

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 632 538**

51 Int. Cl.:

C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **10.04.2001 E 10012384 (3)**

97 Fecha y número de publicación de la concesión europea: **24.05.2017 EP 2278030**

54 Título: **Métodos para el estudio y análisis genético de poblaciones**

30 Prioridad:

10.04.2000 US 196063 P
11.04.2000 US 196258 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
14.09.2017

73 Titular/es:

TAXON BIOSCIENCES, INC. (100.0%)
3150 Paradise Drive
Tiburon, CA 94920, US

72 Inventor/es:

ASHBY, MATTHEW

74 Agente/Representante:

PONS ARIÑO, Ángel

Observaciones:

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 632 538 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 el estudio y análisis genético de poblaciones

5 **Campo técnico de la invención**

La presente invención se refiere a métodos para realizar estudios de la diversidad genética de una población. La invención también se refiere a métodos para realizar análisis genéticos de una población. Adicionalmente, la invención se refiere a métodos para la creación de bases de datos que comprenden la información del estudio y las bases de datos creadas por estos métodos. La invención también se refiere a métodos para analizar la información para correlacionar la presencia de marcadores de ácido nucleico con parámetros deseados en una muestra. Estos métodos se aplican en los campos de exploración geoquímica, agricultura, biorremediación, análisis ambiental, microbiología clínica, ciencia forense y medicina.

15 **Antecedentes de la invención**

Anteriormente los microbios se han utilizado como biosensores para identificar productos químicos en el ambiente. Por ejemplo, los microbios se han utilizado como biosensores para detectar la presencia de nitratos (Larsen, L. H. *et al.*, 1997, A microscale NO₃ biosensor for environmental applications. *Anal. Chem.* 69: 3527-3531), metales (Virta M. *et al.*, 1998, Bioluminescence- based metal detectors. *Methods Mol. Biol.* 102: 219-229) y de diversos hidrocarburos (Sticher P. *et al.*, 1997, Development and characterization of a whole-cell bioluminescent sensor for bioavailable middle-chain alkanes in contaminated groundwater samples. *Appl. Environ. Microbiol.* 63(10): 4053-4060). Sin embargo, en estos ejemplos, los microbios indicadores no son especies naturales, sino más bien el producto de manipulaciones recombinantes diseñadas para aplicaciones específicas. Estas modificaciones implican el acoplamiento de la maquinaria detectora de nutrientes de cepas bacterianas bien caracterizadas con genes indicadores para ayudar a la identificación. Sin embargo, esta estrategia está limitada por la diversidad metabólica de algunas cepas bacterianas bien caracterizadas. En cambio, el gran y diverso grupo de microbios en el medio representa una fuente de biosensores para una serie más grande de aplicaciones que las actualmente existentes. Por tanto, hay una necesidad de identificar y utilizar otros microbios, especialmente los encontrados *in situ*, como biosensores.

Los microbios también tienen un importante impacto sobre la salud y la medicina. Se han realizado estimaciones de que puede haber diez veces el número de células microbianas asociadas con el cuerpo humano en comparación con las células humanas. Muchas poblaciones de células microbianas que están asociadas al cuerpo humano desempeñan una función beneficiosa en el mantenimiento de la salud. Por ejemplo, la microflora intestinal es importante para realizar una digestión y absorción de nutrientes adecuadas y para la producción de determinados factores, incluyendo algunas vitaminas. En general, el sistema inmunitario humano puede mantener el control de las poblaciones bacterianas del cuerpo humano y prevenir el crecimiento excesivo de poblaciones microbianas beneficiosas e infección por poblaciones microbianas perjudiciales. No obstante, la lista de enfermedades humanas que actualmente se atribuye a patógenos microbianos está aumentando. Sin embargo, casi toda la información con respecto a las relaciones ente los microbios y las enfermedades humanas, se ha incrementado a partir de estrategias que requieren el cultivo de especies microbianas.

Dos ejemplos de enfermedades en las que se identificaron agentes causantes a través de métodos moleculares incluyen, la angiomatosis bacilar (Relman, D. A. *et al.*, 1990, *New Engl. J. Med.* 323: 1573) y la enfermedad de Whipple (Wilson, K. H. *et al.*, 1991, *Lancet* 338: 474). Además, los aspectos centrales de la aterosclerosis son coherentes con la inflamación que se produce debido a infección. Se han identificado secuencias de ADN de *Chlamydia* en lesiones ateroscleróticas, lo cual ha llevado a sugerir que este organismo desempeña una función en la enfermedad.

Además, con la llegada de cepas de bacterias resistentes a bacterias, las infecciones bacterianas se han convertido en un problema para la salud. Además, las infecciones microbianas causadas por bacterias u hongos que normalmente no infectan a seres humanos, pueden ser un problema en individuos inmunocomprometidos. Adicionalmente, los individuos de países en vías de desarrollo que pueden estar mal nutridos o que carecen de instalaciones sanitarias adecuadas, también pueden soportar una gran carga de bacterias oportunistas, muchas de las cuales pueden causar afecciones y enfermedades. En la medicina veterinaria, el ganado que vive en dependencias cercanas también puede ser presa de infecciones causadas por diversos tipos de microbios diferentes. Por tanto, hay una necesidad de desarrollar métodos de identificación sensibles de muchos tipos de microbios diferentes sin tener que cultivarlos primero para tratar o prevenir infecciones microbianas en seres humanos y otros animales.

También son un componente importante los ensayos de contaminación microbiana para analizar productos alimentarios. Una gran cantidad de tipos de microbios diferentes puede contaminar los alimentos de seres humanos o animales. Por tanto, la capacidad de analizar la contaminación de los alimentos de una manera rápida y eficaz es crítica para conservar seguros los alimentos. Sin embargo, muchos de los microbios responsables, causantes de afecciones en seres humanos y animales, son difíciles de aislar o identificar.

También se utilizan ensayos de poblaciones microbianas en campos tales como ciencia forense. En los últimos diez a veinte años, los científicos han determinado que las poblaciones microbianas cambian cuando el organismo empieza a descomponerse, y han comenzado a identificar determinadas especies microbianas que son indicativas de descomposición (Lawrence Osborne, Crime-Scene Forensics; Dead Men Talking, New York Times, 3 de diciembre de 2000). Sin embargo, solo se han identificado algunas especies microbianas que pueden ser útiles en estos análisis.

El problema de determinar la diversidad genética no está limitado a las poblaciones microbianas. La diversidad de los anticuerpos es crítica para una respuesta inmunitaria apropiada. Durante la proliferación de células B, se genera diversidad de anticuerpos en las cadenas pesada y ligera de la inmunoglobulina por mecanismos que incluyen genes variables (V) de línea germinal múltiple, recombinación de segmentos génicos V con segmentos génicos de unión (J) (recombinación V-J) y recombinación de segmentos génicos V con segmentos génicos D y segmentos génicos J (recombinación V-D-J) así como imprecisiones recombinatorias. Además, mutaciones puntuales somáticas que se producen durante toda la vida del individuo, también conducen a la diversidad de anticuerpos. Por tanto, puede generarse un enorme número de genes de anticuerpos diferentes que codifican anticuerpos con especificidad exquisita. La diversidad de receptores de células T (RCT) se genera de una manera similar a través de recombinación entre numerosos segmentos V, D y J e imprecisiones recombinatorias. Se ha estimado que pueden constituirse 10^{14} cadenas V δ , más de 10^{13} cadenas β y más de 10^{12} formas de cadenas V α (Roitt, I. *et al.*, Immunology, 3ª Ed., 1993, páginas 5.1 - 5.14). Un conocimiento de la diversidad de los anticuerpos o de los TCR en un individuo particular sería útil para el diagnóstico de enfermedades, tal como una enfermedad autoinmunitaria, o para un posible tratamiento.

La identificación de microbios, especialmente microbios del suelo, se ha basado tradicionalmente en métodos dependientes de cultivo, por lo cual la detección de una especie microbiana depende de la capacidad de encontrar condiciones de laboratorio que sustenten su crecimiento. Para esto, en el comercio se han desarrollado placas de 96 pocillos, para identificar microbios con diferentes requisitos metabólicos. Por ejemplo, las placas BioLog incorporan 96 formulaciones de medios diferentes en los pocillos de una placa de 96 pocillos. A pesar de estos esfuerzos, actualmente se acepta que mucho menos del 1 % de los microbios puede propagarse en condiciones de laboratorio (Amann, R. I. *et al.*, 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. 59: 143-169).

El amplio interés en genómica ha creado mucha tecnología nueva, excitante, para la cuantificación en paralelo de miles de secuencias distintas de ácido nucleico de manera simultánea. Aunque aún en pañales, estas tecnologías han proporcionado un profundo conocimiento sin precedentes en la biología. Hasta ahora, estas tecnologías se han utilizado predominantemente en aplicaciones farmacéuticas y agrícolas. El perfil de expresión del genoma ha obtenido una aceptación general en la biología y es posible que comience a ser habitual en todas las instituciones académicas, de biotecnología y farmacéuticas, en el siglo 21. Por ejemplo, el Análisis en Serie de la Expresión Genica (SAGE, *Serial Analysis of Gene Expression*) es un método independiente de la hibridación diseñado para cuantificar cambios en la expresión génica (Velculescu, V. E. *et al.*, 1995, *Serial analysis of gene expression*. Science 270: 484-487 y patente de Estados Unidos 5.866.330). Sin embargo, el SAGE solo mide niveles de ARN de tejidos u organismos y no es adecuado para examinar la diversidad genética.

El amplio interés en genómica también ha conducido al desarrollo de muchas tecnologías para el análisis rápido de decenas de miles de secuencias de ácido nucleico. Una tecnología de este tipo es la microplaca (*chip*) de ADN. Aunque esta estrategia se ha utilizado como un diagnóstico para diferenciar entre varias especies del género *Mycobacterium* (Troesch, A., *et al.*, 1999, *Mycobacterium species identification and rifampin resistance testing with high-density DNA probe arrays*. J. Clin. Microbiol. 37: 49-55), tiene utilidad limitada para un estudio microbiano ambiental por dos razones. La primera de ellas es que la secuencia de los ADN diana a analizar debe conocerse para sintetizar las sondas complementarias en la microplaca. Sin embargo, la inmensa mayoría de los microbios ambientales no se han caracterizado. En segundo lugar, las microplacas de ADN se basan en la hibridación de ácidos nucleicos que está sometida a hibridación cruzada a partir de moléculas de ADN con secuencia similar. Sin embargo, el poder resolutivo de una estrategia basada en hibridación es limitado porque se deben identificar regiones de ADN que no se hibriden en cruzado, lo que puede ser difícil para especies microbianas relacionadas.

Las tecnologías genómicas y bioinformáticas soportan mucho potencial desaprovechado para su aplicación en otras áreas de biología, especialmente en el campo de la microbiología. Sin embargo, hasta ahora no ha habido ningún método para determinar de un modo rápido y fácil la diversidad genómica de una población, tal como una población microbiana o de virus. Además, no ha habido ningún método para determinar fácilmente la diversidad de anticuerpos o RCT de una población de células B o T, respectivamente. Por tanto, aún persiste la necesidad de desarrollar dichos métodos en estas áreas.

En "La caracterización de genes de ARNr 16S de comunidades microbianas de campos petrolíferos indica la presencia de diversas bacterias reductoras de sulfato, fermentativas y oxidadoras de sulfuro" de Gerrit Voordouw *et al.* se desvela la caracterización de bacterias de campos petrolíferos, clonando y secuenciando genes de ARNr 16S amplificados por PCR para detectar una bacteria determinada.

En *Proyecto de bases de datos de ribosomas de Niels Larsen et al.* se desvela una base de datos curada que ofrece datos de ribosomas junto con programas y servicios relacionados. La oferta incluye alineaciones filogenéticamente ordenadas de secuencias de ARN ribosómico (ARNr), árboles filogenéticos derivados, diagramas de estructura secundaria de ARNr y diversos paquetes informáticos para manipular, analizar y presentar alineaciones y árboles.

En "*Determinación de la diversidad microbiana en muestras ambientales: fallos de análisis de ARNr basados en PCR*" de Friedrich v. Wintzingerode et al. se desvelan aspectos específicos de conjuntos de muestras, lisis celular, extracción de ácido nucleico, amplificación por PCR, separación de ADN amplificado, aplicación de sondas nucleicas y análisis de datos.

En "*Filogenia de las principales secuencias bacterianas de ARNr 16S en suelos de pastizales de Drentse A (Países Bajos)*" de Andreas Felske et al. se desvela la investigación de principales bacterias en suelos de turba, de pastizales ácidos en los Países Bajos, por aislamiento de ribosomas, electroforesis en gel con gradiente de temperatura, hibridación, clonación y secuenciación utilizando ARNr para clasificar y cuantificar las bacterias más activas.

En "*Frecuencia de formación de moléculas quiméricas como consecuencia de coamplificación PCT de genes de ARNr 16S de genomas bacterianos mixtos*" de Grace C. Y. Wang y Yue Wang se desvela un método para cuantificar la frecuencia de la formación de moléculas quiméricas, en el que los ADN genómicos mixtos de ocho especies de actinomiceto cuyas secuencias de ARNr 16S se han determinado se utilizaron para la coamplificación por PCR de genes de ARNr 16S. Una gran cantidad de ADN ribosómico 16S clonado se examinó por análisis de secuencias y se identificaron moléculas quiméricas por alineamiento de secuencias múltiples con especies de referencia.

En "*Análisis en Serie de la Expresión Génica*" de Victor E. Velculescu et al. se desvela un método denominado análisis en serie de la expresión génica (SAGE), que permite analizar, de un modo cuantitativo y simultáneo, una gran cantidad de transcritos. El SAGE podría proporcionar medios ampliamente aplicables para catalogar y comparar cuantitativamente genes expresados en una variedad de estados normales, de desarrollo y con enfermedad.

