte la incorporación de nucleótidos, proceder a proporcionar el nucleótido siguiente en el orden predeterminado; o
- f) en el caso de que no se detecte la incorporación de nucleótidos, revisar la secuencia prevista para la cadena de polinucleótidos y seleccionar un nuevo orden predeterminado de nucleótidos, en donde el nuevo orden predeterminado se selecciona para correlacionarse con la secuencia prevista revisada.
- 30

En determinadas realizaciones, la etapa de controlar la reacción para detectar la incorporación de un nucleótido también puede incluir, en el caso de que se detecte la incorporación, agregar datos que representen ese nucleótido a los datos de secuencia registrados que representan el polinucleótido que se va a secuenciar. Los datos que representan el nucleótido pueden representar el complemento de ese nucleótido. Esto permite que la información de secuencia se registre a medida que avanza la secuenciación. En el caso de que no se detecte la incorporación de nucleótidos, los datos que representan la ausencia de ese nucleótido (o su complemento) de la secuencia también pueden añadirse a los datos de secuencia registrados.

35

La etapa de revisar los datos de la secuencia prevista puede incluir comparar los datos de la secuencia registrada (incluso, cuando sea relevante, la información de la ausencia de un nucleótido determinado) con los datos de la secuencia de polinucleótidos almacenados en una base de datos, y seleccionar la secuencia candidata más probable de la base de datos que coincida con los datos de secuencia registrados como una secuencia prevista revisada. La base de datos puede ser una base de datos remota; por ejemplo, accesible a través de una red informática como internet; puede estar disponible

40

5 públicamente (por ejemplo, la base de datos de secuencias de GenBank). Alternativamente, la base de datos puede ser una base de datos local; por ejemplo, almacenados en la memoria de una computadora local o en un dispositivo de almacenamiento de datos local. La base de datos puede representar un subconjunto restringido de información de secuencia, tal como solo aquellas secuencias que se consideran probables de ser de interés. Por ejemplo, una base de datos puede incluir solo datos de las secuencias 16S y 23S de patógenos comunes; el uso de una base de datos restringida puede reducir el tiempo necesario para comparar los datos de secuencia registrados con los datos de secuencia en la base de datos.

En la presente memoria se describe además un método para secuenciar una cadena de polinucleótidos mediante secuenciación por síntesis, donde el método comprende las etapas de:

- 10 a) proporcionar en una reacción una pluralidad de cadenas de polinucleótidos diferentes a secuenciar, uno o más cebadores, y una enzima polimerasa;
- b) proporcionar a la reacción nucleótidos en un orden predeterminado, en donde el orden predeterminado se selecciona para correlacionarse con secuencias previstas para la pluralidad de cadenas de polinucleótidos diferentes; y
- 15 c) controlar la reacción para detectar la incorporación de los nucleótidos del orden predeterminado en cadenas de polinucleótidos sintetizados;

en donde, el orden predeterminado de los nucleótidos se selecciona para permitir una secuenciación eficiente de cada una de la pluralidad de secuencias de polinucleótidos diferentes.

"Eficiente" puede definirse como un orden de nucleótidos que proporciona una proporción mejorada de bases incorporadas a nucleótidos proporcionados en todas las secuencias en consideración, en comparación con un orden en el que las cuatro bases se ciclan de forma simple (p. ej., A, C, G, T repetidas). Por ejemplo, las diferentes secuencias pueden compararse y determinarse una secuencia de consenso; el orden de los nucleótidos puede corresponder con la secuencia de consenso (o su complemento). Se pueden usar otros métodos para determinar el orden de nucleótidos. No es necesario que el orden proporcione la secuenciación más eficiente, ya que el orden más eficiente puede no ser posible por otras razones prácticas.

## 25 DESCRIPCIÓN DETALLADA DE LA INVENCION

La invención se refiere a reducir el tiempo necesario para determinar la secuencia de nucleótidos de una muestra de polinucleótidos particular mediante la alteración de la secuencia de suministro de nucleótidos. Típicamente, en la secuenciación mediante métodos de síntesis, los nucleótidos se proporcionan individualmente a la reacción de secuenciación en un orden fijo (por ejemplo, A, C, G, T); si el siguiente nucleótido que se añadirá a la cadena que se está sintetizando está presente, la cadena se extenderá y se detectará la extensión. Los nucleótidos no utilizados se lavan de la reacción y se añade el nucleótido siguiente de la secuencia. La secuenciación del ADN utilizando sensores ISFET generalmente implica el flujo secuencial de nucleótidos (T, G, C, A) en el chip. Si el nucleótido particular que fluye a través del chip encuentra su nucleótido complementario en la cadena molde, entonces se libera un ion de hidrógeno y el sensor ISFET lo detecta como un cambio en mV. En general, cada sensor individual detecta los iones de hidrógeno liberados de una población clonal de plantilla de ADN, que puede generarse mediante una técnica de amplificación clonal como la emPCR e inmovilizarse en una esfera dentro de un pocillo. Cada población clonal puede contener muchos millones de copias idénticas de ADN.

