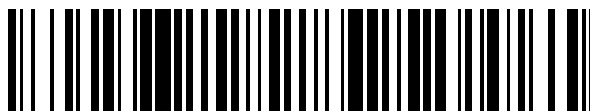


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 716 094**

51 Int. Cl.:

**C12Q 1/689** (2008.01)

**C12Q 1/6869** (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **12.03.2014 PCT/US2014/024516**

87 Fecha y número de publicación internacional: **25.09.2014 WO14150910**

96 Fecha de presentación y número de la solicitud europea: **12.03.2014 E 14767628 (2)**

97 Fecha y número de publicación de la concesión europea: **16.01.2019 EP 2971140**

54 Título: **Métodos para analizar la contaminación en la secuenciación del ADN**

30 Prioridad:  
**15.03.2013 US 201361787451 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**10.06.2019**

73 Titular/es:  
**IBIS BIOSCIENCES, INC. (100.0%)  
2251 Faraday Avenue Suite 150  
Carlsbad, CA 92008, US**

72 Inventor/es:  
**ESHOO, MARK W. y  
MOTLEY, STANLEY**

74 Agente/Representante:  
**IZQUIERDO BLANCO, María Alicia**

ES 2 716 094 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 para analizar la contaminación en la secuenciación del ADN

5 **CAMPO DE LA INVENCION**

[0001] La presente invención está relacionada con los métodos y sistemas para detectar la contaminación en los análisis de ácido nucleico. Más particularmente, la presente invención está relacionada con los métodos y sistemas para detectar la contaminación en las pruebas o ensayos de secuenciación de ADN.

10

**ANTECEDENTES**

[0002] Puesto que es ultra-sensible, la secuenciación de ADN de nueva generación (o NGS, por las siglas en inglés de 'Next Generation DNA sequencing') está expuesta a la contaminación debida a múltiples fuentes. Muchas operaciones de secuenciación NGS muestran niveles detectables de secuencias contaminantes. La presencia de contaminantes puede provocar resultados de falsos positivos cuando se secuencian ácidos nucleicos que están presentes en bajos niveles. Por ejemplo, si en una muestra clínica se busca una rara variante de cáncer que se encuentra en un 1% de las secuencias de un locus de interés, puede ser difícil determinar si pertenece a la muestra o a ADN humano contaminante.

15

20

[0003] Se ha demostrado que un 20% de los genomas de GenBank finalizados contienen secuencias contaminantes de ADN humano. Cada paso del análisis es vulnerable a la contaminación (Mark S. Longo et al., Febrero, 2011; PLoS ONE; Otto Erlwein et al., 2011, PLoS ONE 6(8)).

25

[0004] Se necesitan métodos para calcular o analizar los niveles de contaminación a fin de reducir los falsos positivos en los ensayos diagnósticos basados en la secuenciación.

30

[0005] La técnica anterior en este campo desvela secuenciaciones paralelas masivas de ácidos nucleicos humanos diana que comprenden una variante de secuencia y ADN mitocondrial en paralelo como control para una contaminación potencial (Rogaev E I et al., Nov. 2009, Science 326:817 y 'Supporting Online Material'). La técnica anterior también desvela métodos para analizar ácidos nucleicos bacterianos contaminantes en secuenciaciones de muestras de sangre humana (Lee J.-E. et al., 2010, 'Biopreservation and Biobanking'; 8(3):127-131).

35

**RESUMEN**

[0006] La presente invención está relacionada con los métodos para detectar la contaminación en los análisis de ácido nucleico. Más particularmente, la presente invención está relacionada con los métodos para detectar la contaminación en las pruebas o ensayos de secuenciación de ADN.

40

[0007] La presente invención proporciona un método para detectar ácido nucleico humano contaminante en una muestra, de manera que el método comprende:

a) secuenciar

i) un ácido nucleico humano diana de la mencionada muestra, de manera que el ácido nucleico diana es un ácido nucleico variable; y

45

ii) un ácido nucleico mitocondrial humano, de manera que el mencionado ácido nucleico mitocondrial humano proviene del mencionado ácido nucleico humano contaminante y de la mencionada muestra;

en una muestra de ácido nucleico para obtener un conjunto de datos de secuenciación;

b) determinar el número de lecturas de secuenciación atribuible a los mencionados ácidos nucleicos mitocondriales y el número de lecturas de secuenciación atribuible al mencionado ácido nucleico diana en el mencionado conjunto de datos de secuenciación;

50

c) comparar el número de lecturas de secuenciación atribuible a los mencionados ácidos nucleicos mitocondriales respecto al número de lecturas de secuenciación atribuible al mencionado ácido nucleico diana; y

d) determinar la presencia de ácido nucleico humano contaminante en la mencionada muestra basándose en la mencionada comparación cuando se identifica un número mayor de lecturas de secuenciación atribuible a los ácidos nucleicos mitocondriales humanos en relación con el ácido nucleico diana.

55

[0008] La presente invención también proporciona un método para detectar ácido nucleico bacteriano contaminante en una muestra, de manera que el método comprende:

a) secuenciar

60

i) un ácido nucleico diana; y  
ii) un ácido nucleico bacteriano 16S, de manera que el mencionado ácido nucleico bacteriano 16S proviene del mencionado ácido nucleico bacteriano contaminante y de la mencionada muestra;

en una muestra de ácido nucleico para obtener un conjunto de datos de secuenciación;

b) determinar el número de lecturas de secuenciación atribuible a los mencionados ácidos nucleicos bacterianos 16S y el número de lecturas de secuenciación atribuible al mencionado ácido nucleico diana en el mencionado conjunto de datos de secuenciación;

65

c) comparar el número de lecturas de secuenciación atribuible a los mencionados ácidos nucleicos bacterianos 16S respecto al número de lecturas de secuenciación atribuible al mencionado ácido nucleico diana; y  
 d) determinar la presencia de ácido nucleico bacteriano contaminante en la mencionada muestra basándose en la mencionada comparación cuando se identifica un número mayor de lecturas de secuenciación atribuible a los ácidos nucleicos bacterianos 16S en relación con el ácido nucleico diana.

[0009] Así, en algunas realizaciones, el ácido nucleico contaminante es humano y el ácido nucleico de control es ácido nucleico mitocondrial humano. En algunas realizaciones, el ácido nucleico contaminante es bacteriano y el ácido nucleico de control es ácido nucleico bacteriano 16S (por ejemplo, el ácido nucleico 16S proviene de *B. anthracis*).

[0010] En algunas realizaciones, para determinar los niveles de los ácidos nucleicos diana y de control se utiliza un procesador informático y software informático. En algunas realizaciones, un nivel elevado de ácido nucleico de control respecto al ácido nucleico diana indica un nivel elevado de ácido nucleico contaminante en la muestra. En algunas realizaciones, el ácido nucleico diana es ácido nucleico humano (por ejemplo, una variante, supresión o amplificación de ácido nucleico). En algunas realizaciones, la muestra es una muestra humana y el ácido nucleico diana es vírico, bacteriano o fúngico. En algunas realizaciones, la secuenciación es una secuenciación de nueva generación (por ejemplo, usando las técnicas que se describen en el presente documento). En algunas realizaciones, los ácidos nucleicos diana y de control se amplifican antes de la secuenciación (por ejemplo, usando una amplificación del genoma completo).

[0011] En el presente documento también se describen kits que comprenden, básicamente se componen de o se componen de los componentes necesarios, útiles o suficientes para determinar la presencia, la ausencia o el nivel de contaminación de ácido nucleico en una prueba o ensayo de secuenciación. Por ejemplo, en algunos aspectos, los kits o equipos contienen cebadores o partidores, fijadores de ácido nucleico que se usan en las pruebas de secuenciación, sondas o reactivos. En algunos aspectos, los kits comprenden un software para determinar la presencia, la ausencia o el nivel de contaminación del ácido nucleico en una reacción de secuenciación.

[0012] En otros aspectos adicionales, la presente divulgación proporciona una mezcla reactiva que comprende una mezcla de ácido nucleico diana de secuenciación de ácido nucleico; y un ácido nucleico contaminante y al menos un cebador de secuenciación emparejado con el ácido nucleico contaminante.