Breve resumen de la invención

La presente invención resuelve este problema proporcionando métodos para determinar rápidamente la diversidad de una población microbiana.

La presente invención también proporciona un método para construir una base de datos de perfiles de diversidad de marcadores (PDM *marker diversity profile*) que comprende las etapas de: a) secuenciar una pluralidad de marcadores de ARNr de una muestra, en el que dicha muestra comprende una población microbiana, y en el que cada marcador de ARNr comprende una secuencia polimórfica del gen del ARNr microbiano y en el que la pluralidad de marcadores de ARNr proporciona una representación exacta de la población microbiana de dicha muestra; b) determinar la abundancia, en dicha muestra, de cada una de dicha pluralidad de marcadores de ARNr; c) transducir la abundancia de cada marcador de ARNr en una señal de salida eléctrica; d) almacenar dichas señales de salida eléctrica producidas en la etapa c) en una estructura de datos matricial y asociar en dicha estructura cada señal de salida con la secuencia correspondiente del marcador de ARNr cuya abundancia se convirtió en una señal de salida; e) proporcionar las abundancias de uno o más parámetros de muestra que están asociados a dicha muestra; f) transducir las abundancias de cada parámetro de muestra en una señal de salida eléctrica; g) almacenar dichas señales de salida eléctrica producidas en la etapa f) en una estructura de datos matricial y asociar en dicha estructura cada señal de salida con el parámetro de muestra cuya abundancia se convirtió en una señal de salida; h) diseñar como PDM las señales de salida eléctrica producidas en la etapa c) correspondientes a la pluralidad de abundancias de marcadores de ARNr y las señales de salida eléctrica producidas en la etapa f) correspondientes a la una o más abundancias de parámetros de muestra; y i) repetir las etapas a-h para una pluralidad de muestras distintas, y diseñar como base de datos PDM la pluralidad de PDM almacenados en dicha estructura de datos.

En una realización, dicha abundancia de marcador de ARNr es una abundancia relativa a la abundancia total de la pluralidad de marcadores de ARNr en la muestra.

En una realización, dicho gen de ARNr microbiano es un gen de ARNr 16S.

En una realización, dicha secuencia polimórfica del gen de ARNr microbiano es de la región intergénica entre un gen de ARNr 16S y un gen de ARNr 23S.

En una realización, la muestra se selecciona del grupo constituido por : una muestra de suelo, una muestra de roca, una muestra de agua, una muestra de aire, una muestra de hidrocarburo, una muestra de petróleo y una muestra de biopelícula.

En una realización, la muestra se obtiene del grupo constituido por: petróleo, hidrocarburos, un depósito de aceite, un depósito de gas, suelo, roca, agua, aire, una construcción y una carretera. En una realización, el parámetro de muestra se selecciona del grupo constituido por: petróleo, hidrocarburos, aceite, gas, un compuesto inorgánico, un

mineral, pH, temperatura y salinidad.

En una realización, la muestra se obtiene del grupo constituido por: un ser humano, una planta, un animal, un producto alimentario, un cultivo celular y un cultivo tisular.

5 En una realización, el parámetro de muestra se selecciona del grupo constituido por: una patología aguda, una patología crónica, un estado fisiológico y una fase de desarrollo.

10 En una realización, el parámetro de muestra es el tiempo, y en el que la pluralidad de muestras comprende una evolución temporal.

En una realización, la pluralidad de muestras comprende una evolución temporal útil para supervisar esfuerzos de remediación ambiental o la producción de campos petrolíferos.

15 En una realización, se construye una base de datos PDM de acuerdo con el método de una cualquiera de las reivindicaciones 1 a 10.

Breve descripción de los dibujos

20 La Figura 1 muestra una representación de un gen de ADNr 16S de bacterias que ilustra regiones polimórficas (mostradas como bandas oscuras) y regiones constantes (mostradas como bandas claras).

La Figura 2 muestra una representación esquemática del método de análisis en serie de ADN ribosómico (SARD, *Serial Analysis Ribosomal ADN*) para aislar etiquetas de secuencia polimórficas de ADNr bacteriano.

La Figura 3 muestra diversos miembros representativos del dominio *Bacteria* con sus relaciones taxonómicas.

25 La Figura 4 muestra diversos miembros representativos del dominio *Arquea* con sus relaciones taxonómicas.

La Figura 5 muestra ejemplos de dispersogramas de correlación entre marcadores. Cada punto representa una sola población de muestra.

La Figura 6 muestra ejemplos de diversos dispersogramas entre parámetros y marcadores. Cada punto representa una sola población de muestra.

30 Las Figuras 7A, 7B y 7C muestran dispersogramas que comparan los perfiles del perfil de diversidad de marcadores (PDM). Cada punto representa un marcador. Fig. 7A: correlación significativa con todos los marcadores dentro del PDM. Fig. 7B: no se encuentra correlación utilizando todos los marcadores dentro del PDM. Fig. 7C: misma gráfica que B excepto que se encuentra correlación significativa utilizando un subconjunto de marcadores.

35 La Figura 8 muestra un esquema para la generación de una base de datos matricial de perfiles de diversidad de marcadores. La primera etapa **100** implica asignar a N un valor entero de 1 que corresponde a la primera estructura de datos del PDM. La segunda etapa **105** implica agrupar marcadores polimórficos de una muestra.

En la presente solicitud se describen ejemplos de métodos para agrupar dichos marcadores y se incluye, pero sin limitación, SARD (Figura 2, Tablas I y II) o recopilación masiva de etiquetas (Tabla III). La siguiente etapa **110**

40 implica detectar los marcadores polimórficos. Los ejemplos de métodos para detectar marcadores polimórficos descritos anteriormente incluyen análisis de secuencia de ADN de etiquetas SARD y análisis MALDI-TOF de etiquetas masivo, respectivamente. Esta etapa de detección incluye detectar la presencia y la abundancia de cada marcador en la muestra. La siguiente etapa **115** implica la conversión o la transducción de PDM en una salida de señal eléctrica. Generalmente, este proceso es una conversión electrónica lineal de los datos en una

45 señal digital. La etapa **120** implica detectar parámetros que están asociados a la muestra N. La etapa **125** implica la transducción de los datos de los parámetros de la muestra, que pueden incluir, sin limitación, parámetros tales como pH, tamaño del grano, análisis elemental y/o análisis orgánico, en una salida de señal eléctrica en forma de una salida digital. La etapa **130** implica el almacenamiento, en la memoria de un ordenador, de la señal de salida de cada PDM en una estructura de datos matricial y asociarla con parámetros de muestra. La siguiente etapa es

50 un bloque de toma de decisiones **135** en el que si no se han completado todas las estructuras de los datos, la rutina avanza a la etapa **140** en la que N está incrementado y las etapas **105-130** generadoras de estructuras de datos se repiten para el perfil de diversidad del marcador n +1.

Una vez completadas todas las estructuras de los datos, la rutina avanza a la etapa **145** para formar la base de datos PDM de estructuras de n datos. Cada marcador se asigna un único identificador junto con su abundancia

55 relativa en la población. Esta información también está opcionalmente indexada con otros parámetros conocidos que están asociados a la muestra que incluyen, por ejemplo, el tiempo, los datos, la altura y la localización geográfica. Estas señales se digitalizan y se almacenan en la memoria de un ordenador.

La Figura 9 muestra un esquema de las etapas implicadas en la determinación de parámetros asociados con un PDM. Se crea un perfil de diversidad de marcadores **200** para una muestra. El perfil de diversidad de marcadores

60 se somete a una función de comparación **205** que compara el perfil con perfiles de diversidad de marcadores residentes en la base de datos. La etapa **210** es un bloque de toma de decisiones en el que se realiza una pregunta de si el nuevo perfil de diversidad de marcadores es igual a un perfil de diversidad de marcadores residente en la base de datos. Si la pregunta retorna a "Si" el nuevo perfil de diversidad de marcadores se deduce **215** que comparte los mismos parámetros que los del perfil de diversidad de marcadores residente. Dado

que los parámetros asociados al perfil de diversidad de marcadores residente están caracterizados, los parámetros asociados al nuevo perfil de diversidad marcadores se identifican.

65

Si la pregunta en el bloque de toma de decisiones 210 retorna a "No" la rutina avanza al bloque de toma de decisiones 220 que pregunta si el nuevo perfil de diversidad de marcadores es un subconjunto de un perfil residente en la base de datos. Si la pregunta retorna a "No" los parámetros permanecen indefinidos **225**. Si la pregunta retorna a "Si" la rutina avanza a la etapa **230**. La etapa **230** es una etapa opcional para determinar la correlación entre miembros del subconjunto común de marcadores y puede o bien realizarse para cada nuevo perfil o puede preguntarse a una tabla matricial de valores precalculados de perfiles existentes. Dichos valores generalmente se mantendrían en una base de datos relacional. Si esta etapa no se realiza todos los marcadores comunes se analizan en grupos de marcadores individuales y se tratan como grupos correlacionados **255**. Y la correlación de marcadores se realiza entre los subconjuntos comunes de marcadores, la rutina avanza al bloque de toma de decisiones 235 que pregunta si todos los marcadores comunes están correlacionados. Si la pregunta retorna a "Si" los marcadores están correlacionados con los parámetros **240** residentes en la base de datos. Si ningún marcador está correlacionado con un parámetro, el parámetro o los parámetros permanecen indefinidos **245** mientras que si los marcadores están correlacionados con un parámetro, el parámetro se deduce que está asociado con el perfil de diversidad de marcadores **250**. Si la pregunta del bloque de toma de decisiones **235** retorna a "No", los marcadores comunes se clasifican en grupos de marcadores correlacionados **255**. El primer grupo marcador correlacionado N **260** se somete a un bloque de toma de decisiones 265 que pregunta si ente grupo los marcadores están correlacionados con un parámetro. Un "No" determina que los parámetros permanecen indefinidos. Si el marcador o los marcadores están correlacionados con un parámetro, se deduce que el parámetro está asociado con el perfil de diversidad de marcadores. Las etapas **260-275** se repiten en las etapas 280-295 para para cada grupo de marcadores correlacionado. Los grupos de marcadores correlacionados pueden comprender un solo marcador o varios marcadores. El nivel de confianza en la deducción de que un parámetro está asociado con un perfil de diversidad de marcadores se determina a través del nivel de correlación entre el marcador o marcadores y el parámetro. Por lo tanto, se espera que los conjuntos de marcadores correlacionados sean indicadores más fuertes que cualquier parámetro determinado.

La Figura 10 muestra un esquema para la generación de una base de datos matricial de perfil de diversidad de marcadores. La primera etapa 300 implica asignar a N un valor entero de 1 que corresponde a la primera estructura de datos PDM. La segunda etapa 305 implica agrupar marcadores polimórficos de una muestra. En la presente solicitud se describen ejemplos de métodos para agrupar dichos marcadores e incluyen, pero sin limitación, SARD (Figura 2, Tablas I y II) o recopilación masiva de etiquetas (Tabla III). La siguiente etapa **310** implica detectar los marcadores polimórficos. Los ejemplos de métodos para detectar marcadores polimórficos descritos anteriormente incluyen análisis de secuencia de ADN de etiquetas SARD y análisis MALDI-TOF de etiquetas masivo respectivamente. La siguiente etapa implica la conversión o transducción del PDM en una salida de señal eléctrica. Generalmente, este proceso es una conversión electrónica lineal de los datos en una señal digital. La etapa **320** implica almacenar en la memoria de un ordenador la señal de salida de cada PDM en una estructura de datos matricial asociando cada PDM a una coordenada geográfica tal como longitud y latitud. La siguiente etapa es un bloque de toma de decisiones **325** en el que si no se han completado todas las estructuras de datos, la rutina avanza a la etapa **330** en la que N se incrementa y las etapas 305-320 generadoras de estructuras de datos se repiten para el perfil de diversidad de marcadores n +1. Una vez que todas las estructuras de datos están completadas, la rutina avanza a la etapa **335** para formar la base de datos PDM de N estructuras de datos. Cada marcador se asigna a un único identificador y se indexa con su abundancia relativa en la población. Estas señales se digitalizan y se almacenan en la memoria de un ordenador.

La Figura 11 muestra un esquema de mapas de aplicaciones utilizando perfiles de diversidad de marcadores. Los datos de los perfiles de diversidad de marcadores **400** pueden procesarse de diversas maneras para crear mapas que proporcionen información ambiental significativa. En un ejemplo **405**, cada perfil de diversidad de marcadores en una base de datos, puede correlacionarse por pares con cada uno de los otros perfiles de diversidad de marcadores, para crear una matriz de correlación. Anexando estos datos con las coordenadas geográficas de cada muestra **410**, puede construirse un mapa que refleje los valores de correlación de sitios de muestras físicamente adyacentes. Preferentemente los valores de correlación se codificarán con colores que reflejen el nivel de correlación. El color se selecciona de un espectro de color de referencia que está indexado con valores de correlación entre 0-1.

Los perfiles de diversidad de marcadores **400** también pueden procesarse en mapas al nivel de marcador individual o de grupo de marcadores correlacionado. Esta estrategia es preferible ya que los subconjuntos de marcadores probablemente se correlacionen con un menor número de parámetros asociados a la muestra. Cada marcador en una base de datos de perfil de diversidad de marcadores se correlaciona por pares con cada uno de los otros marcadores en la base de datos para crear una matriz de correlación **415**. La base de datos original puede estar compuesta por perfiles de diversidad de marcadores de una sola área geográfica o de diversas áreas geográfica. En la etapa **420**, los marcadores de un área geográfica se clasifican en grupos basándose en su nivel de correlación. En la etapa **425**, la representación relativa del grupo marcador correlacionado N se determina junto con sus coordenadas geográficas para cada perfil de diversidad de marcadores en un área geográfica. Se construye un mapa **430** en el que la abundancia relativa de cada grupo marcador correlacionado se codifica con un color con sus coordenadas geográficas. Las etapas **425** y **430** se repiten como en **435** y **440** para cada grupo de marcadores correlacionado.