En la bibliografía se informa que cuando se usa un ciclo estándar de 4 nucleótidos, una secuencia de ADN aleatoria se extenderá en aproximadamente 2.4 bases por ciclo. Esto se debe a que no todos los flujos de nucleótidos tienen una base complementaria correspondiente en la cadena molde, lo que da como resultado aproximadamente un 40 % de flujos que devuelven resultados nulos o de 0 mer.

En la presente memoria, los inventores describen una serie de enfoques que pueden usarse para optimizar el orden de los nucleótidos suministrados a la cadena en crecimiento y, por lo tanto, reducir el tiempo necesario para secuenciar una muestra. Los enfoques se describen en la presente memoria individualmente, pero se apreciará que se pueden usar uno o más en combinación. La invención está destinada a reducir el tiempo necesario para secuenciar una muestra objetivo particular. La secuencia puede permitir la lectura de una secuencia de cierta longitud en un tiempo más corto, o permitir lecturas de mayor longitud en el mismo período de tiempo. Todos los enfoques comparten la característica común de que el orden de los nucleótidos que se proporcionarán a la reacción se determina previamente, en función de algún conocimiento de la muestra que se va a secuenciar. En algunos de los enfoques, el orden de los nucleótidos puede modificarse durante la secuenciación, en función de los resultados de la secuenciación hasta el momento. El orden de los

nucleótidos puede seleccionarse a priori en función de un algoritmo probabilístico, teniendo en cuenta varios factores clínicos.

Si se adapta el flujo de nucleótidos para que coincida con una secuencia prevista, la eficiencia de cada evento de flujo aumenta hasta un 100 %, en el caso de que se prediga la secuencia correcta.

5 En un ejemplo de la invención, la muestra de interés puede ser una secuencia objetivo, donde se pueden hacer suposiciones sobre la probabilidad de que la muestra coincida con una determinada secuencia. Por ejemplo, en un entorno clínico donde otros datos de pacientes pueden ingresar al sistema para determinar la posible presencia de determinados agentes infecciosos. En otro ejemplo, la secuencia de suministro de nucleótidos al chip puede determinarse y actualizarse de manera inteligente a lo largo del curso de una prueba, dependiendo de la retroalimentación con respecto a la secuencia de ADN que ya se ha establecido.

10 La invención se considera de particular relevancia cuando el objetivo de la muestra es una región genética particular, en contraposición a la secuenciación de ADN de novo o de muy alto rendimiento. Se pueden hacer suposiciones con respecto a la probabilidad de que la secuencia de muestra coincida con una secuencia de referencia conocida.

15 Para una aplicación de secuenciación objetivo, hay margen para mejorar el rendimiento reduciendo el tiempo necesario para realizar la prueba. Debido al hecho de que las secuencias objetivo pueden ser conocidas, o pueden predecirse hasta cierto punto, el patrón de flujo de nucleótidos puede adaptarse para aumentar el rendimiento y la eficiencia de la prueba. Se identifican cuatro formas principales en las que se puede aplicar este concepto, como se analiza en este documento:

**1. Flujo de nucleótidos predictivo**

**2. Flujo de nucleótidos adaptable**

20 **3. Flujo de multinucleótidos**

**4. Flujo optimizado para múltiples moldes**

El presente método podría implementarse de varias formas, como se presenta a continuación.

1) Flujo de secuencia predictivo

25 En el caso de que sea posible realizar una evaluación de probabilidad antes de ejecutar la prueba con respecto a la muestra particular cuestionada. Esto puede incluir tanto regiones "desconocidas" de muestras como regiones "conocidas" tales como secuencias de cebadores utilizadas para la preamplificación o la preparación de bibliotecas. La evaluación de probabilidad puede incluir más datos clínicos del paciente obtenidos de varias otras fuentes, que luego se ingresan en el instrumento de secuenciación. La evaluación de probabilidad puede simplemente realizarla el médico (por ejemplo, al revisar el historial y los síntomas del paciente, concluir que es más probable que una infección sea la *Bacteria #1*, pero que hay menos posibilidades de que la infección sea la *Bacteria # 2* o la *Bacteria #3*). Alternativamente, la evaluación de probabilidad puede hacer uso de un sistema experto o medios automatizados similares para determinar la probabilidad relativa de diversas infecciones.

35 Un sistema experto puede implementarse como una evaluación de probabilidad realizada en el instrumento de secuenciación (o de hecho como un sistema aparte, por ejemplo, un sistema informático que ejecuta un programa informático adecuado) para identificar el agente infeccioso "más probable" de un paciente determinado.

- El ingreso de datos al instrumento por parte del médico, incluidos, por ejemplo, los antecedentes relevantes del paciente (cirugía o traumatismos recientes, infecciones recientes, medicación, etc.) y datos nosocomiales locales (perfiles de enfermedades infecciosas, brotes recientes de resistencia a antibióticos, etc.)

40 - El algoritmo predice el agente infeccioso más probable según los datos ingresados, como en la tabla siguiente (que es simplemente un ejemplo ilustrativo).

- El flujo de nucleótidos inicial adaptado para coincidir exactamente con el del agente infeccioso previsto.