[0013] Otras realizaciones adicionales resultarán evidentes para las personas versadas en este campo gracias a los principios que se exponen en el presente documento.

## BREVE DESCRIPCIÓN DE LAS ILUSTRACIONES

### [0014]

La Figura 1 muestra los resultados de una NGS que usa una secuencia mitocondrial humana para detectar la presencia de ADN humano.

La Figura 2 muestra que se hallaron muchos alineamientos de secuencias 16S con secuencias 16S de *B. anthracis* en el conjunto de datos de *B. cereus*.

## DESCRIPCIÓN DETALLADA

[0015] La presente invención está relacionada con los métodos para detectar la contaminación en los análisis de ácido nucleico. Más particularmente, la presente invención está relacionada con los métodos y sistemas para detectar la contaminación en las pruebas o ensayos de secuenciación de ADN.

### Definiciones

[0016] Para facilitar la comprensión de la presente tecnología, a continuación se definen diversos términos y expresiones. A lo largo de la descripción detallada se ofrecen otras definiciones adicionales.

[0017] Tal y como se utiliza en el presente documento, 'un', 'una', 'el' o 'la' pueden significar 'uno o más de uno'. Por ejemplo, 'un' aparato puede significar 'un aparato o varios aparatos'.

[0018] Tal y como se utiliza en el presente texto, el término 'marcador' (o 'etiqueta') hace referencia a cualquier átomo o molécula que puede usarse para proporcionar un efecto detectable (y, preferiblemente, cuantificable) y que puede unirse a un ácido nucleico o una proteína. Los marcadores incluyen -pero no se limitan a- tintes o tinciones; radiomarcadores como <sup>32</sup>P; componentes o fracciones conectoras como la biotina; haptenos como la digoxigenina; fracciones luminogénicas, fosforescentes o fluorogénicas; y tinciones fluorescentes, solas o en combinación con porciones o fracciones que pueden suprimir o alterar los espectros de emisiones mediante FRET (transferencia de energía de resonancia de Förster). Los marcadores pueden proporcionar señales detectables mediante fluorescencia, radioactividad, colorimetría, gravimetría, difracción o absorción de rayos X, magnetismo, actividad

enzimática y similares. Un marcador puede ser una fracción o porción cargada (carga positiva o negativa) o, de manera alternativa, puede tener una carga neutra. Los marcadores pueden incluir o estar compuestos de secuencias de proteínas o ácido nucleico, siempre y cuando la secuencia que contenga el marcador sea detectable.

5 **[0019]** El término 'cebador', 'partidor' o 'primer' hace referencia a un oligonucleótido, ya sea natural -como en una digestión de restricción purificada- o producido sintéticamente, que es capaz de actuar como punto de comienzo de la síntesis cuando se somete a unas condiciones en las que se induce o provoca la síntesis de un producto de extensión de un cebador que es complementario respecto a una cadena de ácido nucleico (por ejemplo, en presencia de nucleótidos y un agente inductor como ADN polimerasa y con una temperatura y un pH adecuados).  
10 Preferiblemente, el cebador es monocatenario a fin de obtener la máxima eficacia de amplificación, pero, de manera alternativa, también puede ser bicatenario o de cadena doble. Si es bicatenario, primero se trata el cebador para separar sus hebras o cadenas antes de usarse para preparar productos de extensión. Preferiblemente, el cebador es un oligodesoxirribonucleótido. El cebador debería ser lo suficientemente largo como para cebar o iniciar la síntesis de los productos de extensión en presencia del agente inductor. La longitud exacta de los cebadores  
15 dependerá de muchos factores, incluyendo la temperatura, la fuente del cebador y el uso del método.

**[0020]** El término 'diana' u 'objetivo', cuando se usa en referencia a la reacción en cadena de la polimerasa, hace referencia a la región de ácido nucleico delimitada por los cebadores utilizados para la reacción en cadena de la polimerasa. Así, se busca la 'diana' para separarla del resto de secuencias de ácido nucleico. Un 'segmento' es una  
20 región de ácido nucleico de la secuencia diana.

**[0021]** Tal y como se utiliza en el presente texto, los términos 'sujeto' y 'paciente' hacen referencia a cualquier animal, como un perro, un gato, un ave, ganado y, particularmente, un mamífero, y, preferiblemente, un humano.

25 **[0022]** Tal y como se utiliza en el presente documento, el término 'sensibilidad' hace referencia a una medición estadística del rendimiento de un ensayo (por ejemplo, un método, un análisis o una prueba), que se calcula dividiendo el número de positivos verdaderos por la suma de positivos verdaderos y falsos negativos.

30 **[0023]** Tal y como se utiliza en el presente documento, el término 'especificidad' hace referencia a una medición estadística del rendimiento de un ensayo (por ejemplo, un método, un análisis o una prueba), que se calcula dividiendo el número de positivos verdaderos por la suma de positivos verdaderos y falsos negativos.

**[0024]** Tal y como se utiliza en el presente documento, el término 'amplicón' hace referencia a un ácido nucleico generado mediante una reacción de amplificación. Normalmente, el amplicón es ADN bicatenario; sin embargo, puede ser ARN y/o ADN:ARN. El amplicón comprende ADN que es complementario a un ácido nucleico de muestra. En algunas realizaciones, se configuran pares o parejas de cebadores para generar amplicones de un ácido nucleico de muestra. De este modo, la composición básica de cualquier amplicón puede incluir el par de cebadores, el complemento del par de cebadores y la región de un ácido nucleico de muestra que se amplificó para producir el amplicón. Una persona versada en la materia comprenderá que la incorporación de las secuencias del par de  
35 cebadores a un amplicón puede sustituir las secuencias nativas en el sitio o punto de unión de cebador, y el complemento de este. En algunas realizaciones, después de amplificar la región diana utilizando los cebadores, los amplicones resultantes que tienen las secuencias de los cebadores se usan para análisis posteriores (por ejemplo, determinación de la composición de las bases). En algunas realizaciones, el amplicón también tiene una longitud que es compatible con posteriores análisis.  
40

45 **[0025]** El término 'amplificación' (o 'amplificado'), en el contexto de los ácidos nucleicos, hace referencia a la producción de múltiples copias de un polinucleótido o una porción del polinucleótido, empezando normalmente a partir de una pequeña cantidad del polinucleótido (por ejemplo, tan poco como una sola molécula del polinucleótido), en la que, normalmente, los productos de amplificación o amplicones son detectables. La amplificación de polinucleótidos abarca diversos procesos químicos y enzimáticos. La producción de múltiples copias de ADN a partir de una copia o unas pocas copias de una molécula de ADN diana o ADN molde durante una reacción en cadena de la polimerasa (PCR, por sus siglas en inglés) o una reacción en cadena de la ligasa (LCR, por sus siglas en inglés) son formas de amplificación. La amplificación no se limita a la mera duplicación de la molécula inicial. Por ejemplo, la producción o generación de múltiples moléculas de ADNc a partir de una cantidad limitada de ARN en una muestra  
50 utilizando una transcripción inversa (RT-PCR, por sus siglas en inglés) es una forma de amplificación. Asimismo, la producción de múltiples moléculas de ARN a partir de una sola molécula de ADN durante el proceso de transcripción también es una forma de amplificación.  
55

60 **[0026]** Tal y como se utiliza en el presente documento, el término 'anclaje' o 'fijación' ('anchor', en inglés) hace referencia a un ácido nucleico que sirve como molde o plantilla para emparejar o hibridar un segundo ácido nucleico. En algunas realizaciones, los anclajes sirven como puntos o sitios de hibridación para los cebadores de secuenciación (por ejemplo, en los métodos de secuenciación de nueva generación).