La Figura 12 muestra oligonucleótidos útiles para la amplificación de moléculas de ácido nucleico para análisis en serie de ADN ribosómico, SARD.

La Figura 13 muestra el uso de la estrategia SARD para *Eubacteria*. La secuencia con doble subrayado y la secuencia con subrayado ondulado representan las etiquetas de secuencia para los dos grupos y la secuencia

con subrayado sencillo define el sitio de reconocimiento *Bpml*.

La Figura 14 es una representación gráfica de un análisis SARD de una población definida.

La Figura 15 muestra la secuencia de etiquetas SARD identificadas de la muestra Wy-1. El número entre paréntesis indica el número de etiquetas que tienen esa secuencia.

La Figura 16 muestra etiquetas SARD identificadas de la muestra Wy-2. El número entre paréntesis indica el número de etiquetas que tienen esa secuencia.

La Figura 17 es una representación gráfica del número y abundancia de etiquetas SARD. El panel superior muestra el perfil de diversidad de etiquetas SARD de la muestra Wy-1 y el panel inferior muestra el perfil de diversidad de etiquetas SARD de la muestra Wy-2.

Descripción detallada de la invención

El grado de diversidad de los microbios en nuestro ambiente se ha reconocido solo recientemente. Con la llegada de la reacción en cadena de la polimerasa (PCR) y del análisis de secuencia de ADN ribosómico (ADNr) de subunidad pequeña, los investigadores han podido detectar y realizar análisis filogenéticos en microbios individuales sin tener que cultivar primero los microbios de interés. Esta estrategia filogenética molecular ha cambiado significativamente nuestra perspectiva de la evolución y de la diversidad microbiana (Woese, C. R, 1987, *Bacterial evolution*. Microbiol Rev. 51 (2): 221-71; Pace, N. R, 1997, A molecular view of microbial diversity and the biosphere. Science. 276 (5313): 734-40). Por ejemplo, ahora se piensa que las primeras formas de vida habían utilizado compuestos inorgánicos para la nutrición en lugar de compuestos basados en carbono orgánico. Además, ahora se sabe que la inmensa proporción de diversidad biológica se debe a especies microbianas. Se han realizado estimaciones de que puede haber más de diez mil especies distintas de microbios en un solo gramo de suelo. Las Figuras 3 y 4 muestran algunos de los miembros representativos de los dominios *Bacteria* y *Archaea*, respectivamente, que pueden encontrarse en muestras ambientales.

Los microbios residen prácticamente en todos los nichos, incluyendo ambientes extremos con temperaturas entre -6,67 y 121,11 °C (20 y 250 °F). Los microbios se han aislado incluso de depósitos de petróleo con una profundidad de más de 1,61 km (1 milla) por debajo de la superficie terrestre (Jeanthon, C. *et al.*, 1995, *Thermotoga subterranea* sp. nov., a new thermophilic bacterium isolated from a continental oil reservoir. Arch. Microbiol. 164: 91-97). Para prevalecer en dichas condiciones diversas, los microbios han realizado adaptaciones notables y han adquirido la capacidad de utilizar fuentes inusuales de carbono y minerales inusuales que están inmediatamente disponibles. Estas adaptaciones fisiológicas y metabólicas que permiten a algunos microbios residir en un nicho particular también pueden limitar su distribución a dichas áreas. Se conocen bien numerosos ejemplos de parámetros ambientales que conducen a restricciones de distribución microbiana y normalmente están estipuladas por un programa metabólico específico de especie (por ejemplo, naturaleza estricta de la fuente de carbono, nitrógeno y energía).

Los microbios que tienen requisitos de nutrientes muy definidos probablemente tienen una distribución restringida en el entorno. Por tanto, la dependencia de los microbios sobre la presencia de una fuente particular para proliferar puede servir como la base para un ensayo para identificar la presencia y caracterizar la distribución de diversas características en el medio, características tales como biológicas, químicas y geoquímicas. En otras palabras, los microbios pueden funcionar como biosensores ambientales.

En un aspecto de la presente invención, la capacidad de los microbios para funcionar como biosensores ambientales se utiliza para identificar estados ambientales particulares. En una realización preferida, se utiliza un perfil de una población microbiana para identificar uno o más parámetros de un estado ambiental particular. En una realización más preferida, se utiliza un perfil de población microbiana para identificar áreas que probablemente tienen reservas minerales y/o petrolíferas. En otra realización preferida, en ciencia forense se utiliza un perfil de población microbiana para identificar la descomposición de un cuerpo o para asociar un individuo con otro individuo, o con un objeto o con una localización. En otra realización aún preferida, para identificar contaminación microbiana de productos alimentarios de seres humanos y animales se utiliza un perfil de población microbiana. En otra realización aún preferida, el perfil se utiliza para diagnosticar enfermedades humanas o animales.

Definiciones

A menos que se defina de otra manera, todos los términos técnicos y científicos utilizados en el presente documento tienen el significado que normalmente entiende un experto habitual en la técnica a la que pertenece la presente invención. La realización práctica de la presente invención emplea, a menos que se indique de otra manera, técnicas de química, biología molecular, microbiología, ADN recombinante, genética e inmunología convencionales. Véase, por ejemplo, Maniatis *et al.*, 1982; Sambrook *et al.*, 1989; Ausubel *et al.*, 1992; Glover, 1985; Anand, 1992; Guthrie y Fink, 1991.

Un microbio se define como cualquier organismo que sea de los dominios *Bacteria*, *Eucaria* o *Arquea*. Los microbios incluyen, sin limitación bacterias, hongos, nemátodos, protozoos, arqueobacterias, algas, dinoflagelados, mohos, bacteriófagos, micoplasma, virus y viroides.

Un marcador es una secuencia de ADN que puede utilizarse para diferenciar o identificar un gen, un genoma o un

organismo particular de otro. En una realización, puede generarse un marcador mediante uno de los métodos descritos en el presente documento. Un marcador representa uno de un número limitado de especies taxonómicas o genes. En una realización preferida, un marcador representa una sola especie taxonómica o un gen. En una realización, el marcador representa una sola especie microbiana. En otra realización, el marcador representa una sola especie o tipo de virus. En otra realización, el marcador representa una sola inmunoglobulina o dominio variable de TCR.

Un perfil de diversidad de marcadores (PDM) es un conjunto de datos que se obtiene de cada muestra poblacional y que contiene una colección de marcadores. En una realización preferida, el PDM también comprende otra información, incluyendo todos los parámetros conocidos asociados a una muestra poblacional particular. Dichos parámetros que se relacionan con muestras ambientales pueden incluir componentes inorgánicos (obtenidos a través de análisis de adsorción atómica), componentes orgánicos (obtenidos a través GC-MS o LC-MS), análisis de tamaño de grano, pH y salinidad. Los parámetros que se relacionan con muestras médicas pueden incluir, pero sin limitación, un historial médico completo del donante. En una realización preferida, los marcadores se obtienen por análisis en serie de ADN ribosómico, SARD.

Métodos para el análisis genético de poblaciones

Los ribosomas, que contienen numerosas proteínas ribosómicas y tres moléculas de ARN ribosómico (ARNr), son un componente clave de la síntesis de proteínas. El ARNr de subunidad 16S, que está codificado por el gen de ADNr 16S, ha sido el centro de gran atención en estudios filogenéticos microbianos. La secuencia de ADNr 16S está muy conservada entre los grupos taxonómicos, incluso también posee regiones que son muy polimórficas (Figura 1). Además, se piensa que la velocidad de cambio en la secuencia de ARN ha sido relativamente constante a lo largo del tiempo evolutivo, lo que permite a los científicos determinar la relación relativa de los diferentes organismos.

Los análisis microbianos moleculares típicos implican la utilización de regiones muy conservadas del ADNr 16S para amplificar el gen de aproximadamente 1.500 pb. La secuencia del producto amplificado por PCR se determina y se compara con la de otras secuencias de ADNr conocidas. Aunque esta estrategia es muy informativa, no permite estudiar rápidamente una comunidad microbiana ambiental.

La presente invención proporciona métodos para calcular rápida y fácilmente la diversidad genética de una población. Los métodos utilizan tecnología genómica independiente de hibridación para superar los problemas de determinación de la diversidad genética identificados anteriormente. Este método puede utilizarse para cualquier población de células, virus u organismos que comprendan al menos una molécula de ADN que comprenda regiones de alta conservación de secuencia intercaladas con secuencias polimórficas, en el que las secuencias polimórficas pueden utilizarse para diferenciar diferentes miembros de la población de interés. Un aspecto de la presente invención describe un método (SARD) que puede capturar una región polimórfica diseñada de una molécula de ADN determinada presente en los miembros de una comunidad microbiana. En una realización preferida, la molécula de ADN es una molécula de ADNr 16S. En otra realización, la molécula de ADN es la región intergénica entre los genes de ADNr 16S y 23S.

El método puede realizarse de la siguiente manera (véase la Figura 2):

Etapas 1. Preparación de muestras y amplificación de ADN por PCR

Las muestras pueden obtenerse de cualquier organismo o región que se desee. Para los análisis microbianos ambientales, pueden obtenerse muestras de, sin limitación, construcciones, carreteras, suelo, rocas, plantas, animales, cultivos celulares o tisulares, desechos orgánicos, aire o agua. Para análisis microbianos médicos, las muestras pueden obtenerse de, sin limitación, seres humanos, animales, parásitos, agua, suelo, aire y productos alimentarios. Para análisis de virus, las muestras pueden obtenerse de, sin limitación, reservas de cultivos de virus, seres humanos, animales, plantas, cultivos celulares o tisulares y microbios. Para análisis de inmunoglobulina o TCR, las muestras pueden obtenerse de, sin limitación, seres humanos, animales o cultivos celulares o tisulares. Las moléculas de ADN de la muestra de interés pueden aislarse mediante cualquier método conocido en la técnica. Véase, por ejemplo, Sambrook *et al.*, 1989 y Ausubel *et al.*, 1992. En una realización preferida, el ADN se obtiene como describen Yeates *et al.*, "Methods for Microbiological DNA Extraction from Soil for PCR Amplification," Biological Procedures Online, volumen 1, 14 de mayo de 1998, disponible en www.science.uwaterloo.ca/bpo; Liu *et al.*, Applied and Environmental Microbiology (1997) 63: 4516-4522; y Tsai *et al.*, Applied and Environmental Microbiology (1992) 58: 2292-2295. Las moléculas de ADN no tienen que estar completamente purificadas sino que solamente requieren aislarse en el punto en el que pueda realizarse la PCR.

Los microbios ambientales a menudo existen en biopelículas (Costerton, JW., *et al.*, 1999, Bacterial biofilms: a common cause of persistent infections. Science 284(5418): 1318-1322) o en estrecha asociación con superficies sólidas. El ADN microbiano de una muestra de interés se aísla mediante cualquiera de los diversos métodos que conocen ampliamente los expertos en la técnica y que se describen en la bibliografía (Gillan, D.C. *et al.*, 1998, Genetic diversity of the biofilm covering Montacuta ferruginosa (Mollusca, bivalvia) as evaluated by denaturing gradient gel electrophoresis analysis and cloning of PCR-amplified gene fragments coding for 16S rRNA Appl.

Environ. Microbial. 64(9): 3464-72).

Las muestras pueden enriquecerse selectivamente antes de aislarse mediante cualquier método conocido en la técnica. En una realización, para una población de microbios que se sabe o que se sospecha que se alimenta de una fuente de hidrocarburo, tal como propano, el hidrocarburo puede añadirse al medio en el cual viven los microbios durante un periodo de tiempo antes que estos se recojan. En otra realización, puede cultivarse una población de virus en células antes de aislarse. En una realización adicional, las células B y T pueden expandirse en el cultivo antes del aislamiento. Puede ser más fácil obtener suficientes cantidades de muestra si una población se expande antes del aislamiento. Sin embargo, esto debe ponderarse frente a la posibilidad de que la expansión altere la proporción de los diferentes miembros de la población entre sí.

En general, los cebadores utilizados para la amplificación se diseñan para hibridarse con una región del ADN que está muy conservada entre los miembros de la población. Además, los cebadores deben flanquear una región polimórfica cuya secuencia parcial debe proporcionar información de diagnóstico con respecto a la diversidad genética de la población. Por ejemplo, para el gen de ADNr 16S, se diseñan cebadores que se hibridan con una región muy conservada del gen de ADNr 16S que flanquea una región polimórfica (véase la Figura 2). En un gen de inmunoglobulina, se diseñan cebadores que se hibridan con una región del ADN de la célula B que flanquea el sitio de recombinación V-J. Como alternativa, pueden diseñarse cebadores que se unan a las regiones relativamente constantes dentro de determinadas regiones del gen de V-J que flanquea una región polimórfica. Véase Roitt *et al.*, Immunology, 3^a Ed., 1993, págs. 5.2-5.14, que muestra regiones de variabilidad y conservación dentro de los genes de inmunoglobulina y TCR. Después de las enseñanzas de la memoria descriptiva, un experto habitual en la materia reconocerá que en el diseño de cebadores, pueden utilizarse otros genes que tengan regiones que estén muy conservadas entre miembros de la población y que flanqueen regiones polimórficas.

Los cebadores también pueden diseñarse para que flanqueen una región de ADN que comprenda un sitio de restricción para una enzima de restricción. En una realización preferida, la enzima de restricción es una enzima de cuatro pares de bases, tal como *AluI*, que corta en un sitio de reconocimiento (véase la Figura 2). Adicionalmente, el sitio de restricción debe estar cerca del gen de interés pero no en la región polimórfica del mismo. En una realización preferida, el sitio de restricción debe ser uno que esté presente en el gen de interés en gran parte de las especies o genes conocidos.