Agente infeccioso	Probabilidad de infección
<b>Bacteria 1</b>	<b>50%</b>
Bacteria 2	20%
Bacteria 3	10%
Bacteria 4	2%
Bacteria 5	2%
Bacteria 6	<1%

5 En una alternativa, el médico puede examinar al paciente y concluir que es más probable que una infección sea causada por un primer agente infeccioso, pero que existe una menor posibilidad de que la infección se deba a un segundo o tercer agente infeccioso (es decir, sin el uso de un sistema experto o sin asignar directamente probabilidades cuantitativas). En este caso, el médico (o un técnico) puede ingresar la lista de tres agentes infecciosos potenciales directamente en un instrumento de secuenciación; el instrumento tomará entonces cada uno de los tres agentes potenciales en orden como el agente infeccioso previsto y, así adaptará el flujo de nucleótidos inicial al primer agente previsto.

10 En la mayoría de los casos, se puede seleccionar un gen informativo para secuenciar (por ejemplo, genes ribosomales de bacterias), y la secuencia de este gen en el agente infeccioso previsto más probable se usa para determinar la secuencia del flujo de nucleótidos inicial.

2) Flujo de secuencia adaptable

15 Se basa en la retroalimentación en tiempo real de la instrumentación de secuenciación y adapta el patrón de nucleótidos en consecuencia. Esto implicaría hacer referencia (de forma continua o en determinados momentos) de los datos generados a una base de datos y adaptar el patrón de flujo de nucleótidos futuro en función de los datos ya generados. Este enfoque es útil en el caso de que se sepa menos acerca de la secuencia cuestionada, o que la secuencia prevista inicialmente resulte ser incorrecta.

Un flujo de secuencia adaptable cambia el patrón de flujo según la retroalimentación en tiempo real del instrumento.

**Ejemplo**

20 La evaluación de probabilidad inicial se lleva a cabo como se describió anteriormente, y el instrumento de secuenciación determina el orden inicial del flujo de nucleótidos, orden que se adapta al agente infeccioso "más probable". Si la secuencia prevista es incorrecta, el instrumento detecta el fallo en la extensión del molde, y se identifica y utiliza una secuencia prevista revisada para determinar un nuevo orden de flujo de nucleótidos.

Por ejemplo, la evaluación inicial da las siguientes probabilidades:

Agente infeccioso	Probabilidad de infección
<b>Bacteria 1</b>	<b>50%</b>
Bacteria 2	20%
Bacteria 3	10%
Bacteria 4	2%
Bacteria 5	2%
Bacteria 6	<1%

25 La infección es en realidad la *Bacteria #3*. El algoritmo predice **incorrectamente** la *Bacteria #1* en función de la evaluación de probabilidad y el instrumento inicia el flujo de nucleótidos adaptado a la secuencia de la *Bacteria #1*

- *Bacteria #3*: secuencia de muestra GCACCTGTCTCAGAGTTCCCGAAGGCACCAAAGCATC

- *Bacteria #1* - en función del patrón de flujo CGTGGACAGAGTGC

A medida que cada nucleótido extiende el molde sobre la cadena objetivo, el instrumento registra datos que representan cada nucleótido y la secuencia de la cadena objetivo y/o sintetizada.

5 Al encontrar la primera base que no coincide, el instrumento recorre los nucleótidos para encontrar la base correcta:

- *Bacteria #3*: secuencia de muestra GCACCTGTCTCAGAGTTCCCGAAGGCACCAAAGCATC

*Bacteria #1* CGTGGACAGAGTGT

10 En este caso, el orden de los nucleótidos predijo que una C sería la siguiente base en la secuencia, para coincidir con la secuencia de la *Bacteria #1*. Sin embargo, la C no extendió la secuencia molde; por lo tanto, el instrumento no detecta un evento de extensión y proporciona los nucleótidos restantes (A, G, T) para determinar cuál extiende el molde. En este caso, la T extiende el molde. Esta información se agrega a un registro de la secuencia sintetizada. En función de la secuencia obtenida, el instrumento compara la secuencia obtenida hasta la fecha con la siguiente secuencia más probable, en este caso, la *Bacteria #2*, y si, como en este caso, la siguiente secuencia más probable coincide con la cadena sintetizada hasta el momento, entonces la siguiente secuencia más probable se convierte en la nueva secuencia prevista, y el orden de flujo de nucleótidos se revisa para que corresponda con la nueva secuencia prevista. Luego, el instrumento reinicia el flujo de nucleótidos adaptado a la *Bacteria #2*:

- *Bacteria #3*: secuencia de muestra GCACCTGTCTCAGAGTTCCCGAAGGCACCAAAGCATC

- *Bacteria #2* - en función del patrón de flujo CGTGGACAGAGTGTCAAGGGCTTCCGTGGTTA

20 Al encontrar una no coincidencia adicional (aquí, una G cuando la secuencia prevista condujo a un orden de flujo de A), el instrumento nuevamente no detecta un evento de extensión y una vez más recorre los nucleótidos restantes para encontrar la inserción de base correcta:

*Bacteria #3*: secuencia de muestra

GCACCTGTCTCAGAGTTCCCGAAGGCACCAAAGCATC

CGTGGACAGAGTGTCAAGGGCTTCCGTGGTTC

25 En función de la secuencia obtenida y las probabilidades relativas de los diversos agentes infecciosos, el algoritmo revisa la secuencia prevista a la de la *Bacteria #3*, y reinicia el flujo de nucleótidos adaptado a la *Bacteria #3*.