65 **[0027]** Tal y como se utiliza en el presente documento, el término 'muestra' se utiliza en su sentido más amplio. En un sentido, incluye un cultivo o porción representativa obtenidos de cualquier fuente, incluyendo fuentes biológicas y ambientales. Las muestras biológicas pueden obtenerse de animales (incluyendo humanos) e incluyen fluidos,

sólidos, tejidos y gases. Las muestras biológicas incluyen productos sanguíneos, como plasma, suero y similares. Las muestras ambientales incluyen materiales del medio ambiente, como materia de la superficie, tierra, barro, lodo, biopelículas, agua y muestras industriales. Sin embargo, no debe entenderse que estos ejemplos limitan los tipos de muestras aplicables a la presente invención.

5

### Realizaciones de la tecnología

[0028] Si bien la presente divulgación hace referencia a algunas realizaciones ilustradas, debe entenderse que estas realizaciones se ofrecen a modo de ejemplo y no son limitativas.

10

[0029] Tal y como se ha explicado previamente, los métodos de secuenciación (por ejemplo, los métodos de secuenciación de nueva generación) proporcionan una sensibilidad elevada. Sin embargo, el alto nivel de sensibilidad puede hacer que la detección de ácidos nucleicos contaminantes humanos o microbianos arroje resultados de falsos positivos en los ensayos de detección o diagnóstico.

15

[0030] Por consiguiente, las realizaciones de la presente invención proporcionan métodos y sistemas que utilizan secuencias de mantenimiento conservadas (por ejemplo, secuencias de ADN mitocondrial o ribosómico) para determinar los niveles de contaminación del ácido nucleico en la secuenciación o en otras reacciones de detección del ácido nucleico (por ejemplo, NGS).

20

[0031] En algunas realizaciones, la presencia de secuencias mitocondriales humanas se utiliza para detectar secuencias humanas contaminantes en una muestra. Por ejemplo, en algunas realizaciones, los métodos y sistemas NGS utilizan cebadores o anclajes para ampliar específicamente y secuenciar secuencias mitocondriales humanas. En algunas realizaciones, se utilizan la secuenciación del genoma completo y/o métodos de amplificación, y el número de lecturas atribuibles a las secuencias mitocondriales humanas se identifica usando métodos y sistemas bioinformáticos.

25

[0032] En algunas realizaciones, la presencia de secuencias bacterianas 16S se usa para detectar ácidos nucleicos bacterianos contaminantes en una muestra. Por ejemplo, en algunas realizaciones, los métodos y sistemas NGS utilizan cebadores para amplificar y secuenciar específicamente secuencias bacterianas 16S. En algunas realizaciones, se utilizan la secuenciación del genoma completo y/o métodos de amplificación, y el número de lecturas atribuibles a las secuencias bacterianas 16S se identifica usando métodos y sistemas bioinformáticos. En algunas realizaciones, la secuencia bacteriana 16S es una secuencia 16S de *B. anthracis*.

30

[0033] En algunas realizaciones, el nivel de ácido nucleico atribuible al ácido nucleico de mantenimiento (por ejemplo, determinado por el número de lecturas de secuenciación) se compara con el nivel de un ácido nucleico diana (por ejemplo, determinado por el número de lecturas de secuenciación). La cantidad de ácido nucleico de control en comparación con la cantidad de ácido nucleico diana secuenciado se usa para determinar la presencia o el nivel de contaminación.

35

40

[0034] En algunas realizaciones, se utilizan métodos bioinformáticos para determinar el nivel de ácidos nucleicos de control (por ejemplo, mitocondrial humano o bacteriano 16S) en una muestra. Después, los niveles se comparan con los niveles de ácido nucleico diana en una muestra para determinar el nivel de contaminación. Después, el nivel de contaminación presente se usa para determinar si los resultados del ensayo (por ejemplo, la presencia o ausencia de una secuencia diana dada) son resultados positivos verdaderos, debido a la contaminación, o son sospechosos.

45

### I. Amplificación

[0035] En algunas realizaciones, se amplifican ácidos nucleicos antes de la secuenciación, después de la secuenciación o al mismo tiempo que la secuenciación. En algunas realizaciones, los métodos de amplificación son métodos de amplificación del genoma completo.

50

[0036] Los ejemplos ilustrativos -y no limitativos- de técnicas de amplificación de ácido nucleico incluyen -pero no se limitan a- la reacción en cadena de la polimerasa (PCR), la reacción en cadena de la polimerasa con transcriptasa inversa (RT-PCR), la amplificación mediada por transcripción (TMA), la reacción en cadena de la ligasa (LCR), la amplificación por desplazamiento de cadena (SDA) y la amplificación basada en la secuencia de ácido nucleico (NASBA). Aquellas personas con conocimientos y habilidades comunes en este campo comprenderán que algunas técnicas de amplificación (por ejemplo, la PCR) requieren que el ARN se transcriba inversamente a ADN antes de la amplificación (por ejemplo, RT-PCR), mientras que otras técnicas de amplificación amplifican directamente el ARN (por ejemplo, TMA y NASBA).

55

60

[0037] La reacción en cadena de la polimerasa (Patentes de EE. UU. n<sup>os</sup> 4,683,195, 4,683,202, 4,800,159 y 4,965,188), denominada habitualmente PCR, utiliza múltiples ciclos de desnaturalización, hibridación (también denominada 'anillamiento' o 'alineamiento') de pares de cebadores con cadenas opuestas, y extensión de cebadores para aumentar exponencialmente el número de copias de una secuencia de ácido nucleico diana. En una variante denominada RT-PCR, se usa transcriptasa inversa (RT, por sus siglas en inglés) para obtener un ADN

65

complementario (ADNc) a partir de ARNm y, posteriormente, el ADNc se amplifica mediante PCR para producir múltiples copias de ADN. Para otras combinaciones o modificaciones de la PCR, ver, por ejemplo, las Patentes de EE. UU. n<sup>os</sup> 4,683,195, 4,683,202 y 4,800,159; Mullis et al., 'Meth. Enzymol.', 155: 335 (1987); y Murakawa et al., DNA 7: 287 (1988).

**[0038]** La amplificación mediada por transcripción (Patentes de EE. UU. n<sup>os</sup> 5,480,784 y 5,399,491), denominada habitualmente TMA, sintetiza autocatalíticamente múltiples copias de una secuencia de ácido nucleico diana en condiciones de temperatura, fuerza iónica y pH básicamente constantes en las que múltiples copias de ARN de la secuencia diana producen copias adicionales autocatalíticamente. Ver, por ejemplo, las Patentes de EE. UU. n<sup>os</sup> 5,399,491 y 5,824,518. En una variante descrita en la Publicación de EE. UU. n<sup>o</sup> 20060046265, la TMA incluye opcionalmente el uso de fracciones bloqueantes, fracciones terminantes y otras fracciones modificadoras para mejorar la precisión y la sensibilidad del proceso de TMA.

**[0039]** La reacción en cadena de la ligasa (Weiss, R., Science, 254: 1292 (1991), comúnmente llamada LCR, utiliza dos conjuntos de oligonucleótidos de ADN complementarios que se hibridan con regiones adyacentes del ácido nucleico diana. Los oligonucleótidos de ADN se unen covalentemente mediante una ADN ligasa en ciclos repetidos de desnaturalización, hibridación y ligación térmica para producir un producto detectable de oligonucleótidos bicatenarios ligados.

**[0040]** La amplificación por desplazamiento de cadena, (Walker, G. et al., 'Proc. Natl. Acad. Sci.', EE. UU., 89: 392-396 (1992); Patentes de EE. UU. n<sup>os</sup> 5,270,184 y 5,455,166), comúnmente llamada SDA, utiliza ciclos para hibridar pares de secuencias de cebadores con cadenas opuestas de una secuencia diana, una extensión del cebador en presencia de un dNTPαS para producir un producto dúplex de extensión de cebadores hemifosforotioados, un corte mediado por endonucleasa de un sitio de reconocimiento de endonucleasa de restricción semimodificada y una extensión de cebador mediada por polimerasa del extremo 3' del corte o muesca a fin de desplazar una cadena existente y producir una cadena para la siguiente ronda de hibridación de cebadores, corte y desplazamiento de cadenas, lo cual da como resultado una amplificación geométrica del producto. La SDA termofílica (tSDA) utiliza endonucleasas y polimerasas termofílicas a temperaturas más elevadas básicamente con el mismo método (EP Pat. N<sup>o</sup> 0 684 315).