Puede utilizarse un solo conjunto de cebadores o múltiples conjuntos de cebadores. Se puede utilizar un solo conjunto de cebadores si se sabe que la región de ADN a la cual se unirán los cebadores está muy conservada. Como alternativa, si se sabe que hay alguna variación en la región conservada, pueden utilizarse múltiples conjuntos de cebadores para unirse a la región de ADN conservada. El uso de múltiples conjuntos de cebadores puede ser útil para identificar más miembros de una población, especialmente aquellos miembros de la población que exhiben menos identidad de secuencia en las áreas conservadas de una secuencia de ácido nucleico. En una realización puede utilizarse de cuatro a diez conjuntos de cebadores para identificar miembros de una población. Como alternativa, los cebadores utilizados pueden ser degenerados, de tal manera que diferentes moléculas dentro de una población de cebadores incluirán una base diferente en uno o más sitios específicos en el cebador. Por ejemplo, un cebador puede tener un sitio que tenga citosina o timidina. El propósito de construir cebadores degenerados es aumentar el número de diferentes moléculas de ADN que se hibriden con un cebador particular. En la técnica se conocen bien métodos de construcción de cebadores degenerados.

Los cebadores utilizados en esta etapa, y en etapas posteriores, generalmente tienen una longitud suficiente para promover una hibridación específica con una secuencia de ADN. Los cebadores generalmente tienen una longitud de al menos 12 bases, más preferentemente de al menos 15 bases, incluso más preferentemente de al menos 18 bases. Los cebadores pueden tener una longitud de hasta 60 bases, aunque normalmente la mayoría tienen 40 bases de longitud. Los cebadores pueden incluir tanto bases que se encuentran de manera natural en el ADN, tales como adenina, guanina, citosina y timidina, como también pueden incluir nucleótidos que no se encuentran normalmente en el ADN, tal como inosina.

Uno de los cebadores (el cebador "aguas arriba") debe modificarse para incorporar una fracción que pueda utilizarse para unir el producto de la PCR a un soporte sólido. El cebador aguas arriba se define como el cebador que se localiza en el lado opuesto de la región polimórfica de interés relativa al sitio de restricción de cuatro bases flanqueante. En la técnica se conocen diversas fracciones de unión diferentes. En una realización preferida, la fracción es biotina. En otra realización preferida, la fracción es digoxigenina o seis histidinas.

La PCR se realiza utilizando los cebadores para amplificar una subregión que contenga un sitio polimórfico de interés. En la técnica se conocen bien métodos para realizar la PCR. En una realización, los productos de la PCR se normalizan o se sustraen por métodos conocidos en la técnica para reducir la representación de las secuencias dominantes. En Sambrook *et al.*, 1989, Ausubel *et al.*, 1992 Glover, 1985; Anand, 1992 se describen métodos ejemplares.

Etapa II. Digestión del fragmento amplificado y unión a un soporte sólido.

El fragmento amplificado se corta con la enzima de restricción como se ha indicado en la etapa I. En esta etapa puede utilizarse cualquier enzima de restricción siempre que corte en un sitio inmediatamente adyacente a la secuencia polimórfica. En una realización preferida, la enzima es una enzima de restricción de cuatro bases. En la técnica se conocen bien ejemplos de enzimas de restricción de cuatro bases e incluyen muchas que están disponibles en el comercio. Véase, por ejemplo, New England Biolabs Catalog 2000. Como ejemplos de enzimas de restricción de cuatro bases se incluyen, sin limitación, *AluI*, *Bsh1236I*, *DpnI*, *HpaI*, *MboI*, *MspI*, *PaiI* (un isosquízómero de *HaeIII*), *RsaI*, *Sau3AI* y *TaqI*. Después de la restricción, el fragmento de ADN se une a un soporte sólido. En la técnica se conocen bien numerosos soportes sólidos para inmovilizar ADN. Como ejemplos se incluyen, sin limitación, perlas de estreptavidina, que podrían unirse a un producto de PCR marcado con biotina, y perlas de anti-digoxigenina, que podrían unirse a un producto de PCR marcado con digoxigenina y perlas conjugadas con níquel, que podrían unirse a un producto marcado con seis histidinas. En una realización preferida, se utilizan perlas de estreptavidina (Figura 2).

Dado que la posición de la etiqueta de SARD (análisis en serie de ADN ribosómico) la dictamina el primer sitio de reconocimiento de la enzima de restricción, distal al cebador biotinilado utilizado en la reacción de PCR inicial, puede haber casos en los que el primer sitio de reconocimiento de la enzima de restricción se localice dentro de una región conservada del gen de interés. En general, esto no supone un problema porque aunque las etiquetas de la región conservada puede que no sean informativas, la mayoría de las etiquetas procedentes de SARD serán de una región polimórfica y serán informativas. Sin embargo, si se desea disminuir el número de etiquetas que contengan información de una región conservada de un gen en lugar de una región polimórfica, después de la restricción se pueden purificar los productos de la PCR deseados. En una realización preferida, esto puede llevarse a cabo purificando en gel aquellos productos de la PCR que tengan el tamaño esperado.

Etapa III. Digestión del producto de amplificación y ligamiento a enlazadores.

Los productos inmovilizados se dividen en dos grupos y los enlazadores se unen a los productos inmovilizados de cada grupo. Cada enlazador es una molécula de ADN sintética bicatenaria (doble cadena) que comprende una secuencia de ADN específica. Ambos enlazadores incorporan un sitio de enzima de restricción de tipo IIS. En una realización preferida, los dos enlazadores incorporan el mismo sitio de enzima de restricción de tipo IIS. Cada uno de los dos enlazadores también comprende una secuencia de ADN que se hibrida específicamente con un cebador. En una realización, los enlazadores son idénticos entre sí y se hibridan con el mismo cebador. En una realización preferida, los enlazadores son diferentes entre sí de tal manera que cada uno se hibrida con un cebador diferente.

El enlazador bicatenario se liga al producto de PCR inmovilizado. El enlazador puede incorporar el sitio de enzima de restricción de tipo IIS o puede incorporar solo una parte del sitio. En este caso, el enlazador se diseñará de tal manera que el ligamiento del enlazador con el ADN restringido reconstituya el sitio de restricción de tipo IIS. En una realización preferida, el enlazador incorpora un sitio *BpmI*. El ligamiento de enlazador es muy conocido y puede realizarse mediante cualquier método conocido en la técnica. Después del ligamiento, el producto PCR inmovilizado se aísla de los enlazadores libres mediante cualquier método conocido en la técnica. Véase, por ejemplo, Velcelescu *et al.*, Science 270: 484-487, 1995; Powell, Nucleic Acids Research 14: 3445-3446, 1998; Sambrook *et al.*, págs. F.8-F.10, 1989.

Las enzimas de escisión de tipo IIS escinden a una distancia definida de hasta 20 pares de bases, alejada de sus sitios de reconocimiento asimétricos. Las enzimas de restricción de tipo IIS que se encuentran disponibles en el comercio incluyen enzimas que dejan salientes en 5' y las que dejan salientes en 3' como productos de ADN bicatenarios. Algunas enzimas de la primera clase incluyen: *BsmFI* (10/14), *Bst71I* (8/12) y *FokI* (9/13), indicando el número entre paréntesis la posición de escisión en la misma cadena de ADN que la de la posición de la secuencia de reconocimiento/escisión en la cadena de ADN complementaria. Las enzimas de la segunda clase incluyen: *BpmI* (16/14), *BsgI* (16/14), *Eco57I* (16/14) y *GsuI* (16/14). El saliente en 3' dejado por estas enzimas debe retirarse para un ligamiento romo (etapa IV). Por lo tanto, las enzimas que escinden en las posiciones 16/14 dan como resultado una etiqueta de 14 pares de bases. Otras enzimas que cortan en una posición más distal pueden crear una etiqueta más grande. Por ejemplo, *MmeI* (20/18) deja un saliente en 3', pero no está disponible en el comercio (Tucholski, J. *et al.*, 1995, *MmeI*, a class-IIS restriction endonuclease: purification and characterization. Gene 157: 87-92).

Etapa IV. Digestión del producto con la enzima de restricción de Tipo IIS

El producto se digiere con la enzima de restricción de Tipo IIS apropiada para liberar un fragmento de ADN de la perla de anclaje y producir un fragmento de ADN híbrido corto que contenga una parte de la región polimórfica del ADN de interés (la etiqueta) y el ADN enlazador. Después de la digestión, el ADN debe rellenarse o digerirse para crear extremos romos. Si la enzima de restricción del Tipo IIS produce un saliente en 3', el fragmento se digiere con ADN polimerasa de T4 ADN para retirar el saliente en 3'. Si la enzima de restricción de Tipo IIS produce un saliente en 5', el saliente debe rellenarse utilizando desoxinucleótidos apropiados y el fragmento Klenow de la ADN polimerasa I. El fragmento de ADN se separa del resto del producto PCR inmovilizado. En una realización preferida, los dos grupos de productos de PCR inmovilizados se digieren con *BpmI* para liberar los marcadores polimórficos y

se digieren con ADN polimerasa de T4 para crear extremos romos (Fig. 2).

Etapa V. Ligamiento de etiquetas y amplificación por PCR de las dietiquetas resultantes

5 Utilizando métodos muy conocidos en la técnica las etiquetas se ligan entre sí por extremos romos para formar dietiquetas. Véase, por ejemplo, Sambrook *et al.*, 1989; Velculescu *et al.*, Science 270:484-487, 1995. Posteriormente, las dietiquetas se amplifican por PCR utilizando cebadores que son exclusivos para los enlazadores utilizados en la Etapa III. En una realización preferida, los cebadores son diferentes entre sí, si los enlazadores utilizados en la Etapa III eran diferentes entre sí. Como alternativa, los cebadores pueden ser iguales si los enlazadores utilizados en la Etapa III eran idénticos entre sí. El número de reacciones de amplificación por PCR variará dependiendo de la cantidad de ADN presente en el material de partida. Si hay una gran cantidad de ADN, entonces únicamente se requiere una amplificación PCR en esta etapa, en la que cada reacción comprende aproximadamente 15-30 ciclos. Si la cantidad de ADN inicial es baja, entonces en esta etapa puede requerirse más de una reacción de amplificación PCR.

15

Etapa VI. Escisión de las Dietiquetas y Ligamiento para formar Concatámeros de Dietiquetas.

Las dietiquetas se escinden con la enzima de restricción de cuatro bases utilizada en la Etapa II. Después, los productos se ligan para crear concatámeros de dietiquetas. En una realización, los concatámeros de dietiquetas varían de 2 a 200 dietiquetas. En una realización más preferida, el concatámero comprende 20-50 etiquetas polimórficas. El concatámero puede secuenciarse directamente, o puede clonarse en un vector de secuenciación. Utilizando un secuenciador de ADN capilar de 96 canales, en un día pueden analizarse fácilmente unas 12.000 dietiquetas. Como alternativa, los concatámeros pueden secuenciarse manualmente.

Métodos para analizar datos de marcadores

La invención se refiere a métodos para analizar la diversidad genética de una población en una muestra. Cada población que se analiza tendrá su propio conjunto de organismos o genes diferentes. El conjunto de datos que se captura de cada muestra debe recapitular la estructura genética en un formato de estudio que incluya un marcador para cada gen u organismo y la abundancia relativa de cada gen u organismo en la población en su totalidad. Los marcadores para una población particular a partir de perfiles de diversidad de marcadores (PDM), pueden introducirse en una base de datos. Véase, por ejemplo, la Fig. 8 que muestra un esquema para la generación de dicha base de datos. El método a través del cual se capturan los datos no es crítico siempre que produzca una representación exacta de cada población.

35

En un aspecto del método, los PDM se introducen en una base de datos. En una realización preferida, la base de datos se mantiene en una forma legible por ordenador, tal como un disquete, o un disquete no extraíble, o un servidor en red, o en una página Web. Sin embargo, el método mediante el cual se capturan los datos no es crítico siempre que se produzca una representación exacta de cada comunidad microbiana.

40

Los sistemas de inteligencia artificial (IA) pueden ejecutar muchas funciones de gestión y análisis de datos. Los ejemplos de sistemas IA incluyen sistemas expertos y redes neuronales. Los sistemas expertos analizan datos de acuerdo con una base de conocimiento junto con una base de datos residente. Las redes neuronales están constituidas por unidades de procesamiento interconectadas que pueden interpretar múltiples señales de entrada y generar una sola señal de salida. Los sistemas IA se adaptan perfectamente para analizar sistemas biológicos complejos que incluyen poblaciones utilizando protocolos de deducción.

45

Un marcador puede correlacionarse con una condición particular o con otro marcador. Véase, por ejemplo, la Fig. 9 para un esquema de las etapas implicadas en la determinación de parámetros particulares asociados con un PDM y la Fig. 10, que muestra un esquema de generación de una base de datos matricial de diversidad de marcadores. Una condición o estado puede ser una condición ambiental, tal como pH, temperatura, salinidad, o la presencia o la ausencia de un compuesto orgánico o inorgánico, tal como hidrocarburos, nitratos o depósitos minerales. Una condición puede ser una condición fisiológica o médica, tal como una patología aguda o crónica, un estado fisiológico, una fase de desarrollo o puede estar asociada a un tejido o fluido corporal particular. Junto con los PDM también puede grabarse información con respecto a todos los parámetros conocidos asociados a las muestras.

50

Cada perfil de diversidad de marcadores (PDM) está compuesto por marcadores que representan un pequeño número de especies o genes, más preferentemente una especie o un gen. Por ejemplo, en el caso del Ejemplo1, cada marcador comprendería una secuencia de ADNr 16S polimórfica de 12 pares de bases. Dichos parámetros que están relacionados con muestras ambientales pueden incluir componentes inorgánicos (obtenidos a través de análisis de adsorción atómica), componentes orgánicos (obtenidos a través de GC-MS o LC-MS), análisis de tamaño de grano, pH y salinidad. Los parámetros que están relacionados con muestras médicas incluirían, pero sin limitación, un historial médico completo del donante. Véase, por ejemplo, la Fig. 11, que muestra un esquema para aplicaciones de mapeo utilizando perfiles de diversidad de marcadores.