- *Bacteria #3*: secuencia de muestra GCACCTGTCTCAGAGTTCCCGAAGGCACCAAAGCATC

- *Bacteria #3* - en función del patrón de flujo CGTGGACAGAGTCTCAAGGGCTTCCGTGGTTTCGTAG

Cuando se alcanza el final de la secuencia sintetizada sin más no coincidencias, se llega a la conclusión de que la infección es la *Bacteria #3*. El instrumento deja de funcionar y muestra una salida que indica la infección detectada.

30 3) Flujo de multinucleótidos

Los enfoques descritos anteriormente hacen uso de métodos que proporcionan un solo nucleótido a la vez en cada evento de flujo. Esto es preferible cuando se secuencia la región de interés, ya que permite comparar la secuencia detectada con la secuencia prevista y revisar la secuencia prevista si es necesario. Sin embargo, en muchos casos habrá al menos alguna región que no sea de interés que necesite secuenciarse antes de que se pueda secuenciar la región de interés; por ejemplo, los cebadores o adaptadores usados para preparar muestras para secuenciación pueden incluirse en las muestras a analizar, y estas regiones no son informativas en cuanto a la identidad de la región de interés. Alternativamente, las regiones altamente conservadas en genes pueden no considerarse de interés, ya que no permiten al usuario distinguir entre secuencias alternativas. Al combinar 2 o 3 nucleótidos en un evento de flujo particular, es posible "pasar más rápido" a través de una región de secuencia conocida. Por ejemplo, puede ser posible extender una secuencia de cebador conocida en pocas cantidades de flujos, sin extender la secuencia "desconocida" de interés. Esto puede ser útil para

extender regiones que están altamente conservadas, por ejemplo, en genes de ARNr 16S y 23S bacterianos, o en regiones raramente mutadas de oncogenes humanos.

5 - Como se menciona anteriormente, se puede realizar una "evaluación de probabilidad" para clasificar los agentes infecciosos en orden de probabilidad, y se puede determinar el orden de los nucleótidos para que coincida con la secuencia del agente infeccioso "más probable". El flujo de nucleótidos se inicia basándose en esta secuencia.

- La parte inicial de la secuencia se puede conservar entre bacterias, por ejemplo

- *Bacteria 4*

**CACCTGTC**ACTCTACTAACGTATGGCTACCCT

- *Bacteria 5* **CACCTGTC**CGAATGAGTATCTTATTACCATTG

10 - *Bacteria 6* **CACCTGTC**ATACGACGCATACGGTTGAAACA

- En este caso, en lugar de hacer fluir los nucleótidos individuales para secuenciar la parte CACCTGTC no informativa inicial, es posible "pasar más rápido" a través de la región conservada mezclando 2 o más nucleótidos. Esto se puede hacer usando un depósito de reactivo dedicado de nucleótidos mixtos o mezclando nucleótidos "en el momento"

- En la nomenclatura estándar de nucleótidos:

- 15
- "C y A" se pueden indicar con "M" (para nucleótidos que contienen aMino)
  - "T y G" se pueden indicar con "K" (para nucleótidos que contienen ceto")

- En este ejemplo, un flujo inicial de M seguido de K extendería 7 bases en la región conservada:

- *Bacteria 4* **CACCTGTC**ACTCTACTAACGTATGGCTACCCT

- *Bacteria 5* **CACCTGTC**CGAATGAGTATCTTATTACCATTG

20 - *Bacteria 6* **CACCTGTC**ATACGACGCATACGGTTGAAACA

- Flujo de multinucleótidos **M >>> K >>**

25 - Si bien el instrumento detectará que se ha producido un evento de extensión, no será evidente qué nucleótido ha extendido la secuencia. Sin embargo, para esta región conservada, eso no es motivo de preocupación. En algunas realizaciones, se pueden proporcionar tres nucleótidos juntos; por ejemplo, proporcionar una mezcla de A, C y T extenderá la secuencia inicial de CACCT y se detendrá en la siguiente G. Después de este "pasar más rápido" inicial a través de una región conservada conocida, el orden de los nucleótidos podría revertirse a un secuencia predictiva o adaptable, como se discutió anteriormente, para continuar la prueba de secuenciación.

4) Flujo optimizado para múltiples secuencias

30 En muchas aplicaciones, habrá > 1 secuencia que requiere secuenciación. En este caso, es posible crear un algoritmo para determinar el patrón de flujo de nucleótidos más eficiente para secuenciar todos los moldes. De hecho, las simulaciones muestran que cuando el patrón de flujo se optimiza según un algoritmo básico, que se describe a continuación, la cantidad de flujos de nucleótidos necesarios para secuenciar tres moldes objetivo de 50 bp se puede reducir hasta en un 28 %.