**[0041]** Otros métodos de amplificación incluyen, por ejemplo: la amplificación basada en secuencias de ácido nucleico (Patente de EE. UU. n<sup>o</sup> 5,130,238), llamada comúnmente NASBA; una que utiliza una ARN replicasa para amplificar la propia molécula de sonda (Lizardi et al., BioTechnol. 6: 1197 (1988)), llamada comúnmente Qβ replicasa; un método de amplificación basado en la transcripción (Kwoh et al., Proc. Natl. Acad. Sci.; EE. UU., 86:1173 (1989)); y una replicación de secuencia autosostenida (Guatelli et al., Proc. Natl. Acad. Sci., EE. UU., 87:1874 (1990)). Para saber más de los métodos de amplificación conocidos, ver Persing, David H., "In Vitro Nucleic Acid Amplification Techniques" en 'Diagnostic Medical Microbiology: Principles and Applications' (Persing et al., Eds.), págs. 51-87 (American Society for Microbiology, Washington, DC, EE. UU. (1993)).

**[0042]** En algunas realizaciones, la amplificación es un método de amplificación isotérmico. En algunas realizaciones, los métodos de amplificación son la amplificación de fase sólida, la amplificación 'Polony' ('colonia de polimerasa'), la amplificación de colonia, la PCR de emulsión, la RCA de cuentas, la SDA superficial, etc., como bien sabrá una persona versada en la materia. En algunas realizaciones, se utilizan métodos de amplificación que dan como resultado la amplificación de moléculas de ADN libre en solución o unidas a una matriz apropiada solamente por un extremo de la molécula de ADN. En algunas realizaciones, se utilizan métodos que se basan en una PCR en puente, de manera que ambos cebadores de PCR quedan unidos a una superficie (ver, por ejemplo, WO 2000/018957, U.S. 7,972,820; 7,790,418 y Adessi et al., 'Nucleic Acids Research' (2000): 28(20): E87). En algunos casos, los métodos de la invención pueden producir una 'tecnología de colonia de polimerasa', o 'polony' (por su nombre en inglés), en referencia a una amplificación múltiple que mantiene el agrupamiento espacial de amplicones idénticos (ver 'Harvard Molecular Technology Group' y la web 'Lippper Center for Computational Genetics'). Estos incluyen, por ejemplo, las 'polonys' in situ (Mitra y Church, 'Nucleic Acid Research' 27, e34, 15 e Diciembre de 1999), la amplificación de círculo rodante (RCA) (Lizardi et al., Nature Genetics 19, 225, Julio de 1998), la PCR en puente (Patente de EE. UU. n<sup>o</sup> 5,641,658), la PCR picotituladora (Leamon et al., 'Electrophoresis' 24, 3769, Noviembre de 2003) y la PCR de emulsión (Dressman et al., PNAS 100, 8817, 22 de Julio de 2003).

**[0043]** En algunas realizaciones se utiliza la amplificación del genoma completo. Los primeros métodos de amplificación del genoma completo se describieron en 1992 y estaban basados en los principios de la reacción en cadena de la polimerasa. Zhang y sus colaboradores (Zhang, L., et al.; 'Proc. Natl. Acad. Sci.', EE. UU., 1992, 89: 5847-5851) desarrollaron la técnica de PCR de extensión de cebadores (PEP, por sus siglas en inglés) y Tellenius y sus colaboradores (Tellenius et al., Genomics. 1992, 13(3):718-25) desarrollaron el método de reacción en cadena de polimerasa de oligonucleótidos degradados (DOP-PCR, por sus siglas en inglés). El PEP incluye un número elevado de ciclos de PCR, normalmente usando polimerasa Taq y cebadores aleatorios de 15 bases que se hibridan a una temperatura de rigurosidad baja. DOP-PCR es un método que normalmente usa polimerasa Taq y oligonucleótidos semidegradados que se unen a una temperatura de hibridación baja en -aproximadamente- un millón de sitios del genoma humano. Después de los primeros ciclos, se sucede un gran número de ciclos con una temperatura de hibridación más elevada, lo cual solo permite la amplificación de los fragmentos que se marcaron en

el primer paso.

[0044] La amplificación por desplazamientos múltiples (o MDA; también conocida como amplificación por desplazamiento de cadena; SDA) es un método isotérmico que no está basado en PCR, sino en la hibridación de hexámeros aleatorios con ADN desnaturalizado, seguida de la síntesis por desplazamiento de cadena a temperatura constante (Blanco et al., 1989, J. Biol. Chem. 264:8935-40; Dean, F.B. et al. (2002) 'Comprehensive human genome amplification using multiple displacement amplification'; Proc. Natl. Acad. Sci. USA 99,5261; y Van, J. et al., (2004) 'Assessment of multiple displacement amplification in molecular epidemiology'. Biotechniques 37, 136)). Se ha aplicado a pequeñas muestras de ADN genómico, dando como resultado la síntesis de ADN con un peso molecular elevado con un sesgo de representación de secuencias limitado (Lizardi et al., 'Nature Genetics'; 1998, 19, 225-232; Dean et al., 'Proc. Natl. Acad. Sci.'; EE. UU., 2002, 99, 5261-5266). A medida que se sintetiza ADN mediante el desplazamiento de cadenas, se sucede un número cada vez más alto de operaciones de cebado, de manera que se forma una red de estructuras de ADN hiper-ramificadas. La reacción puede catalizarse mediante la ADN polimerasa Phi29 o mediante grandes fragmentos de la ADN polimerasa Bst. La ADN polimerasa Phi29 tiene una actividad correctora que da como resultado unos índices de error 100 veces inferiores a la polimerasa Taq.

[0045] Los ejemplos de polimerasas de ácidos nucleicos que son adecuadas para usarse en las realizaciones de la presente invención incluyen -pero no se limitan a- la ADN polimerasa (fragmento de Klenow, ADN polimerasa T4), las polimerasas de ADN termoestables (Perler F. B. et al., 'Adv. Protein Chem.', 1996, 48:377-435) identificadas y clonadas en diversas bacterias termoestables (como las ADN polimerasas Taq, VENT, Pfu, Tfi), así como sus derivados genéticamente modificados (TaqGold, VENTexo, Pfu exo). Preferiblemente, la polimerasa de ácido nucleico utilizada para la extensión de cebadores de colonia es estable a una temperatura en la que la hibridación del cebador y la plantilla es lo suficientemente específica como para evitar amplificaciones falsas o incompletas de la plantilla.

[0046] Preferiblemente, la solución de amplificación contiene, como precursores de nucleótidos, desoxirribonucleótidos trifosfato, por ejemplo dATP, dTTP, dCTP, dGTP, que existen de forma natural o no natural, por ejemplo modificados con un grupo fluorescente o radioactivo. Se ha desarrollado una gran variedad de ácidos nucleicos modificados sintéticamente para métodos químicos y biológicos a fin de aumentar la detectabilidad y/o la diversidad funcional de los ácidos nucleicos. Estas moléculas funcionalizadas/modificadas pueden ser completamente compatibles con las enzimas naturales polimerizantes, de manera que se mantienen el emparejamiento de bases y las propiedades de replicación de los equivalentes naturales, tal y como se ha analizado recientemente (Thum O et al., 'Angew. Chem. Int.', Ed. 2001, 40 (21):3990-3993).

[0047] Otros componentes de la solución de amplificación se añaden de manera consecutiva con la elección de la polimerasa de ácido nucleico, y básicamente se corresponden con los compuestos en este campo que se sabe son eficaces para ayudar en la actividad de cada polimerasa. La concentración de compuestos como el dimetilsulfóxido (DMSO), la albúmina de suero bovino (BSA), el Tritón X-100 o el MgCl<sub>2</sub> es muy conocida en la técnica anterior y se sabe que es importante para obtener una amplificación óptima; así, el técnico u operador puede ajustar fácilmente estas concentraciones a los métodos de la presente invención tomando como referencia los ejemplos expuestos previamente.