60

En otro aspecto de la invención, los PDM se recogen durante una evolución temporal, y cada momento es uno de los

65

parámetros incluidos. Las evoluciones temporales pueden ser útiles para rastrear cambios con el tiempo para una amplia variedad de indicaciones. Por ejemplo, las evoluciones temporales pueden ser útiles para rastrear la progresión de una enfermedad, durante esfuerzos de remediación ambiental y durante la producción de campos petrolíferos.

5 En otro aspecto de la invención, los PDM se recogen durante una evolución temporal, y cada momento es uno de los parámetros incluidos. Las evoluciones temporales pueden ser útiles para rastrear cambios con el tiempo para una amplia variedad de indicaciones. Por ejemplo, las evoluciones temporales pueden ser útiles para rastrear la progresión de una enfermedad, durante esfuerzos de remediación ambiental y durante la producción de campos petrolíferos.

10 En otro aspecto de la invención, los PDM se recogen en diversas localizaciones distintas, tal como en diversas localizaciones geográficas o en diversos tejidos corporales. La comparación de los PDM recopilados de diversas localizaciones distintas es útil para diferenciar cambios entre estas diversas localizaciones, lo cual puede ser indicativo de condiciones ambientales o patologías particulares.

15 La comparación de perfiles de diversidad de marcadores puede revelar tendencias en poblaciones ya sea relativas al tiempo o a una localización geográfica. En este primer caso, las comparaciones de poblaciones microbianas pueden resolver información espacial sobre el ambiente que de otra manera no podrían detectarse. Como ejemplos de dicha información se incluyen patrones de migración de agua, de compuestos orgánicos y de minerales. Por ejemplo, los depósitos placeres de minerales están causados por la acción tanto del agua como del viento haciendo que los minerales migren desde un depósito de lodo a una localización únicamente para depositarse en otra localización. La migración de dichos minerales puede conducir a una traza detectable sobre las poblaciones microbianas en la trayectoria de migración. También pueden detectarse atributos físicos del entorno, tales como estructuras, formaciones y líneas de falla. Normalmente se entiende que las fallas ofrecen una vía de migración vertical significativa para gases, tales como metano, que se sabe los microbios utilizan diferencialmente. Combinando los datos PDM con coordenadas geográficas, tales como altura, longitud y latitud, que pueden obtenerse fácilmente con dispositivos de sistema de posicionamiento global, es posible crear mapas que delimiten la distribución de diversos microbios en el entorno.

20 Los análisis de correlación entre un marcador y todos los marcadores restantes en la base de datos, revelarán pares de marcadores con propensión a coincidir. Este proceso puede repetirse de una manera repetitiva en todos los marcadores para producir una matriz de coeficientes de correlación entre todos los marcadores observados. La Fig. 5 muestra un dispersograma de dos pares de marcadores, presentando uno de los pares un alto grado de correlación. Esta estrategia también puede utilizarse para crear un dendrograma que refleje el nivel de correlación relativo entre cada marcador. Por lo tanto, en cualquier nivel de correlación seleccionado, todos los marcadores observados pueden dividirse en grupos en los que los marcadores de cada grupo comparten el mismo nivel de correlación con cada otro miembro del grupo. Si se selecciona un valor alto de coeficiente de correlación (p, ej., 0,8), los marcadores de cada grupo podrían, en la mayoría de los casos, encontrarse en la misma muestra. Por lo tanto, este ejercicio dividirá una población determinada en grupos de genes u organismos que tengan una propensión a localizarse conjuntamente entre sí. En una realización preferida, el ejercicio dividirá una comunidad microbiana en grupos de microbios que tengan una propensión a localizarse conjuntamente.

45 Los análisis de correlación entre un marcador (variable 1) y un parámetro de muestra (variable 2) identificarán marcadores cuya presencia a menudo, o invariablemente, coincida con un componente presente en las muestras. En La Fig. 6 se muestran algunos tipos de relaciones entre marcadores y componentes de muestra (o parámetros). Un fuerte valor de correlación entre un marcador y un parámetro de muestra permitirá realizar predicciones sobre la abundancia de cualquier variable (marcador o parámetro de muestra) siempre que se conozca una las variables.

50 En algunos casos, un marcador no será específico para una sola especie o gen. Por ejemplo, la secuencia etiqueta que se identificaría por la estrategia representada en el Ejemplo 1 identificaría *Denitrobacter permanens* y *Legionella anisa*. En los casos en los que se encuentra una correlación significativa entre un marcador y un parámetro de muestra de interés, la acción preferida es utilizar la información de secuencia de etiqueta para identificar la secuencia génica completa. Después la secuencia puede utilizarse para identificar las secuencias y para identificar sondas específicas de especie para verificar la correlación. Esto puede realizarse utilizando métodos conocidos en la técnica, tales como mediante PCR o mediante hibridación a moléculas de ADN aisladas de la muestra de interés, seguido de secuenciación o de otros métodos de análisis.

60 Las sondas específicas de especie, que se identifican de marcadores con una fuerte correlación con un parámetro de muestra de interés, pueden utilizarse después como un diagnóstico, o para prospectar el parámetro de interés. Dichos ensayos estarían basados preferentemente en PCR y serían muy sensibles, rápidos y asequibles. En una realización preferida, un marcador identificado por estos métodos puede utilizarse como una sonda de hibridación para identificar un trozo de ADN más grande del cual procede este marcador. La secuencia de la molécula de ADN más grande puede después utilizarse para diseñar cebadores que se hibriden específicamente con la molécula de ADN de interés y que pueden utilizarse para amplificar específicamente la molécula de ADN por PCR. Como alternativa, se puede utilizar un ensayo basado en hibridación utilizando una sonda que se una específicamente a la

molécula de ADN de interés. El uso de cebadores o de sondas es especialmente útil para determinar rápidamente si una gran cantidad de muestras contiene la molécula de ADN que se correlaciona con el parámetro de interés.

En una realización preferida, se identifica un marcador que se correlaciona con un parámetro deseado. El marcador puede identificarse utilizando SARD, o puede identificarse utilizando otro método, tal como polimorfismos de longitud de fragmentos de restricción (RFLP, del inglés *Restriction Fragment Length Polymorphism*) o *polimorfismos* de longitud de fragmentos de restricción terminales (T-RFLP, del inglés *Terminal Restriction Fragment Length Polymorphism*, Liu *et al.*, Applied and Environmental Microbiology 63:4516-4522, 1997). Para identificar diferencias de tamaño, puede utilizarse un método tal como electroforesis en gel con gradiente desnaturante (DGGE, del inglés *Denaturing gradient gel electrophoresis*). En una realización preferida, para identificar el marcador se utiliza SARD. Otras muestras se exploran para determinar si tienen el marcador de interés. En una realización preferida, la exploración utilizada es PCR o hibridación, más preferentemente es PCR. En una realización aún más preferida, el marcador se correlaciona con la presencia de reservas minerales o petrolíferas.

El análisis de correlación entre los PDM (PDM_n, variable 1; PDM_{n+1}, variable 2) puede revelar las similitudes relativas entre muestras. Se espera que muestras tomadas del mismo individuo o de sitios ambientales próximos, que tienen similar composición, muestren un fuerte coeficiente de correlación (Fig. 7A). Sin embargo, se espera que las muestras que comparten solo un parámetro, o solo algunos parámetros en común, no muestren una correlación significativa cuando se consideran todos los marcadores (Fig. 7B). Incorporando el conocimiento aprendido de las correlaciones entre Marcadores y entre Marcador/Parámetro, los PDM pueden compararse utilizando en el análisis marcadores individuales o preferentemente, subconjuntos de marcadores correlacionados (Fig. 7C). Esta estrategia puede eliminar mucha interferencia y permite identificar relaciones ocultas.

El análisis de correlación puede realizarse mediante cualquier método o cálculo conocido en la técnica. El análisis de correlación para r y r^2 puede realizarse como describe M. J. Schmidt en Understanding and Using Statistics, 1975 (D.C. Health and Company), págs. 131-147. El grado de correlación para r puede definirse de la siguiente manera:

1,0	Perfecto
0,8-0,99	Alto
0,5-0,7	Moderado
0,3-0,4	Bajo
0,1-0,2	Insignificante

En una realización, la correlación entre los dos marcadores o entre un marcador y un parámetro es al menos baja (r es de 0,3 a 0,4). En una realización preferida, la correlación es al menos moderada (r es de 0,5 a 0,7). En una realización más preferida, la correlación es alta (r es de 0,8 a 0,99).

Con el desarrollo de numerosas tecnologías genómicas para analizar conjuntos complejos de ácidos nucleicos, se tiene la oportunidad de comenzar a catalogar el reservorio de la diversidad microbiana, y por tanto, metabólica. Dado que la proliferación de un microbio en una localización determinada dependerá de la presencia de los nutrientes metabólicos necesarios, la información como la abundancia de ese microbio puede servir como un biosensor para un conjunto de parámetros determinado. Cuando se considera como un conjunto, la estructura de la comunidad microbiana en una determinada localización, contendrá un potencial biosensor intrínseco para una amplia serie de parámetros. La fiabilidad predictiva de los datos de una comunidad microbiana completa también aumentará significativamente. Por ejemplo, si un microbio determinado estuviese presente en el 50 % de las muestras de suelo extraídas por encima de los depósitos de petróleo y no se encontrase en ningún otro lugar, entonces la presencia de diez de dichos microbios podría crear un valor predictivo con una precisión de 99,9 %.

Aplicaciones de la invención

45 *Exploración Geoquímica y Mineral*

Los métodos descritos en la presente invención tienen diversos beneficios sobre las tecnologías existentes. Por ejemplo, en el campo de la exploración geoquímica, análisis basados en ADN_r genómico posiblemente podrán resolver un amplio conjunto de parámetros geoquímicos de interés con respecto a las industrias del petróleo y minera. Actualmente, se requieren muchas tecnologías diferentes para medir estos parámetros. Dado que esta invención se basa en una medición universal, la detección de ácido nucleico, esta puede reducir enormemente la instrumentación y los costes externalización de muestras.

Los depósitos de aceite y gas se localizan muy cerca de la superficie de la tierra a profundidades que varían desde unos 30,4 cm (100 pies) a más de 3048 ms (10.000 pies). Cuando se forma el aceite, este experimenta una migración en la que se produce una de dos cosas. El aceite puede continuar migrando hasta que finalmente alcanza la superficie, donde se evapora con el tiempo. Como alternativa esta migración puede bloquearse mediante una estructura impermeable, denominada "trampa". Los métodos geofísicos (tales como métodos sísmicos tridimensionales) para la exploración de petróleo se basan en el hallazgo de estas estructuras trampa con la esperanza de que contengan aceite.

El aceite crudo está constituido por una diversidad de longitudes de cadena de hidrocarburo. Los hidrocarburos más ligeros (en particular, metano, etano, propano y butano) a menudo pueden difundirse a través de las estructuras trampa y, como resultado de los gradientes de presión, experimentar una migración vertical hacia la superficie. Determinados microbios presentes en la superficie o en la capa de la superficie pueden utilizar estos hidrocarburos en migración, lo que ocasionalmente produce cambios mineralógicos que son detectables en la superficie. Por tanto, cabe esperar que estos hidrocarburos en migración afecten a las poblaciones microbianas, de tal manera que la capacidad de determinar la diversidad genética de una población microbiana puede revelar firmas microbianas que sean indicativas de la presencia de aceite.

Avances recientes en microfluídica en la industria genómica han dado como resultado el desarrollo de instrumentos que pueden detectar ácidos nucleicos específicos en pocos minutos. La utilización de dichos instrumentos permitirá realizar en el campo mediciones de diversos parámetros. En cambio, los ensayos químicos convencionales requieren análisis de laboratorio e interpretación.

Se han creado biosensores con capacidad de detectar hidrocarburos que están presentes al nivel de detección, o por debajo de este, de instrumentación analítica de tipo GC-MS sofisticada (Sticher, P. *et al.*, 1997, Development and characterization of a wholecell bioluminescent sensor for bioavailable middle-chain alkanes in contaminated groundwater samples. *Appl Environ Microbiol.* 63:4053-4060). Sticher *et al.* demostraron que utilizando un solo gen indicador en un microbio modificado genéticamente que comprende un gen indicador, se podían detectar cambios extremadamente pequeños en su ambiente en respuesta a un tratamiento agudo con un hidrocarburo particular. La presente invención puede documentar el efecto sobre una población que comprende miles de microbios durante el tiempo geológico y por tanto tiene la posibilidad de ser más sensible que los instrumentos analíticos actuales.

La presente invención también puede utilizarse para crear un estudio de entidades biológicas que esté limitado solo por el requisito previo de que estas entidades contengan ácidos nucleicos que estén dispuestos en regiones que están conservadas y en regiones que sean polimórficas cuando se comparan con secuencias de organismos relacionados. Más adelante se describen algunos ejemplos adicionales de la aplicación de la presente invención.