35 - Como se menciona anteriormente, se puede realizar una "evaluación de probabilidad" para clasificar los agentes infecciosos en orden de probabilidad, y se puede determinar el orden de los nucleótidos para que coincida con la secuencia del agente infeccioso "más probable". El flujo de nucleótidos se inicia basándose en esta secuencia.

- Hay escenarios en los que habrá > 1 secuencia cuestionada. Por ejemplo, en una aplicación de oncología, es probable que haya que secuenciar varios oncogenes.

- Para las infecciones del torrente sanguíneo, si se requiere que la prueba proporcione una identificación detallada, puede haber > 1 región genómica cuestionada, incluso dentro de un solo organismo.

- Las siguientes secuencias son representativas de múltiples secuencias que pueden requerir secuenciación simultánea.

- Secuencia 1 ACTCTACTAACGTATGGCTACCCTTAGTGGGGATGCTACCTAAAACCCTTC
- Secuencia 2 CGAATGAGTATCTTATTACCATTTTGCAGTCCAATGTTTTAATTGTGTTGT
- Secuencia 3 ATACGACGCATACGGTTCGAAACAAGAACGTACAATGTACGGAACCTCGACA
- El uso de un patrón de flujo cíclico T, G, C, A requiere **100 eventos de flujo** para secuenciar los tres objetivos.
- Un patrón de flujo óptimo secuencia todos los moldes en **72 eventos de flujo**:

- ACTGACTGACGTACGTATCGACTAGTCAGTACAGTGACTGAC TACGACTGATCGATCGATCGTTCGATCGAT

10 - 28 % de reducción en la duración del flujo

- El patrón óptimo se puede determinar como el patrón que minimiza la cantidad de eventos de flujo para secuenciar todos los moldes de interés. Se pueden utilizar definiciones alternativas.

15 - En este escenario, las secuencias no se parecen entre sí ni comparten regiones de homología significativa. Sin embargo, es posible determinar un patrón de flujo de nucleótidos no cíclico que complete la extensión de todas las secuencias en el menor tiempo posible.

- Un algoritmo predictivo y adaptativo utilizado junto con este aspecto optimizaría aún más la prueba y reduciría la duración.

Otro ejemplo de múltiples regiones objetivo, y la reducción de ciclos necesarios para secuenciar todos los objetivos, se da por lo siguiente:

Secuencias objetivo representativas:

20 Secuencia 1 ACTCTACTAACGTATGGCTACCCTTAGTGGGGATGCTACCTAAAACCCTTC

**Secuencia 2** CGAATGAGTATCTTATTACCATTTTGCAGTCCAATGTTTTAATTGTGTTGT  
ATACGACGCATACGGTTCGAAACAAGAACGTACAATGTACGGAACCTCGAC

Secuencia 3 A

25 Si se utiliza un ciclo simple de los cuatro nucleótidos, cada uno de estos tomaría, respectivamente, 84, 72 y 100 nucleótidos para completar la secuenciación. Si se utiliza una secuencia optimizada, ACTGACTGACGTACGTATCGACTAGTCAGTACAGTGACTGACTGACTGACTGATCG ATCGATCGTTCGATCGAT, la secuenciación de todas las secuencias objetivo podría completarse en 72 nucleótidos.

A continuación se ofrece un ejemplo del proceso de optimización. Se debe considerar que este es simplemente un ejemplo; se pueden utilizar otros algoritmos y es posible que dichos algoritmos proporcionen un mayor grado de optimización.

30 En resumen, el algoritmo analiza la siguiente base que se secuenciará en cada una de las cadenas de ADN y selecciona la que es más probable que ocurra. Si hay dos o más bases con la misma probabilidad de surgir, escoge la base que ha estado esperando más tiempo para fluir. Como punto de partida para describir el algoritmo, supongamos que queremos secuenciar una muestra que creemos que contiene una de las siguientes secuencias de ADN:

Secuencia 1: ACTCTACTAACGTATGGCTACCCTTAGTGGGGATGCTACCTAAAACCCTTC

Secuencia 2: CGAATGAGTATCTTATTACCATTTTGCAGTCCAATGTTTTAATTGTGTTGT

35 Secuencia 3: ATACGACGCATACGGTTCGAAACAAGAACGTACAATGTACGGAACCTCGACA

# ES 2 813 825 T3

Para cada secuencia, el algoritmo mantiene un indicador de la siguiente base a secuenciar. En los detalles a continuación, estos indicadores se representan como flechas debajo de cada secuencia. Las siguientes secciones muestran las primeras diez etapas del algoritmo para ilustrar cómo funciona.

## Etapas de flujo 1

5 Antes de que fluya el primer nucleótido, las posiciones actuales en cada secuencia son las que se muestran a continuación:

```
Secuencia 1  ACTCTACTAACGTATGGCTACCCTTAGTGGGGATGCTACCTAAAACCCCTTC
                ^
Secuencia 2  CGAATGAGTATCTTATTACCATTTTGCAGTCCAATGTTTAAATTGTGTGT
                ^
Secuencia 3  ATACGACGCATACGGTTCGAAACAAGAACGTACAATGTACGGAACTCGACA
                ^
```

10 En este caso, dos de las secuencias son A y una es C. Por lo tanto, el primer nucleótido que fluirá será A, ya que es más probable que produzca una extensión.