## II. Secuenciación

[0048] En algunas realizaciones, los métodos de secuenciación del ácido nucleico se utilizan para la detección. En algunas realizaciones, la tecnología que se proporciona en el presente documento puede usarse en tecnologías de secuenciación de segunda generación (también conocidas como 'de nueva generación', 'de próxima generación' o 'Next-gen'), de tercera generación (también conocidas como 'Next-next-gen') o de cuarta generación (también conocidas como 'N3-Gen'), incluyendo -pero sin limitarse a- la pirosecuenciación, la secuenciación por ligación, la secuenciación de una sola molécula, la secuenciación mediante síntesis (SBS), la secuenciación clónica paralela masiva, la SBS paralela masiva de una sola molécula, la secuenciación paralela masiva de una sola molécula en tiempo real, la tecnología de nanoporos paralela masiva de una sola molécula en tiempo real, etc. Morozova y Marra ofrecen un análisis de estas tecnologías en 'Genomics', 92: 255 (2008). Aquellas personas versadas en la materia comprenderán que, puesto que el ARN es menos estable en la célula y más propenso a los ataques de nucleasa, en los experimentos normalmente el ARN se transcribe inversamente a ADN antes de la secuenciación.

[0049] En este campo se conocen diversas técnicas de secuenciación del ADN, incluyendo las metodologías de secuenciación basadas en la fluorescencia (ver, por ejemplo, Birren et al., 'Genome Analysis: Analyzing DNA', 1, Cold Spring Harbor, Nueva York, Estados Unidos). En algunas realizaciones, la tecnología puede usarse en técnicas de secuenciación automatizada conocidas en este campo. En algunas realizaciones, la presente tecnología puede usarse en la secuenciación paralela de amplicones divididos (Publicación PCT N°: WO2006084132, de Kevin McKernan et al.). En algunas realizaciones, la tecnología puede usarse para la secuenciación de ADN mediante extensión de oligonucleótidos en paralelo (ver, por ejemplo, la Patente de EE. UU. N° 5,750,341 de Macevicz et al., y la Patente de EE. UU. N° 6,306,597 de Macevicz et al.).

**[0050]** Los ejemplos adicionales de técnicas de secuenciación que pueden usarse en la tecnología incluyen la tecnología 'Church polony' (Mitra et al., 2003, 'Analytical Biochemistry', 320, 55-65; Shendure et al., 2005 Science 309, 1728-1732; Patente de EE. UU. N.º. 6,432,360, Patente de EE. UU. N.º. 6,485,944, Patente de EE. UU. N.º. 6,511,803), la tecnología de pirosecuenciación picotituladora 454 (Margulies et al., 2005, Nature, 437, 376-380; EE. UU. 20050130173), la tecnología de adición de bases solas Solexa (Bennett et al., 2005, 'Pharmacogenomics', 6, 373-382; Patente de EE. UU. N.º. 6,787,308; Patente de EE. UU. N.º. 6,833,246), la tecnología de secuenciación de firmas paralela masiva de Lynx (Brenner et al. (2000). 'Nat. Biotechnol.', 18:630-634; Patente de EE. UU. N.º. 5,695,934; Patente de EE. UU. N.º. 5,714,330), y la tecnología de colonia PCR de Adessi (Adessi et al. (2000)., 'Nucleic Acid Res.', 28, E87; WO 00018957).

**[0051]** Los métodos de secuenciación de nueva generación (NGS, por sus siglas en inglés) comparten la característica común de estrategias de alto rendimiento y masivamente paralelas, con un objetivo de costes más bajos en comparación con los métodos de secuenciación más viejos (ver, por ejemplo, Voelkerding et al., 'Clinical Chem.', 55: 641-658, 2009; MacLean et al., 'Nature Rev. Microbiol.', 7: 287-296). Los métodos de NGS pueden dividirse a grandes rasgos entre los que normalmente utilizan la amplificación de plantillas o moldes y los que no. Los métodos que requieren amplificación incluyen la pirosecuenciación comercializada por Roche como las plataformas de tecnología 454 (por ejemplo, GS 20 y GS FLX), la plataforma Solexa comercializada por Illumina, y la plataforma 'Supported Oligonucleotide Ligation and Detection' (Detección y ligación de oligonucleótidos de apoyo o 'SOLiD') comercializada por Applied Biosystems. Los enfoques de no amplificación, también conocidos como 'secuenciación de una sola molécula', se ejemplifican mediante la plataforma HeliScope, comercializada por Helicos Bio-Sciences, y las plataformas emergentes comercializadas por VisiGen, Oxford Nanopore Technologies Ltd., Life Technologies/Ion Torrent y Pacific Biosciences, respectivamente.

**[0052]** En la pirosecuenciación (Voelkerding et al., 'Clinical Chem.', 55: 641-658, 2009; MacLean et al., 'Nature Rev. Microbiol.', 7: 287-296; Patente de EE. UU. n.º 6,210,891; Patente de EE. UU. n.º 6,258,568), la plantilla de ADN se fragmenta, se repara por los extremos, se liga a los adaptadores y se amplifica clonalmente 'in situ' capturando moléculas de plantilla únicas con cuentas o bolitas que tienen oligonucleótidos complementarios a los adaptadores. Cada cuenta o bolita, que lleva un único tipo de plantilla, se compatimenta en una microvesícula de agua en aceite, y la plantilla se amplifica clonalmente usando una técnica llamada PCR de emulsión. La emulsión se interrumpe tras la amplificación y las bolitas de depositan en los pocillos individuales de una placa picotituladora que funciona como una celda de flujo durante las reacciones de secuenciación. En la celda de flujo se produce una introducción ordenada e iterativa de cada uno de los cuatro reactivos dNTP en presencia de enzimas de secuenciación y reporteros luminiscentes como la luciferasa. En caso de que se añada un dNTP adecuado añ extremo 3' del cebador de secuenciación, la consiguiente producción de ATP provoca un estallido de luminiscencia en el pocillo, el cual se graba usando una cámara CCD. Es posible obtener longitudes de lectura mayores o iguales a 400 bases, y pueden obtenerse lecturas de secuencia de 10<sup>6</sup>, lo que da como resultado una secuencia de hasta 500 millones de pares de bases (Mb).

**[0053]** En la plataforma Solexa/Illumina (Voelkerding et al., 'Clinical Chem.', 55: 641-658, 2009; MacLean et al., 'Nature Rev. Microbiol.', 7: 287-296; Patente de EE. UU. n.º 6,833,246; Patente de EE. UU. n.º 7,115,400; Patente de EE. UU. n.º 6,969,488), los datos de secuenciación se producen en forma de lecturas de menos longitud. En este método, el ADN fragmentado monocatenario se repara por los extremos para obtener extremos romos 5'-fosforilados, y después se realiza una adición mediada de Klenow de una sola base A al extremo 3' de los fragmentos. La adición de A facilita la adición de oligonucleótidos adaptadores de T sobrantes, que se usan posteriormente para capturar las moléculas adaptadoras de plantillas en la superficie de una celda de flujo que está salpicada de anclajes de oligonucleótidos. El anclaje se usa como un cebador de PCR, pero, debido a la longitud de la plantilla y su proximidad a otros oligonucleótidos de anclaje cercanos, la extensión por PCR hace que la molécula se 'arquee' y se hibride con un oligonucleótido de anclaje adyacente y forme una estructura de puente en la superficie de la célula de flujo. Estos bucles de ADN se desnaturalizan y se rompen. Después, las hebras o cadenas delanteras se secuencian con terminadores de tinción reversibles. La secuencia de nucleótidos incorporados se determina mediante la detección de fluorescencia de post-incorporación, de manera que cada fluoróforo y bloque se elimina antes del siguiente ciclo de adición de dNTP. La longitud de las lecturas de secuencia va desde 36 nucleótidos hasta más de 50 nucleótidos, de manera que el resultado total es superior a 1000 millones de pares de nucleótidos por cada operación analítica.