Desarrollo de depósitos de aceite y gas

Además de la aplicación de la presente invención en la exploración del petróleo, la presente invención también puede utilizarse en el desarrollo de depósitos de aceite y gas. Diversas propiedades de depósitos de aceite que afectan directamente a la viabilidad comercial de los depósitos las modulan, a un determinado nivel, los microbios. Algunas veces hay sulfuro de hidrógeno en el aceite crudo y puede convertir, de otra manera, el aceite 'dulce' en aceite 'avinagrado'. Además de su efecto corrosivo sobre el equipo del campo petrolífero, el H₂S también plantea un riesgo para los trabajadores y reduce significativamente el valor de un depósito de aceite debido a que para eliminar el gas debe instalarse una planta de lavado. Los niveles de H₂S pueden cambiar durante el desarrollo de un depósito y ahora se piensa que es el resultado de bacterias reductoras de sulfato (Leu, J.-Y. *et al.*, 1999, 1999, The same species of sulphate-reducing Desulfomicrobium occur in different oil field environments in the north sea. *Lett. Appl. Microbiol.* 29(4):246-252). Identificando la presencia de microbios que pueden conducir a la producción de H₂S, la valoración de nuevos depósitos y las estrategias de desarrollo resultantes podría volverse más eficaces.

El aceite crudo y el gas natural están compuestos por una mezcla compleja de hidrocarburos que incluyen hidrocarburos de cadena lineal de longitudes generalmente comprendidas entre 2 y 40 átomos de carbono. Los hidrocarburos de longitud de cadena más corta son más valiosos (p. ej., gasolina, C₄-C). En algunos depósitos de aceite, los hidrocarburos más ligeros se extraen selectivamente durante o antes del desarrollo del depósito. Se sospechaba que los microbios desempeñaban una función en este proceso ya que los hidrocarburos de longitud de cadena más corta están más biodisponibles. La presente invención puede identificar microbios que están implicados en este proceso y por lo tanto puede hacer predicciones en cuanto a la susceptibilidad de determinados depósitos del empobrecimiento de dichos hidrocarburos de cadena corta. La presente invención también permite identificar microbios capaces de acortar los hidrocarburos de cadena larga aumentando por lo tanto el valor de los depósitos existentes.

Detección de insectos y parásitos

El significativo impacto negativo que los insectos pueden tener sobre la agricultura es muy conocido. Los insectos también sirven como vectores para la transmisión de muchos microbios causantes de enfermedades. Se han descrito numerosas relaciones entre insectos y microbios. Por ejemplo, el género de bacterias *Wolbachia* se encuentra asociado con muchas especies de hormigas y se ha observado que altera la determinación del sexo y la fecundidad en el hospedador (Wenseleers, T. *et al.*, 1998, Widespread occurrence of the micro-organism *Wolbachia* in ants. *Proc. R Soc. Lond. B. Biol. Sci.* 265(1404):1447-52). Además, en las hormigas se han identificado muchas especies de bacterias endosimbióticas intracelulares (Schroder, D. *et al.*, 1996, Intracellular endosymbiotic bacteria of *Camponotus* species (carpenter ants): systematics, evolution and ultrastructural characterization. *Mol. Microbiol.* 21:479-89). Probablemente, la mayoría de los insectos, si no todos, tienen relaciones íntimas, específicas de especie, con microbios, lo que podría representar un talón de Aquiles para el control de poblaciones de insectos. La invención descrita en esta solicitud proporcionaría un medio para identificar microbios que modulen el bienestar de

una especie de insecto determinada.

Posiblemente, la identificación de microbios que están específicamente asociados a un insecto determinado, también podría servir como una base para un ensayo muy sensible para determinar la presencia del insecto. Por ejemplo, los métodos actuales para identificar la presencia de termitas en estructuras de madera se basan en inspección visual y no son muy adecuados. Un ensayo para detectar la presencia de un microbio asociado a termitas que se base en amplificación por PCR sería tanto muy sensible como no invasivo.

Adicionalmente, la capacidad de crear inventarios exhaustivos de diversidad microbiana tiene diversas aplicaciones en lo que respecta a ecología microbiana que podría ser útil en la industria agrícola. Por ejemplo, la industria agrícola utiliza enormes cantidades de pesticidas para prevenir profilácticamente la pérdida de cultivos. La presente invención proporciona la capacidad de realizar estudios exhaustivos de poblaciones microbianas y puede conducir a predicciones tales como la susceptibilidad de un campo determinado contra fitopatógenos particulares. Este conocimiento conduciría a mejorar las estrategias de las aplicaciones pesticidas.

Adicionalmente, el estudio de la diversidad microbiana en un entorno, tal como un campo agrícola en diversos momentos del año, revelaría los cambios dinámicos que se producen en las poblaciones microbianas como resultado de fluctuaciones estacionales (temperatura y humedad), aplicación de pesticidas y proliferación de determinados organismos.

La comparación de los perfiles de diversidad extraídos del mismo sitio o de un sitio similar en diferentes momentos revelaría interacciones entre especies en una población. Estas interacciones productivas pueden manifestarse en sí mismas bien en el aumento o disminución en la representación de un marcador relativo a la disminución o aumento en la representación de un segundo marcador, respectivamente.

Dicha información proporcionaría una advertencia temprana de la proliferación de especies de plagas, así como la identificación de especies que son patógenas contra especies de plagas. Esos organismos patógenos pueden tener valor bien como un agente biológico para controlar la proliferación de especies de plagas y/o como una fuente para genes o compuestos que actuarían como pesticidas. Este fenómeno no está limitado a interacciones entre microbios. Los huevos, así como las fases larvares de los insectos, probablemente interactúan con los microbios del suelo y crean un impacto detectable sobre las poblaciones microbianas. Un ejemplo de tal organismo es *Bacillus thuringiensis*, que en sí mismo se utiliza normalmente en la agricultura orgánica como un agente insecticida. El gen responsable de esta actividad insecticida (toxina Bt) se ha utilizado en gran medida para crear plantas transgénicas con resistencia al ataque causado por insectos.

Biorremediación

Se ha realizado una cantidad de esfuerzo considerable en el desarrollo de métodos para la eliminación, basada en microbios, de productos químicos del medio. Dichos productos químicos incluyen metales pesados, aromáticos polinucleares (PNA "polynuclear aromatics"), aromáticos halogenados, aceite crudo y diversos otros compuestos orgánicos tales como MTBE. Las consideraciones reguladas para la eliminación de microorganismos en el medio han redirigido esfuerzos hacia la identificación y aumento del crecimiento de organismos endémicos que tienen la capacidad de metabolizar o eliminar del medio compuestos de interés.

La presente invención puede facilitar esfuerzos de biorremediación en dos sentidos. En primer lugar, pueden identificarse organismos en un lugar determinado que previamente se haya demostrado que pueden eliminar compuestos de interés o que tengan una probabilidad significativa de tener la capacidad de metabolizar los compuestos relevantes basándose en su coincidencia en el medio con los productos químicos en numerosos escenarios geológicos. En segundo lugar, la presente invención puede identificar tendencias de tipos de suelos y especies microbianas particulares. En este caso, las correlaciones que se extraen desde las bases de datos entre la distribución microbiana y los tipos de suelo pueden combinarse con la base del conocimiento de geoquímica existente. Por ejemplo, el Servicio Geológico de los Estados Unidos, USGS, proporciona muchos mapas disponibles al público que describen numerosos parámetros geoquímicos y geofísicos. Las extrapolaciones de la distribución de las especies microbianas pueden realizarse a nivel regional, y posiblemente mundial. Estas distribuciones microbianas extrapoladas pueden servir como base para regímenes de tratamiento específicos de sitio para aumentar el crecimiento de determinadas especies relevantes sin realizar primero un estudio microbiano.

Ciencia forense

La presente invención, particularmente el análisis SARD, también puede utilizarse en aplicaciones forenses. En los últimos diez a veinte años se han enfocado estudios sobre los cambios que se producen en el cuerpo de una persona o de un animal después de su muerte. Muchos de estos cambios implican cambios en las poblaciones microbianas que aparecen durante la descomposición del organismo. Los cambios en dichas poblaciones microbianas se han correlacionado con el tiempo que transcurre desde que una persona muere y con las condiciones que el cuerpo experimenta después de la muerte, por ejemplo, calor, exposición solar, inhumación parcial o completa, exposición a la lluvia, etc. Los análisis SARD permitirían a los científicos forenses determinar de

un modo rápido y exacto el tamaño y los tipos de las poblaciones microbianas, que a su vez pueden utilizarse para determinar tiempos más precisos de muerte así como las condiciones a las que el cuerpo ha estado expuesto.

Otras aplicaciones

5 Esta estrategia puede utilizarse para cualquier región polimórfica en un genoma microbiano. Este método tampoco requiere limitarse a genes. La secuencia de ADN de regiones intergénicas de genomas no está bajo un alto nivel de presión selectiva y por tanto, representa una secuencia de ADN muy polimórfica. Un ejemplo de dicha región es la región intergénica entre las regiones que codifican las subunidades de ADNr grande (23S) y pequeña (16S). Los métodos descritos anteriormente pueden utilizarse para diferenciar miembros de una población basándose en diferencias de tamaño de la región intergénica entre los genes de ADNr 16S-23S o 23S-5S. La región espaciadora entre estos genes se ha descubierto que es hipervariable en poblaciones microbianas. En el caso de la región intergénica 16S-23S, el tamaño del espaciador varía entre aproximadamente 200-1500 pares de bases dependiendo de la presencia o de la ausencia de diversos genes de ARNt. (Nour M. 1998, 16S-23S and 23S-5S intergenic spacer regions of lactobacilli: nucleotide sequence, secondary structure and comparative analysis, Res. Microbial. 149(6):433-448; Berthier F. *et al.*, 1998, Rapid species identification within two groups of closely related lactobacilli using PCR primers that target the 16S/23S rRNA spacer region, FEMSMicrobiol Lett. 161(1):97-106; Tilsala-Timisjarvi A. *et al.*, 1997, Development of oligonucleotide primers from the 16S-23S rRNA intergenic sequences for identifying different dairy and probiotic lactic acid bacteria by PCR Int. J. Food Microbiol. 35(1):49-56). Estos dos genes de ADNr se transcriben en el mismo operón. Por lo tanto, las regiones conservadas de la secuencia codificante de ADNr de estas subunidades pueden utilizarse para amplificar las regiones intergénicas.

Para que la presente invención pueda entenderse mejor, se exponen los siguientes ejemplos. La finalidad de estos ejemplos es únicamente ilustrativa y en ningún caso deben considerarse como limitantes del alcance de la invención.

25 **EJEMPLO 1**

Análisis en serie de etiquetas polimórficas de ADNr del dominio *Eubacteria*

30 Se obtuvo una muestra que comprendía bacterias ambientales y se extrajo ADN total de la muestra. Para amplificar el ADN, se realizó PCR mezclando tampón de reacción Advantage2 5 µl 10X (Clontech), los dNTP 2,5 µl, cebadores TX9/TX16 5 µl, 8 µM, 50 ng de ADN de muestra, Taq polimerasa Advantage2 0,5 µl (Clontech) y agua hasta 50 µl. El cebador TX9 se marcó con biotina. Los cebadores se muestran en la Fig. 12 y en la Fig. 13 se muestra una estrategia general de este ejemplo. Después, la mezcla de reacción se sometió a PCR con las siguientes condiciones:

- a) 94 °C durante 5 minutos;
- b) 94 °C durante 1 minuto, 10 segundos;
- c) 55 °C durante 50 segundos;
- 40 d) 68 °C durante 1 minuto;
- e) repetir las etapas (b) a (d) 20 veces; y
- f) 68 °C durante 2 minutos.

45 El producto de la PCR, de aproximadamente 600 pares de bases, se purificó en gel utilizando agarosa al 1 % y un kit Qiaquick (Qiagen) de acuerdo con las instrucciones del fabricante. El ADN se eluyó con Tris-EDTA (TE; Tris 10 mM pH 8; EDTA 1 mM) 50 µl. Para restringir el ADN amplificado, se añadió tampón n.º 2 New England Biolabs (NEB n.º 2) 10X para llevar la concentración del tampón a 1X y se añadieron 25 unidades de *A**l**u**I*. La mezcla de reacción se incubó durante 2 horas a 37 °C; se añadieron 25 unidades MÁS de *A**l**u**I* y la mezcla de reacción se incubó durante una hora más seguido de inactivación de la enzima a 65 °C durante 20 minutos.

50 Para inmovilizar en una perla el fragmento de ADN restringido, se añadió el mismo volumen de tampón BW 2X (2X BW = Tris 10 mM, pH 7,5; EDTA 1 mM; NaCl 2 M) a la mezcla de reacción, seguido de la adición de perlas de Estreptavidina (SA) M-280 (Dynal) lavadas 50 µl. La mezcla de reacción se incubó durante 20 minutos a temperatura ambiente. Después, las perlas se lavaron dos veces en BW 1X y dos veces en tampón de lavado (Tris 10 mM, pH 8, MgSO₄ 10 mM; NaCl 50 mM). Durante la última etapa de lavado, las perlas se dividieron en dos grupos.

60 Para añadir los enlazadores al ADN inmovilizado, un grupo de perlas se resuspendió en 4 µl de enlazador TX-12/13 10 µM (una molécula de ADN bicatenario que comprende los cebadores TX-012 y TX-013) para formar el grupo A, mientras que el otro grupo (grupo B) se resuspendió en 4 µl de enlazador TX-14/15 10 µM (una molécula de ADN bicatenario que comprende los cebadores TX-014 y TX-015) al que se añadió el otro grupo. Se añadieron 36 µl de mezcla de ligasa de T4 (tampón ligasa 10X, 4 µl; agua 32 µl, ligasa T4 0,2 µl [400 U]) a la mezcla de enlazador/perla y la mezcla se incubó durante una noche a 16 °C.

65 La molécula de ADN unida a las perlas de estreptavidina se incubó después con *B**p**m**I*, que reconoce su secuencia de restricción específica (GAGGTC) en la molécula de ADN. *B**p**m**I* escinde el ADN de tal manera que libera la perla de estreptavidina de la molécula de ADN e incorpora una parte de la región polimórfica de la primera molécula de

ADN amplificada. Las perlas se lavaron dos veces con 0,5 ml cada vez con 1 x BW y dos veces con tampón de lavado. Las perlas se resuspendieron en una mezcla de *Bpml* 10 µl (10 X NEB n.º 3 1 ml [New England Biolabs], albúmina de suero bovino 1 µg/µl 1 µl, agua 8 µl y *Bpml* 1 µl [New England Biolabs]). La mezcla de reacción se incubó a 37 °C durante 2 horas y después a 65 °C para inactivar la enzima. El sobrenadante que contenía las etiquetas de ADN se aisló después.