## Etapas de flujo 2

Antes de que fluya este nucleótido, las posiciones actuales en cada secuencia son las que se muestran a continuación:

```
Secuencia 1  ACTCTACTAACGTATGGCTACCCTTAGTGGGGATGCTACCTAAAACCCCTTC
                - ^
Secuencia 2  CGAATGAGTATCTTATTACCATTTTGCAGTCCAATGTTTAAATTGTGTGT
                ^
Secuencia 3  ATACGACGCATACGGTTCGAAACAAGAACGTACAATGTACGGAACTCGACA
                - ^
```

En este caso, dos de los nucleótidos son C y uno es T. El algoritmo elige C, ya que es más probable que dé una extensión.

15 **Etapas de flujo 3**

Antes de que fluya este nucleótido, las posiciones actuales en cada secuencia son las que se muestran a continuación:

```
Secuencia 1  ACTCTACTAACGTATGGCTACCCTTAGTGGGGATGCTACCTAAAACCCCTTC
                - - ^
Secuencia 2  CGAATGAGTATCTTATTACCATTTTGCAGTCCAATGTTTAAATTGTGTGT
                - ^
Secuencia 3  ATACGACGCATACGGTTCGAAACAAGAACGTACAATGTACGGAACTCGACA
                - - ^
```

En este caso, dos de los nucleótidos son T y uno es G. El algoritmo elige T, ya que es más probable que dé una extensión.

## Etapas de flujo 4

# ES 2 813 825 T3

Antes de que fluya este nucleótido, las posiciones actuales en cada secuencia son las que se muestran a continuación:

```
Secuencia 1  ACTCTACTAACGTATGGCTACCCTTAGTGGGGATGCTACCTAAAACCCCTTC
                ____^
Secuencia 2  CGAATGAGTATCTTATTACCATTTGTCAGTCCAATGTTTAAATTGTGTGT
                _ ^
Secuencia 3  ATACGACGCATACGGTTCGAAACAAGAACGTACAATGTACGGAACTCGACA
                ____^
```

En este caso, los tres nucleótidos C, G y A son todos igualmente probables. El nucleótido que espera más tiempo es G, por lo que se elige como nucleótido nido en la secuencia.

## 5 Etapa de flujo 5

Antes de que fluya este nucleótido, las posiciones actuales en cada secuencia son las que se muestran a continuación:

```
Secuencia 1  ACTCTACTAACGTATGGCTACCCTTAGTGGGGATGCTACCTAAAACCCCTTC
                ____^
                CGAATGAGTATCTTATTACCATTTGTCAGTCCAATGTTTAAATTGTGTGT
Secuencia 2  ____^
                ATACGACGCATACGGTTCGAAACAAGAACGTACAATGTACGGAACTCGACA
Secuencia 3  ____^
```

En este caso, el nucleótido A es más probable, por lo que se elige este.

## Etapa de flujo 6

10 Antes de que fluya este nucleótido, las posiciones actuales en cada secuencia son las que se muestran a continuación:

```
Secuencia 1  ACTCTACTAACGTATGGCTACCCTTAGTGGGGATGCTACCTAAAACCCCTTC
                ____^
                CGAATGAGTATCTTATTACCATTTGTCAGTCCAATGTTTAAATTGTGTGT
Secuencia 2  ____^
                ATACGACGCATACGGTTCGAAACAAGAACGTACAATGTACGGAACTCGACA
Secuencia 3  ____^
```

En este caso, el nucleótido c es más probable, por lo que se elige este.

## Etapa de flujo 7

15 Antes de que fluya este nucleótido, las posiciones actuales en cada secuencia son las que se muestran a continuación:

```
Secuencia 1  ACTCTACTAACGTATGGCTACCCTTAGTGGGGATGCTACCTAAAACCCCTTC
                ____^
                CGAATGAGTATCTTATTACCATTTGTCAGTCCAATGTTTAAATTGTGTGT
Secuencia 2  ____^
                ATACGACGCATACGGTTCGAAACAAGAACGTACAATGTACGGAACTCGACA
Secuencia 3  ____^
```

En este caso, el nucleótido T es más probable, por lo que se elige este.

**Etapas de flujo 8**

Antes de que fluya este nucleótido, las posiciones actuales en cada secuencia son las que se muestran a continuación:

```

Secuencia 1  ACTCTACTAACGTATGGCTACCCTTAGTGGGGATGCTACCTAAAACCCCTTC
                _____ ^
Secuencia 2  CGAATGAGTATCTTATTACCATTTTGCAGTCCAATGTTTTAATTGTGTTGT
                _____ ^
Secuencia 3  ATACGACGCATACGGTTCGAAACAAGAACGTACAATGTACGGAACTCGACA
                _____ ^
    
```

5 En este caso, el nucleótido G es más probable, por lo que se elige este.

**Etapas de flujo 9**

Antes de que fluya este nucleótido, las posiciones actuales en cada secuencia son las que se muestran a continuación:

```

Secuencia 1  ACTCTACTAACGTATGGCTACCCTTAGTGGGGATGCTACCTAAAACCCCTTC
                _____ ^
Secuencia 2  CGAATGAGTATCTTATTACCATTTTGCAGTCCAATGTTTTAATTGTGTTGT
                _____ ^
Secuencia 3  ATACGACGCATACGGTTCGAAACAAGAACGTACAATGTACGGAACTCGACA
    
```

En este caso, todas las secuencias requieren el nucleótido A, por lo que se elige este.