**[0054]** Secuenciar moléculas de ácido nucleico usando tecnología SOLiD (Voelkerding et al., 'Clinical Chem.', 55:641-658, 2009; MacLean et al., 'Nature Rev. Microbiol.', 7: 287-296; Patente de EE. UU. n.º 5,912,148; Patente de EE. UU. n.º 6,130,073) también incluye la fragmentación del molde o plantilla, la ligación a los adaptadores de nucleótidos, la unión a las cuentas o bolitas y la amplificación clonal mediante PCR de emulsión. Después de esto, las cuentas o bolitas que portan la plantilla se inmovilizan en una superficie derivatizada de una 'celda de flujo' de vidrio y se hibrida o aparea un cebador complementario al oligonucleótido adaptador. Sin embargo, en vez de utilizar este cebador para la extensión 3', se usa para proporcionar un grupo fosfato 5' para la ligación con sondas de interrogación que contienen dos bases específicas para pruebas seguidas de 6 bases degeneradas y uno de los cuatro marcadores fluorescentes. En el sistema SOLiD, las sondas de interrogación tienen 16 combinaciones posibles de las dos bases en el extremo 3' de cada sonda, y uno de los cuatro fluorescentes en el extremo 5'. El color del fluoróforo o fluorescente -y, por lo tanto, la identidad de cada sonda- se corresponde con unos esquemas de



codificación de espacio de color especificados. Se desarrollan múltiples rondas (normalmente 7) de hibridación, ligación y detección fluorescente de las sondas, seguidas de la desnaturalización y, después, una segunda ronda de secuenciación que utiliza un cebador que difiere en una base respecto al cebador inicial. De esta manera, la secuencia de plantilla puede reconstruirse computacionalmente y las bases de la plantilla se interrogan dos veces, lo cual da como resultado una mayor precisión. La lectura de secuencia tiene una longitud promedio de 35 nucleótidos, y el resultado general es superior a 4000 millones de bases por cada operación de secuenciación.

**[0055]** En algunas realizaciones, la tecnología puede usarse en la secuenciación de nanoporos (ver, por ejemplo, Astier et al., 'J. Am. Chem. Soc.', 8 de Febrero de 2006; 128(5):1705-10). La teoría tras la secuenciación por nanoporos está relacionada con lo que sucede cuando se sumerge un nanoporo en un fluido conductor y se aplica un potencial (voltaje). En estas condiciones, puede observarse una ligera corriente eléctrica debida a la conducción de iones a través del nanoporo, y la cantidad de corriente es extremadamente sensible al tamaño del nanoporo. Cuando cada base de un ácido nucleico pasa a través del nanoporo, esto provoca un cambio en la magnitud de la corriente a través del nanoporo que es distinto en cada una de las cuatro bases, lo cual hace posible determinar la secuencia de la molécula de ADN.

**[0056]** En algunas realizaciones, la tecnología puede usarse con HeliScope de Helicos BioSciences (Voelkerding et al., 'Clinical Chem.', 55: 641-658, 2009; MacLean et al., 'Nature Rev. Microbiol.', 7: 287-296; Patente de EE. UU. n° 7,169,560; Patente de EE. UU. n° 7,282,337; Patente de EE. UU. n° 7,482,120; Patente de EE. UU. n° 7,501,245; Patente de EE. UU. n° 6,818,395; Patente de EE. UU. n° 6,911,345; Patente de EE. UU. n° 7,501,245). El ADN molde se fragmenta y se poliadenila en el extremo 3', de manera que la última adenosina porta un marcador fluorescente. Los fragmentos molde poliadenilados y desnaturalizados se ligan con poli(dT) oligonucleótidos en la superficie de una celda de flujo. Las localizaciones físicas iniciales de las moléculas molde capturadas se graban con una cámara CCD y, después, el marcador se escinde y se elimina. La secuenciación se obtiene mediante la adición de polimerasa y la adición en serie de reactivos dNTP marcados fluorescentemente. Las incorporaciones provocan señales fluorescentes que corresponden al dNTP, y la señal se captura con una cámara CCD antes de cada ronda de adición de dNTP. La longitud de las lecturas de secuencia es de entre 25 y 50 nucleótidos, y el resultado general es superior a 1000 millones de pares de nucleótidos por cada operación de análisis.

**[0057]** La tecnología de torrente de iones es un método de secuenciación de ADN basado en la detección de los iones de hidrógeno que se liberan durante la polimerización del ADN (ver, por ejemplo, Science 327(5970): 1190 (2010); Solicitudes de Patente de EE. UU. n°s 20090026082, 20090127589, 20100301398, 20100197507, 20100188073 y 20100137143). Un micropocillo contiene una cadena de ADN molde que se va a secuenciar. Bajo la capa de micropocillos hay un sensor de iones hipersensible ISFET. Todas las capas están contenidas en un chip semiconductor CMOS, similar al que se utiliza en la industria de la electrónica. Cuando se incorpora un dNTP a la cadena complementaria en crecimiento, se libera un ion de hidrógeno, lo que activa un sensor de iones hipersensible. Si hay repeticiones de homopolímeros presentes en la secuencia molde, se incorporarán múltiples moléculas de dNTP en un solo ciclo. Esto provoca un número correspondiente de hidrógenos liberados y una señal electrónica proporcionalmente más alta. Esta tecnología difiere de otras tecnologías de secuenciación por el hecho de que no se utiliza ningún nucleótido modificado ni ningún sistema óptico. La precisión por base del secuenciador Ion Torrent es de ~99,6% por cada 50 lecturas de base, de manera que se generan ~100 Mb en cada tanda u operación. La longitud de lectura es de 100 pares de bases. La precisión para las repeticiones de homopolímeros de 5 repeticiones de longitud es de ~98%. Los beneficios de la secuenciación semiconductor de iones son una alta velocidad de secuenciación y unos bajos costes iniciales y de operación.

**[0058]** La tecnología puede usarse en otro enfoque de secuenciación de ácido nucleico desarrollado por Stratos Genomics, Inc., e incluye el uso de 'expandómeros'. Normalmente, este proceso de secuenciación incluye proporcionar una cadena o hebra hija producida mediante una síntesis dirigida por plantilla. Generalmente, la cadena hija incluye diversas subunidades unidas en una secuencia correspondiente a una secuencia de nucleótidos contiguos de un ácido nucleico diana (o una parte de este) en la que las subunidades individuales comprenden una unión, al menos una sonda o residuo de nucleobase, y al menos un enlace escindible de forma selectiva. El o los enlaces escindibles de forma selectiva se rompen para obtener un expandómero con una longitud mayor que las diversas subunidades de la cadena hija. Normalmente, el expandómero incluye las uniones (o 'tethers', en inglés) y los elementos reporteros para analizar la información genética en una secuencia correspondiente a la secuencia de nucleótidos contiguos del ácido nucleico diana o una porción de este. Posteriormente, se detectan los elementos reporteros del expandómero. Los detalles adicionales de los enfoques basados en los expandómeros se describen, por ejemplo, en la Publicación de Patente de EE. UU. n° 20090035777, titulada "High Throughput Nucleic Acid Sequencing by Expansion", presentada el 19 de Junio de 2008, que se incorpora en su totalidad en el presente documento.

**[0059]** Otros métodos emergentes de secuenciación de una sola molécula incluye la secuenciación en tiempo real mediante síntesis utilizando una plataforma VisiGen (Voelkerding et al., 'Clinical Chem.', 55: 641-58, 2009; Patente de EE. UU. n° 7,329,492; Publicación de Patente de EE. UU. n° 20070250274 A1; Publicación de Patente de EE. UU. n° 2008241951 A1), en la que el molde o plantilla de ADN cebado e inmovilizado se somete a una extensión de cadenas usando una polimerasa modificada fluorescentemente y moléculas aceptoras fluorescentes, lo cual da como resultado una transferencia de energía de resonancia de Förster (FRET) detectable tras la adición de

nucleótidos.

### III. Análisis de datos

5 **[0060]** En algunas realizaciones, se utiliza un programa de análisis computerizado para convertir los datos en bruto generados mediante el ensayo de detección (por ejemplo, el nivel de contaminación humana o los ácidos nucleicos microbianos) en datos con valor predictivo para el usuario final (por ejemplo, personal médico). El usuario puede acceder a los datos predictivos mediante cualquier medio apropiado. Así, en algunas realizaciones preferidas, la presente invención proporciona el beneficio adicional de que el usuario, que probablemente no está familiarizado con la genética o la biología molecular, no tenga por qué comprender los datos en bruto. Los datos se presentan directamente al usuario en su forma más útil. Luego, el usuario puede utilizar la información inmediatamente para obtener más información útil (por ejemplo, en el diagnóstico médico, la investigación o el rastreo.