Para eliminar el saliente 3' de las etiquetas de ADN y hacer romos los extremos, al sobrenadante (10 µl) en hielo se le añadió la mezcla de polimerasa de T4 10 µl (1 µl 10X NEB n.º 2; dNTP 4 mM 0,5 µl, agua 8,5 µl y polimerasa de T4 0,33 µl [1 U, New England Biolabs]). La reacción se incubó a 12 °C durante 20 minutos y a 65 °C durante 20 minutos para inactivar la polimerasa. Para formar dietiquetas, se combinaron el grupo A y el grupo B para dar un volumen total de 40 µl. Se añadieron 4 µl de ATP_r 10 mM y 0,2 µl de ADN ligasa de T4 (400 U) y la reacción se incubó desde 4 horas hasta toda la noche a 16 °C.

Como una etapa de amplificación intermedia, las dietiquetas se amplificaron después en una reacción PCR 300 µl (tampón Taq Advantage2 10X 30 µl, dNTP 4 mM 15 µl; mezcla de cebadores TX111/TX12130 30 µl 8 µM; Taq polimerasa Advantage2 3 µl; molde de dietiqueta 3 µl; agua 219 µl). La mezcla de reacción de 300 µl se dividió en tres reacciones de 100 µl y se amplificó utilizando las siguientes condiciones:

- a) 94 °C durante 5 minutos;
- b) 94 °C durante 30 segundos;
- c) 56 °C durante 30 segundos;
- d) 68 °C durante 40 segundos;
- e) repetir las etapas (b) a (d) durante 15 ciclos; y
- f) 68 °C durante 2 minutos.

Después de la amplificación, se añadieron tres volúmenes (900 µl) de tampón QG (Qiagen) y un volumen (300 µl) de isopropanol, la mezcla se unió a una columna de espín de extracción en gel (Qiagen) y el ADN se eluyó con TE 50 µl

El ADN amplificado se sometió después a electroforesis en gel de poliacrilamida no desnaturante (PAGE; Novex 1 mm 10 % gel TBS-PAGE). El gel se tiñó con bromuro de etidio 5 µg/ml, las dietiquetas de aproximadamente 106 pares de bases se escindieron del gel, el gel escindido se fragmentó y se añadieron 0,3 ml de TE al gel y se incubó a 65 °C durante 30 minutos. La mezcla de gel/TE se transfirió a una columna de espín miniprep (Qiagen) y el eluato que contenía las dietiquetas amplificadas se recogió (molde de dietiqueta). El ADN amplificado se denominó molde de dietiqueta P300.

Para determinar el número óptimo de ciclos PCR para la amplificación a gran escala (titulación del ciclo PCR), se mezclaron 6,4 µl de molde de dietiqueta P300 con 8 µl de tampón Taq Advantage2 10X, dNTP 4 µl 4 mM, cebadores TX111/TX121 8 µl, Taq polimerasa Advantage2 0,8 µl y agua 52,8 µl. La mezcla de reacción se dividió en tres reacciones de 25 µl y se amplificó durante 6, 8 o 10 ciclos utilizando las condiciones PCR descritas anteriormente para la etapa de amplificación intermedia. Después de la amplificación, los productos de ADN se desalinizaron y purificaron utilizando una Columna de espín de Extracción en Gel de Qiagen como se describe anteriormente excepto que se únicamente se utilizaron 75 µl de tampón QG y 25 µl de isopropanol, y los productos de ADN se eluyeron con 20 µl de TE. El número de ciclos que dio lugar a la dietiqueta más larga de 106 pares de bases se produjo sin ninguna mancha vertical detectable por análisis TBS-PAGE al 10 % que se utilizó para la amplificación a gran escala.

Para la amplificación de dietiquetas a gran escala, se mezclaron 180 µl de tampón Taq Advantage2 10X, dNTP 90 µl 4 mM, mezcla de TX111/TX121 180 µl 8 µM, 18 µl de Taq polimerasa Advantage2, 144 µl de molde de dietiqueta P300 y 11881 µl de agua conjuntamente y después se dividió en 18 reacciones de 100 µl. La PCR se realizó como se ha descrito en líneas anteriores utilizando el número de ciclos que se determinó para la etapa de titulación anterior. Las reacciones de la PCR se desalinizaron añadiendo tres volúmenes de tampón QG y un volumen de isopropanol a las reacciones y se pasaron a través de un total de cuatro columnas de espín de extracción en gel de Qiagen. Cada columna se eluyó con 50 µl de TE. Las muestras se sometieron a PAGE preparativa no desnaturante (Novex 1,5 mm 10 % TBS-PAGE). Las bandas de ADN se tiñeron y se escindieron como se ha indicado anteriormente y se dividieron en tres tubos. La poliacrilamida se fragmentó y se eluyó utilizando TE como se ha descrito anteriormente y el ADN se purificó pasando el contenido de la mezcla gel/TE a través de una columna Miniprep Spin Plasmid Qiagen como se ha descrito anteriormente.

Para precipitar el ADN, se añadieron 1 µl de glicógeno, 150 µl de acetato de amonio 10M y 1125 µl de etanol a cada eluato de aproximadamente 300 µl. La mezcla se incubó a -80 °C durante 20 minutos y se microcentrifugó a 13000 rpm a 4 °C durante 15 minutos. El sedimento se lavó con etanol al 70 % 1 ml y se secó a 37 °C durante 5 minutos.

Los sedimentos de ADN se resuspendieron en mezcla de *Alul* 150 µl (15 µl de NEB n.º 2, 135 µl de agua y 15 µl de *Alul* [150 U]). La mezcla de reacción se incubó a 37 °C durante 2 horas. Se añadieron 150 µl de tampón BW 2X y la muestra se transfirió a un nuevo tubo que contenía 150 perlas SA compactadas. Las perlas se incubaron con la

mezcla de ADN durante 30 minutos a temperatura ambiente y las perlas se sedimentaron magnéticamente. El sobrenadante se retiró, se extrajo con fenol/cloroformo/alcohol isoamílico 250 µl y la fase acuosa se precipitó con acetato de amonio 150 µl 10M y etanol 1125 µl durante una noche a -80 °C. El ADN se centrifugó a 13000 rpm a 4 °C durante 15 minutos, el sedimento se lavó con etanol al 70 % y se secó. Esta etapa elimina cualquier enlazador libre que pueda haber, así como cualquier producto de ADN incompletamente digerido que aún permaneciese unido a los enlazadores marcados con biotina.

Para concatenar las dietiquetas, el sedimento de ADN se resuspendió en mezcla de ligasa 10 µl (tampón ligasa 10X 1 µl; agua 9 µl y T4 ligasa 1 µl [100 U]) y se incubó a 16 °C durante 30 minutos. El ADN se precipitó añadiendo 40 µl de TE, 25 µl de acetato de amonio 10M y etanol 180 µl. El ADN se precipitó durante 15 minutos a -80 °C y se centrifugó, se lavó y secó como se ha indicado anteriormente.

Las dietiquetas concatenadas se purificaron, resuspendiendo el sobrenadante en tampón de carga TBE 1X y sometiendo la muestra a PAGE no desnaturalizante (TBE-PAGE al 8 %). El área entre 0,5 a 1,2 kb se escindió y el ADN se eluyó del gel en TE 300 µl como se ha descrito anteriormente. El ADN se precipitó utilizando glicógeno 1 µl, acetato de amonio 150 µl 10M y 1125 µl de etanol. El ADN se precipitó, se centrifugó, se lavó y se secó como se ha indicado anteriormente.

Las dietiquetas concatenadas se clonaron en pUC19 resuspendiendo el sedimento de ADN en pUC19 (ADN plasmídico, aproximadamente 100 ng) cortado con *Sma*I 3 µl y añadiendo mezcla de T4 ligasa 7 µl (tampón ligasa 10X 1 µl, agua 6 µl y T4 ligasa 0,2 µl [400 U]). La mezcla de plásmido/dietiqueta se incubó durante 2 horas a 16 °C y se utilizó 1 µl de la mezcla para transformar células DH10B químicamente competentes 100 µl. Los transformantes resistentes a amp se exploraron por PCR utilizando, como cebadores de la PCR, pUC/m13 directo/inverso de 17 meros. Después, se secuenciaron los transformantes que contenían las dietiquetas concatenerizadas.

Ejemplo 2

Análisis SARD de una población definida

Se realizó SARD esencialmente como se describe en el Ejemplo 1. En este ejemplo, se mezclaron muestras de ADN genómico bacteriano disponibles en el comercio, a la misma concentración (peso/volumen). Las muestras de ADN bacteriano utilizadas incluyeron *Bacillus subtilis*, *Clostridium perfringens*, *Escherchia coli*, *Lactococcus lactis* y *Streptomyces coelicolor*. Se mezclaron los mismos volúmenes de las muestras de ADN (ADN genómico total) a una concentración de 50 nD/µl, cada uno. El Cuadro I muestra el tamaño de los genomas de cada especie bacteriana, el número de copias de ADNr 16S por genoma y el porcentaje molar de copias de 16S por cada especie bacteriana en la muestra de ADN total.

Cuadro I

Bacteria	Genoma (Mb)	Copias de 16S/Genoma	% Molar
<i>B. subtilis</i>	4,2	10	17,2
<i>C. perfringens</i>	4,4	10	16,4
<i>E. coli</i>	4,6	7	11,0
<i>L. lactis</i>	2,4	10	30,7
<i>S. coelicolor</i>	8,0	6	5,4

Después del SARD, se secuenciaron 120 etiquetas. El Cuadro II muestra el número y el porcentaje de etiquetas esperado en comparación con el número y el porcentaje de etiquetas observado para la población. Véase también la Fig. 14.

Cuadro II

Bacteria	Número de etiquetas esperado	Número de etiquetas observado	Porcentaje de etiquetas esperado	Porcentaje de etiquetas observado
<i>B. subtilis</i>	21	44	17,2 %	36,7 %
<i>C. perfringens</i>	20	18	16,4 %	15,0 %
<i>E. coli</i>	13	8	11,0 %	6,7 %
<i>L. lactis</i>	37	35	30,7 %	29,2 %
<i>S. coelicolor</i>	6	6	5,4 %	5,0 %

Cada uno de los genes de ADNr de estas especies produjo una etiqueta SARD que era diferenciable de los otros miembros del conjunto. Como puede observarse, se encontraron aproximadamente el doble de etiquetas que correspondían al ADNr de 16S de *B. subtilis* en comparación con lo esperado, basándose en el porcentaje molar de ADNr 16S de *B. subtilis*. La observación de que *B. subtilis* parecía ser dos veces más abundante que lo esperado ya se había descrito anteriormente. Farrelly *et al.* (Farrelly, V. F. *et al.*, 1995, Effect of genome size and *rrn* gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. *Appl. Environ. Microbiol.* 61(7): 2798-2801) describieron resultados similares cuando se amplificaba por PCR el gen de ADNr 16S de poblaciones

mixtas de ADN genómico. Llegaron a la conclusión de que este fenómeno era el resultado de la organización en tándem de los operones *rrn* en el genoma de *B. subtilis* donde genes de ADNr múltiples pueden amplificarse como un solo producto. Las etiquetas restantes se descubrieron a abundancias próximas a los valores esperados (desviación < 40 %).

5

Ejemplo 3

Análisis SARD de diversidad bacteriana ambiental

10 Para demostrar que el método SARD podía utilizarse para estudiar la diversidad bacteriana ambiental, se extrajo ADN total de dos muestras de suelo (Wy-1 y Wy-2) tomadas del Centro Rocky Mountain Oilfield Testing (RMOTC, Casper, Wyoming) en octubre del 2000. Las muestras se recogieron a una distancia de aproximadamente 8,05 km (0,5 millas) y a una profundidad de 0,36-0,46 m (14-18 pulgadas). Las muestras ambientales de ADN se sometieron a análisis SARD como se describe en el Ejemplo 1. En un análisis preliminar, se identificaron 148 etiquetas de la muestra Wy-1 y 234 etiquetas de la muestra Wy-2 (Figuras 15 y 16, respectivamente).

15

En la muestra Wy-1, se identificaron 58 etiquetas distintas y la abundancia de cada etiqueta variaba. La etiqueta más abundante (ATGGCTGTCGTCAGCT) constituía aproximadamente el 34 % de la población. Esta secuencia de etiqueta era idéntica a la de muchas secuencias bacterianas del GenBank y su posición dentro del gen de ADNr 16S indica que se localiza en una región conservada localizada distal al sitio de restricción *AluI* dirigido. En otras palabras, la contribución del gen, o de los genes, de 16S en esta etiqueta no contenía el sitio *AluI* conservado. Dado que la posición de etiqueta SARD la dictamina el primer sitio *AluI* distal al cebador marcado con biotina utilizado en la reacción de PCR inicial, es probable que el primer sitio *AluI* en el gen o genes 16S contribuyentes se localice aguas abajo dentro de una región conservada. Para disminuir el número de etiquetas que no contienen el sitio *AluI* conservado cerca de la región polimórfica, se puede purificar en gel los productos de la PCR de aproximadamente 100 pares de bases después de la primera etapa de restricción con *AluI*. Sin embargo, esto puede dar como resultado la pérdida de alguna información. No obstante, el 39 % de las etiquetas (58/148) en este conjunto era diferente entre sí. Véanse las Figuras 15 y 17.