10 **Etapas de flujo 10**

Antes de que fluya este nucleótido, las posiciones actuales en cada secuencia son las que se muestran a continuación:

```

Secuencia 1  ACTCTACTAACGTATGGCTACCCTTAGTGGGGATGCTACCTAAAACCCCTTC
Secuencia 2  CGAATGAGTATCTTATTACCATTTTGCAGTCCAATGTTTTAATTGTGTTGT
Secuencia 3  ATACGACGCATACGGTTCGAAACAAGAACGTACAATGTACGGAACTCGACA
    
```

En este caso, el nucleótido c es más probable, por lo que se elige este.

15 En muchas aplicaciones, sería beneficioso combinar varios de los métodos de implementación anteriores. Por ejemplo, en una herramienta de infección del torrente sanguíneo, un algoritmo probabilístico podría predecir el agente infeccioso más probable. Al predecir la secuencia de varios amplicones 16S, se pudo determinar un patrón de flujo optimizado. Si, durante la prueba, el agente infeccioso previsto resulta ser incorrecto, un flujo adaptativo inteligente podría ajustar el patrón de flujo de nucleótidos óptimo basándose en un algoritmo probabilístico revisado. Dadas las limitaciones para determinar una secuencia de flujo óptima para múltiples secuencias objetivo, es probable que los diversos aspectos de la presente descripción sean principalmente adecuados para situaciones en las que es probable que estén presentes solo una o una pequeña cantidad de secuencias.

LISTADO DE SECUENCIAS

- <110> DNA Electronics Ltd
- 25 <120> Métodos de secuenciación
- <130> PC927610WO
- <150> GB 1407334.0

< 151> 2014-04-25  
<150> GB 1503465.5  
< 151> 02/03/2015  
<160> 20  
5 <170> PatentIn versión 3.5  
<210> 1  
< 211> 37  
< 212> ADN  
<213> Secuencia artificial  
10 <220>  
<223> secuencia hipotética  
<400> 1  
gcacctgtct cagagttccc gaaggcacca aagcatc  
37  
15 <210> 2  
< 211> 14  
< 212> ADN  
<213> Secuencia artificial  
<220>  
20 <223> secuencia hipotética  
<400> 2  
cgtggacaga gtgc  
14  
<210> 3  
25 < 211> 14  
< 212> ADN  
<213> Secuencia artificial  
<220>  
<223> secuencia hipotética  
30 <400> 3  
cgtggacaga gtgt  
14  
<210> 4  
< 211> 32  
35 < 212> ADN

<213> Secuencia artificial  
<220>  
<223> secuencia hipotética  
<400> 4  
5 cgtggacaga ggtcaaggg cttcgtggt ta  
32  
<210> 5  
< 211> 32  
< 212> ADN  
10 <213> Secuencia artificial  
<220>  
<223> secuencia hipotética  
<400> 5  
cgtggacaga ggtcaaggg cttcgtggt tc  
15 32  
<210> 6  
< 211> 37  
< 212> ADN  
<213> Secuencia artificial  
20 <220>  
<223> secuencia hipotética  
<400> 6  
cgtggacaga ggtcaaggg cttcgtggt ttogtag  
37  
25 <210> 7  
< 211> 32  
< 212> ADN  
<213> Secuencia artificial  
<220>  
30 <223> secuencia hipotética  
<400> 7  
cacctgtcac tctactaacg tatggctacc ct  
32  
<210> 8  
35 < 211> 32

< 212> ADN  
<213> Secuencia artificial  
<220>  
<223> secuencia hipotética  
5 <400> 8  
cacctgtccg aatgagtatc ttattacat tg  
32  
<210> 9  
< 211> 32  
10 < 212> ADN  
<213> Secuencia artificial  
<220>  
<223> secuencia hipotética  
<400> 9  
15 cacctgtcat acgacgcata cggttcgaaa ca  
32  
<210> 10  
< 211> 51  
< 212> ADN  
20 <213> Secuencia artificial  
<220>  
<223> secuencia hipotética  
<400> 10  
actctactaa cgtatggcta cccttagtgg ggatgctacc taaaaccctt c  
25 51  
<210> 11  
< 211> 51  
< 212> ADN  
<213> Secuencia artificial  
30 <220>  
<223> secuencia hipotética  
<400> 11  
cgaatgagta tcttattacc atttgcagt ccaatgtttt aattgtgtg t  
51  
35 <210> 12