15 **[0061]** La presente invención contempla cualquier método que pueda recibir, procesar y transmitir la información a y desde los laboratorios que realicen los ensayos, los informes informativos, el personal médico y los sujetos. Por ejemplo, en algunas realizaciones de la presente invención, se obtiene una muestra (por ejemplo, sangre, células o tejido) de un sujeto y se envía a un servicio de perfiles (por ejemplo, un laboratorio de unas instalaciones médicas, un negocio de perfiles genómicos, etc.) situado en cualquier parte del mundo (por ejemplo, en un país diferente al país en el que reside el sujeto o en el que se va a usar la información en última instancia) para generar los datos en bruto. Cuando la muestra comprenda un tejido u otra muestra biológica, el sujeto puede visitar un centro médico para que obtengan la muestra y se envíe al centro de perfiles, o el sujeto puede recoger la muestra él mismo (por ejemplo, un frotis de mejilla) y enviarla directamente al centro de perfiles. Cuando la muestra comprenda información biológica determinada previamente, el sujeto puede enviar la información directamente al centro de perfiles (por ejemplo, un ordenador puede escanear una hoja o tarjeta de información que contenga la información y transmitir los datos a un ordenador del centro de perfiles mediante un sistema de comunicación electrónico). Una vez que el servicio de perfiles la ha recibido, la muestra se procesa y se obtiene un perfil (por ejemplo, el nivel de contaminación) específico para la información deseada para el sujeto.

30 **[0062]** Después, los datos se preparan en un formato que sea adecuado para que lo interprete el usuario final (por ejemplo, personal médico). Por ejemplo, en vez de proporcionar datos en bruto, el formato preparado puede ofrecer una conclusión o un análisis o evaluación (por ejemplo, el nivel de contaminación de los ácidos nucleicos; la presencia de ácido nucleico microbiano o ácido nucleico humano variable, etc.) al sujeto. Los datos se pueden mostrar al usuario mediante cualquier método adecuado. Por ejemplo, en algunas realizaciones, el servicio de perfiles elabora un informe que puede imprimirse para el usuario o puede mostrarse en la pantalla de un ordenador.

35 **[0063]** En algunas realizaciones, la información se analiza primero en una instalación médica local o regional. Después, los datos en bruto se envían a una instalación central de procesamiento para posteriores análisis y/o para convertir los datos en bruto en información útil para el usuario. La instalación central de procesamiento ofrece las ventajas de la privacidad (todos los datos se guardan en una instalación central con protocolos de seguridad uniformes), la velocidad y la uniformidad de los análisis de datos. Así, la instalación central de procesamiento puede controlar el destino de los datos tras el tratamiento del sujeto. Por ejemplo, si utiliza un sistema de comunicación electrónico, la instalación central puede proporcionar información al usuario final, el sujeto o los investigadores.

### IV. Kits y sistemas

45 **[0064]** En algunos aspectos, la presente divulgación proporciona kits y sistemas para la amplificación y/o el análisis de ácidos nucleicos. En algunos aspectos, los kits incluyen los reactivos necesarios, suficientes o apropiados para el análisis y la detección de secuencias de ácido nucleico contaminantes (por ejemplo, cebadores, sondas, anclajes, soportes sólidos, reactivos, controles, instrucciones, etc.). Por ejemplo, en algunos aspectos, los kits contienen cebadores y anclajes para la amplificación y la secuenciación de muestras (por ejemplo, muestras potencialmente contaminadas). En algunos aspectos, los kits incluyen software de análisis (por ejemplo, para analizar los datos de secuenciación y determinar el nivel de ácidos nucleicos contaminantes).

55 **[0065]** En algunos aspectos, los kits contienen uno o más recipientes o contenedores que contienen reactivos, cebadores, sondas, anclajes, soportes sólidos, 'búffers' o amortiguadores y similares. En algunos aspectos, cada componente del kit está embalado en un contenedor separado. En algunos aspectos, los contenedores se embalan y/o transportan en el mismo kit o en la misma caja para usarse conjuntamente. En algunos aspectos, uno o más componentes del kit se transportan y/o embalan de forma separada.

60 **[0066]** En algunos aspectos, los sistemas incluyen dispositivos automáticos para el manejo de muestras y reactivos (por ejemplo, robóticos).

### EXPERIMENTO

65 **[0067]** Los siguientes ejemplos sirven para ilustrar algunas realizaciones preferidas y también aspectos preferidos de la presente invención y no debe interpretarse que limitan el alcance de esta.

**Ejemplo 1**

**[0068]** Secuencias mitocondriales humanas como marcador de contaminación

5 **[0069]** Se usó NextGENe (Software NGS) para buscar secuencias mitocondriales humanas en conjuntos de datos que podían contener ADN humano.

10 **[0070]** El conjunto de datos de análisis pertenecía a una muestra de suero humano disminuido que se sometió a WGA y NGS y dio positivo en herpes.

15 **[0071]** La Figura 1 muestra los resultados de NGS utilizando una secuencia mitocondrial humana para detectar la presencia de ADN humano. Se hallaron muchos alineamientos con secuencias mitocondriales humanas en el conjunto de datos de la muestra de suero humano disminuido que había dado positivo en herpes.

20 **[0072]** Secuencias bacterianas 16S como marcador de contaminación

**[0073]** Se usó NextGENe (Software NGS) para buscar secuencias bacterianas 16S en conjuntos de datos que podían contener contaminación de ADN bacteriano desconocido.

25 **[0074]** El conjunto de datos de 10 fg de ADN de *Bacillus cereus* se sometió a WGA y NGS. Se usaron once secuencias 16S diferentes de *B.anthraxis* para detectar secuencias bacterianas contaminantes.

30 **[0075]** Los resultados se muestran en la Figura 2. La Figura 2 muestra que se hallaron muchos alineamientos de secuencias 16S y secuencias 16S de *B.anthraxis* en el conjunto de datos de *B. cereus*.

35 **[0076]** Si bien la tecnología se ha descrito en relación con las realizaciones ejemplares específicas, debe entenderse que la invención, tal y como se reivindica, no debe limitarse erróneamente a dichas realizaciones específicas. De hecho, se pretende que las diversas modificaciones de las maneras de llevar a cabo la invención, que resultan evidentes para aquellas personas versadas en los campos pertinentes, estén dentro del alcance de las siguientes reivindicaciones.

40

45

50

55

60

65

**REIVINDICACIONES**

5 **1.** Un método para detectar ácido nucleico humano contaminante en una muestra, de manera que el método comprende:

a) secuenciar:

10 i) un ácido nucleico humano diana de la mencionada muestra, de manera que el ácido nucleico diana es un ácido nucleico variable; y

ii) un ácido nucleico mitocondrial humano, de manera que el mencionado ácido nucleico mitocondrial humano proviene del mencionado ácido nucleico humano contaminante y de la mencionada muestra;

en una muestra de ácido nucleico para obtener un conjunto de datos de secuenciación;

15 b) determinar el número de lecturas de secuenciación atribuibles a los mencionados ácidos nucleicos mitocondriales humanos y el número de lecturas de secuenciación atribuibles al mencionado ácido nucleico diana en el mencionado conjunto de datos de secuenciación;

20 c) comparar el número de lecturas de secuenciación atribuibles a los mencionados ácidos nucleicos mitocondriales humanos respecto al número de lecturas de secuenciación atribuibles al mencionado ácido nucleico diana; y

d) determinar la presencia de ácido nucleico humano contaminante en la mencionada muestra basándose en la mencionada comparación cuando se identifica un número mayor de lecturas de secuenciación atribuibles a los ácidos nucleicos mitocondriales humanos en relación con el ácido nucleico diana.

25 **2.** Un método para detectar ácido nucleico bacteriano contaminante en una muestra, de manera que el método comprende:

a) secuenciar:

30 i) un ácido nucleico diana; y

ii) un ácido nucleico bacteriano 16S, de manera que el mencionado ácido nucleico bacteriano 16S proviene del mencionado ácido nucleico bacteriano contaminante y de la mencionada muestra;

en una muestra de ácido nucleico para obtener un conjunto de datos de secuenciación;

35 b) determinar el número de lecturas de secuenciación atribuibles a los mencionados ácidos nucleicos bacterianos 16S y el número de lecturas de secuenciación atribuibles al mencionado ácido nucleico diana en el mencionado conjunto de datos de secuenciación;

40 c) comparar el número de lecturas de secuenciación atribuibles a los mencionados ácidos nucleicos bacterianos 16S respecto al número de lecturas de secuenciación atribuibles al mencionado ácido nucleico diana; y

d) determinar la presencia de ácido nucleico bacteriano contaminante en la mencionada muestra basándose en la mencionada comparación cuando se identifica un número mayor de lecturas de secuenciación atribuibles a los ácidos nucleicos bacterianos 16S en relación con el ácido nucleico diana.

45 **3.** El método de las reivindicaciones 1 o 2, de manera que la mencionada determinación de los niveles de los mencionados ácidos nucleicos diana y de control se realiza utilizando un procesador informático y software informático.

50 **4.** El método de la reivindicación 1, de manera que el mencionado ácido nucleico diana es una variante, supresión o amplificación de una secuencia de ácido nucleico.

**5.** El método de cualquiera de las reivindicaciones 1 a 4, de manera que la mencionada secuenciación es una secuenciación de nueva generación (o de próxima generación).

55 **6.** El método de cualquiera de las reivindicaciones 1 a 5, de manera que los mencionados ácidos nucleicos diana y de control se amplifican antes de la secuenciación.

**7.** El método de la reivindicación 6, de manera que la mencionada amplificación comprende la amplificación del genoma completo.

60

65

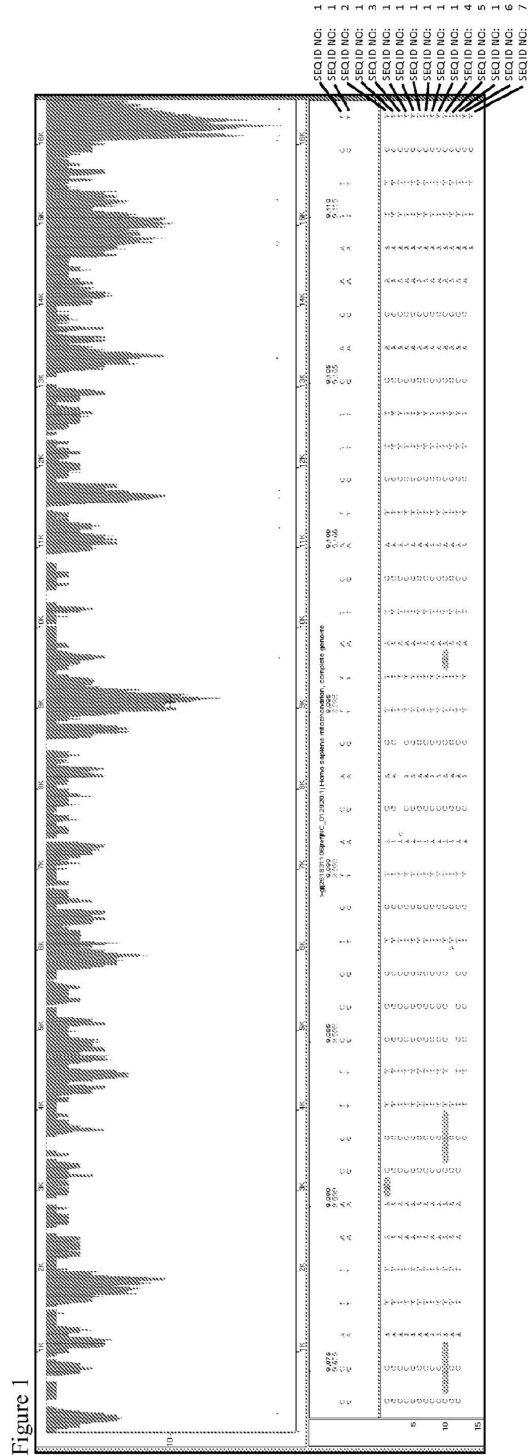


Figura 1

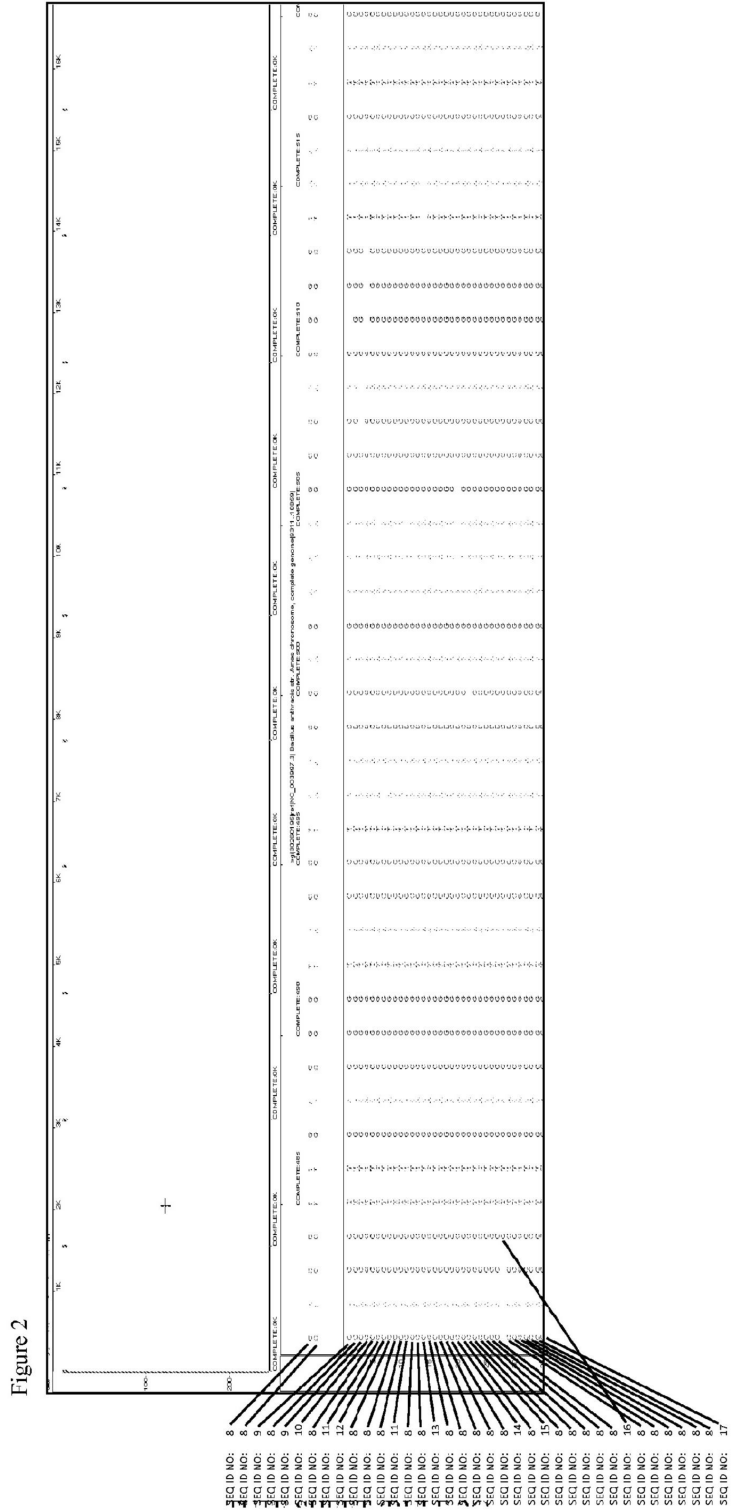


Figura 2