< 211> 51  
 < 212> ADN  
 <213> Secuencia artificial  
 <220>  
 5 <223> secuencia hipotética  
 <400> 12  
 atacgacgca tacgggtcga aacaagaacg tacaatgtac ggaactcgac a  
 51  
 <210> 13  
 10 < 211> 72  
 < 212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> secuencia hipotética  
 15 <400> 13  
 actgactgac gtacgtatcg actagtcagt acagtgactg actacgtact gatcgatcga 60  
 tcgtcgatcg at 72  
 <210> 14  
 < 211> 51  
 < 212> ADN  
 20 <213> Secuencia artificial  
 <220>  
 <223> secuencia hipotética  
 <400> 14  
 actctactaa cgtatggcta cccttagtgg ggatgctacc taaaaccctt c  
 25 51  
 <210> 15  
 < 211> 51  
 < 212> ADN  
 <213> Secuencia artificial  
 30 <220>  
 <223> secuencia hipotética  
 <400> 15  
 cgaatgagta tcttattacc atttgcagt ccaatgtttt aattgtgtg t  
 51

ES 2 813 825 T3

<210> 16  
< 211> 51  
< 212> ADN  
<213> Secuencia artificial  
5 <220>  
<223> secuencia hipotética  
<400> 16  
atacagcgca tacggtcga aacaagaacg tacaatgtac ggaactcgac a  
51  
10 <210> 17  
< 211> 72  
< 212> ADN  
<213> Secuencia artificial  
<220>  
15 <223> secuencia hipotética  
<400> 17  
actgactgac gtagctatcg actagtcagt acagtgactg actacgtact gatcgatcga  
60  
tcgtcgatcg at  
20 72  
<210> 18  
< 211> 51  
< 212> ADN  
<213> Secuencia artificial  
25 <220>  
<223> secuencia hipotética  
<400> 18  
actctactaa cgtatggcta cccttagtgg ggatgctacc taaaaccctt c  
51  
30 <210> 19  
< 211> 51  
< 212> ADN  
<213> Secuencia artificial  
<220>  
35 <223> secuencia hipotética

# ES 2 813 825 T3

<400> 19

cgaatgagta tcttattacc atttgcagt ccaatgtttt aattgtgtg t

51

<210> 20

5

< 211> 51

< 212> ADN

<213> Secuencia artificial

<220>

<223> secuencia hipotética

10

<400> 20

atacgacgca tacggtcga aacaagaacg tacaatgtac ggaactcgac a

51

**REIVINDICACIONES**

- 5
1. Un método para secuenciar una cadena de polinucleótidos mediante secuenciación por síntesis, comprendiendo el método las etapas de:
- 10
- a) proporcionar en una reacción una cadena de polinucleótidos a secuenciar, un cebador, y una enzima polimerasa;
  - b) proporcionar a la reacción nucleótidos en un orden predeterminado, en donde el orden predeterminado se selecciona para correlacionarse con una secuencia prevista para la cadena de polinucleótidos;
  - c) controlar la reacción para detectar la incorporación de un nucleótido en una cadena de polinucleótidos sintetizados;
- 15
- en donde, en el caso de que se detecte la incorporación de nucleótidos, proceder a proporcionar el nucleótido siguiente en el orden predeterminado; y
- en donde en el caso de que no se detecte la incorporación de nucleótidos, se revisa la secuencia prevista para la cadena de polinucleótidos para seleccionar un nuevo orden predeterminado de nucleótidos, en donde el nuevo orden predeterminado se selecciona para correlacionarse con la secuencia prevista revisada.
- 20
2. El método de la reivindicación 1, en donde no se detecta la incorporación de nucleótidos, y la secuencia prevista para la cadena de polinucleótidos se revisa para seleccionar un nuevo orden predeterminado de nucleótidos, en donde el nuevo orden predeterminado se selecciona para correlacionarse con la secuencia prevista revisada.
- 25
3. El método de la reivindicación 1, en donde, en el caso de que no se detecte la incorporación de nucleótidos, se detiene la reacción.
4. El método de cualquier reivindicación precedente, en donde la secuencia de polinucleótidos prevista se predice basándose en información *a priori*.
- 30
5. El método de cualquier reivindicación precedente, en donde la etapa de controlar la reacción para detectar la incorporación de un nucleótido también incluye, en el caso de que se detecte la incorporación, agregar datos que representen ese nucleótido a los datos de secuencia registrados que representan el polinucleótido que se va a secuenciar.
- 35
6. El método de cualquier reivindicación precedente, en donde la etapa de controlar la reacción para detectar la incorporación de un nucleótido también incluye, en el caso de que no se detecte la incorporación de nucleótidos, agregar datos que representen la ausencia de ese nucleótido a los datos de secuencia registrados que representan el polinucleótido que se va a secuenciar.
- 40
7. El método de cualquier reivindicación precedente que comprende además proporcionar múltiples nucleótidos diferentes a la reacción simultáneamente, en donde los múltiples nucleótidos diferentes carecen de uno, o dos, de los cuatro nucleótidos A, G, C, T.
- 45
8. El método de cualquier reivindicación precedente, en donde el orden predeterminado de nucleótidos se selecciona para correlacionarse con las secuencias previstas para dos o más polinucleótidos diferentes a secuenciar.
9. El método de la reivindicación 8, en donde el orden de los nucleótidos se selecciona para permitir una secuenciación eficiente de cada una de las dos o más secuencias de polinucleótidos diferentes.