20

25

30 Se descubrió que la muestra Wy-2 contenía 79 etiquetas diferentes de un total de 234 que se examinaron. Por tanto, el 34 % de las etiquetas (79/234) en este conjunto eran diferentes entre sí. Véanse las Figuras 16 y 17. Como ocurría en el caso de la muestra Wy-1, la etiqueta ATGGCTGTCGTCAGCT, que representa una secuencia conservada en un gen de ADN 16S, era más abundante y constituía aproximadamente el 30 % de la población.

35

40

45

Ejemplo 4

Análisis en serie de etiquetas polimórficas de ADNr del dominio *Arquea*

50 El método descrito en el Ejemplo 1 también puede aplicarse para el dominio *Arquea*. El dominio *Arquea* está constituido por dos reinos conocidos, Euriarqueota y Crenarqueota (Pace, N. R. 1997, A molecular view of microbial diversity and the biosphere. Science 276: 734-740). Para estudiar estos dos dominios puede utilizarse un conjunto de oligonucleótidos y de enzimas de restricción.

55

En este ejemplo, se diseñaron los siguientes oligonucleótidos: 5' biotina-TA(CT)T(CT)CCCA(GA)GCGG(CT)(GCT)(GC)(GA)CTT(AGCT)-3' que corresponde a la posición 817-838 del gen de ADNr 16S de *Methanococcus jannaschii* (número de registro del GenBank M59126) y (5'-GGTG(TGC)CA(GC)C(CA)GCCGCGGTAA(TC)ACC(AGCT)-3' que corresponde a la posición 457-481 del gen de ADNr 16S de *Methanococcus jannaschii*.

60

En este ejemplo, se utilizó un sitio *Bfal* (CTAG) que corresponde a la posición 768-771 del gen de ADNr 16S de *Methanococcus jannaschii*. Este sitio flanquea inmediatamente una región polimórfica.

65 El método SARD se ensayó por ordenador utilizando 17 representantes del dominio *Arquea* (Figura 4). Quince de las etiquetas SARD que pudieron identificarse eran todas exclusivas para este conjunto (Tabla II). Dos especies del género *Methanobacterium* no poseían ningún sitio *Bjal* en la región que podría amplificarse y por lo tanto no

produciría ninguna etiqueta.

Ejemplo 5

5 Estudio de regiones de ADNr polimórficas amplificadas por PCR

Los cebadores oligonucleotídicos que son complementarios a regiones conservadas que flanquean inmediatamente una región polimórfica pueden utilizarse para amplificar las secuencias de ADN polimórficas. Las secuencias de cebadores oligonucleotídicos pueden incluir sitios de restricción existentes o introducidos que permiten la clonación subsecuencia en un vector bacteriano. Utilizando o introduciendo diferentes sitios de restricción en cada cebador, los productos de la PCR digeridos por enzimas de restricción pueden concatenerizarse de una manera unidireccional antes de la clonación. Esta etapa permitiría realizar un análisis de secuencias en serie de múltiples regiones polimórficas a partir de un solo producto recombinante.

15 Ejemplo 6

Estudio de productos de traducción de regiones polimórficas de ADNr 16S

Las regiones polimórficas de ADNr 16S también pueden investigarse identificando en paralelo los productos de traducción de una región polimórfica determinada. Podrían diseñarse oligonucleótidos de tal manera que pudieran servir para amplificar una región polimórfica. El cebador en 5' también podría incluir un sitio de unión a la polimerasa de T7 u otro sitio de unión a la polimerasa, una secuencia consenso Kozak, un codón ATG iniciador y un epítipo para facilitar la purificación. Como ejemplos de epítipos se incluyen hemaglutinina (HA), *myc*, Flag o polihistidina. Después de la amplificación, los productos se someten a transcripción/traducción *in vitro* para producir los productos peptídicos. Estos péptidos se purifican del extracto celular y se analizan por espectrometría de masas. Este tipo de una estrategia se ha aplicado para la identificación de mutaciones en el gen de BRCA1 (Garvin, A M. *et al.*, 2000 MALDI-TOF based mutation detection using tagged in vitro synthesized peptides. Nature Biotechnol. 18: 95-97).

Aunque este proceso no proporciona rápidamente información de secuencia de ADN a partir de la cual deducir la taxonomía, permitiría la creación de perfiles de diversidad microbiana constituidos por 'etiquetas de masa'. Estas etiquetas de masa podrían utilizarse para identificar correlaciones entre etiquetas específicas y diversos parámetros de muestra. Para ensayar el contenido de la información de esta estrategia, se tradujo por ordenador una región polimórfica de los genes de ADNr 16S de las especies de la Figura 3. De las 34 regiones polimórficas examinadas, 32 produjeron una etiqueta con una masa única en este conjunto (Tabla III).

Otra estrategia para traducir las regiones polimórficas amplificadas podría ser clonar y expresar las secuencias en células completas. Por ejemplo, podrían diseñarse oligonucleótidos para amplificar regiones polimórficas que incluyesen secuencias en los extremos 5' que permitiesen la clonación por recombinación homóloga en levaduras. Estas secuencias podrían clonarse en un casete de expresión para crear una fusión entre una proteína secretada, tal como factor alfa o invertasa, y un epítipo para facilitar la purificación y la secuencia de aminoácidos traducida o de la región polimórfica de ADNr. La recombinación homóloga en levaduras es bastante sólida y puede permitir fácilmente el aislamiento de 10^3 - 10^4 recombinantes independientes en una sola placa de transformación. Los productos secretados pueden aislarse del medio e identificarse por espectrometría de masas.

45 Ejemplo 7

Hibridación de ADNr microbiano con oligonucleótidos inmovilizados

Una complicación con el uso de sondas oligonucleotídicas cortas en micromatrices de ADN es la inestabilidad de los dúplex de oligonucleótidos cortos. Una posible solución a este problema es sintetizar sondas que incluyan secuencias degeneradas para alojar secuencias desconocidas. La longitud de oligonucleótidos la dictamina el número de degeneraciones, o de permutaciones por secuencia, que han de incluirse. Por ejemplo, un oligonucleótido nonamérico completamente degenerado requiere 4^9 o 262.144 secuencias oligonucleotídicas diferentes. La hibridación eficiente de oligonucleótidos nonaméricos no es posible utilizando condiciones estándar. Una solución ha sido incorporar una etapa de ligamiento en la hibridación de una secuencia diana con una sonda oligonucleotídica nonamérica (Gunderson, K. L. *et al.*, 1998, Mutation detection by ligation to complete n-mer DNA arrays. Genome Res. 8: 1142-1153).

Otra solución sería aumentar de un modo eficaz la longitud del oligonucleótido incluyendo en el cebador nucleótidos que no sean degenerados. Esta estrategia podría aplicarse a un estudio de una comunidad microbiana construyendo sondas oligonucleotídicas que están compuestas por una región constante que corresponde a una región bien conservada de un gen de ADNr 16S junto con una secuencia degenerada que corresponde a la región polimórfica flanqueante.

Un ejemplo de dicha colección de oligonucleótidos degenerados para el dominio *Bacteria* podría incluir permutaciones del siguiente cebador: 5'-ACGAGCGCAACCCNNNNNNNN-3', donde N indica cualquier nucleótido

ES 2 632 538 T3

en esa posición. Esta secuencia corresponde a la posición 1101-1122 del gen de ADNr 16S de *E. coli* (número de registro del GenBank E05133). Como alternativa, podrían diseñarse cebadores de tal manera que estuviesen compuestos por una mezcla de posiciones semidegeneradas (por ejemplo, A o G) y posiciones degeneradas (por ejemplo, A, G, C o T), de secuencia constante.

5

Tabla 1*

Especie	N.º de registro del GenBank	Secuencia de etiqueta	Posición
<i>Desulfurobacterium thermolithotrophum</i>	AJ001049	GTCAGTTGCCGAAGCT	814-829
Aquificales OPS132 no cultivada	AF027104	GTCCGTGCCGTAAGCT	810-825
<i>Bacteroides caccae</i>	XB3951	TTTCCGGCCCGTAGCT	1021-1036
<i>Actinomyces bovis</i>	X81061	TTTCCGGCCCGTAGCT	834-849
<i>Actinomyces meyeri</i>	X82451	TTTCTGCCCGTAGCT	828-843
<i>Denitrobacterium detoxificans</i>	AF079507	CCTCCGGCCCGCAGCT	788-803
Bacteria BPC110 GNS no cultivada	AF154084	CCCGGTAGTCCTAGCT	765-780
Bacteria GCA004 GNS no cultivada	AF154104	CATCGGTGCCGAGCT	824-839
Bacteria GCA112 GNS no cultivada	AF154100	CGCGGTGCCGTAAGCT	826-841
<i>Acetobacter acetii</i>	AF127399	ACTCAGTGTCTAGCT	782-797
<i>Glinconobacter asati</i>	AB024492	ACTCAGTGTCTAGCT	783-798
<i>Burkholderia</i> sp. 1B1	X92188	CCTTAGTAACGAAAGCT	837-852
<i>Denitrobacter permamnis</i>	Y12639	TTTCCGGCCCGTAGCT	789-804
<i>Desulfobacter curvatus</i>	M34413	CTGCTGTGCCNAAGCT	861-876
<i>Desulfobulbids</i> sp. BG25	U85473	CCTCTGTGTCCGAGCT	854-869
<i>Legionella anisa</i>	X73394	TTTCCGGCCCGTAGCT	790-805
Clon SB-1 mineralizante de benceno	AF029039	TTTCCGGCCCGTAGCT	1029-1044
<i>Escherichia coli</i>	B05133	CGXGGCTTCCGGAGCT	848-863
<i>Acidobacterium</i> Sub. Div-1 no cultivada	X68464	CCGCCGTGCCGAAGCT	813-828
<i>Acidobacterium</i> Sub. Div-1 no cultivada	Z73363	CGGCTGTGCCGAAGCT	521-536
<i>Acidobacterium</i> Sub. Div-1 no cultivada	Z73365	CCACTGTGCCGTAAGCT	521-536
<i>Acidobacterium</i> Sub. Div-1 no cultivada	Z73368	CTGCTGTGCCGAGCT	521-536
<i>Acidobacterium</i> Sub. Div-1 no cultivada	Z73364	CTGCCGXGCCGGAGCT	521-536
<i>Acidobacterium</i> Sub. Div-1 no cultivada	U68659	CCAATGTGCCGGAGCT	319-334
<i>Acidobacterium</i> Sub. Div-1 no cultivada	D2617X	CCGTCGTGCCGTAAGCT	779-794
<i>Acidobacterium</i> Sub. Div-1 no cultivada	X97101	CCGTCGTGCCGTAAGCT	687-702
<i>Acidobacterium</i> Sub. Div-1 no cultivada	X97098	CTGCCGTGTCTAGCT	798-813
<i>Acidobacterium</i> Sub. Div-1 no cultivada	AF04764S	CTCCCGTGTCTAGCT	779-794
<i>Acidobacterium</i> Sub. Div-2 no cultivada	AF05054B	CCGCCGTGCCGGAGCT	316-331
<i>Acidobacterium</i> Sub. Div-2 no cultivada	U68612	CTGAGGAACGAAAGCT	226-241
<i>Acidobacterium</i> Sub. Div-3 no cultivada	Y07646	GTGTCGTGCCGGAGCT	830-845
<i>Acidobacterium</i> Sub. Div-3 no cultivada	X97097	GGGCTGTGCCGAAGCT	804 819
<i>Acidobacterium</i> Sub. Div-3 no cultivada	X68466	GGTCGGTGCCGGAGCT	796-811
<i>Acidobacterium</i> Sub. Div-3 no cultivada	X68468	GGTCGGTGCCAGAGCT	796-811
<i>Acidobacterium</i> Sub. Div-1 no cultivada	U68648	GGTTCGTGCCGGAGCT	317-332
<i>Acidobacterium</i> Sub. Div-1 no cultivada	X684G7	TGCTGTGCCGGAGCT	796-811
<i>Acidobacterium</i> Sub. Div-1 no cultivada	AFQ13515	TATCCGTGCCGGAGCT	799-814
<i>Acidobacterium</i> Sub. Div-1 no cultivada	AF027004	GGTCGGTGCCGGAGCT	778-793

* Las secuencias mostradas en negrita con sombra indican que no son únicas para este conjunto.

Tabla II

Especie	N.º de registro del GenBank	Secuencia de etiqueta	Posición
Crenarqueota			
<i>Aeropyrum pernix</i>	D83259	CTAGGGGGCGGGAG	614-627
<i>Desulfurococcus mobiles</i>	M36474	CTAGGTGTTGGGTG	856-869
<i>Staphylothermus marinus</i>	X99560	CTAGGTGTTGGGCG	770-783
<i>Metallosphaera sedula</i>	X90481	CTAGGTGTCGCGTA	756-769
<i>Sulfolobus acidocaldarius</i>	D14053	CTAGGTGTCGAGTA	785-798
<i>Sulfolobus metallicus</i>	D85519	CTAGGTGTCACGTG	744-757
<i>Caldivirga maquilingensis</i>	AB013926	CTAGCTGTTGGGTG	773-786
<i>Pyrobaculum islandicum</i>	L07511	CTAGCTGTCGGCCG	781-794
Euriarqueota			
<i>Archaeoglobus fulgidus</i>	X05567	CTAGGTGTCACCGA	780-793
<i>Archaeoglobus veneficus</i>	Y10011	CTAGGTGTCACCGG	758-771
<i>Haloarcula japonica</i>	D28872	CTAGGTGTGGCGTA	762-775
<i>Halococcus morrhuae</i>	D11106	CTAGGTGTGGCGTT	765-778
<i>Methanococcus jannaschii</i>	M59126	CTAGGTGTCGCGTC	768-781
<i>Methanobacterium bryantii</i>	AF028688	Ninguna	
<i>Methanobacterium subterraneum</i>	X99045	Ninguna	
<i>Pyrococcus abyssi</i>	Z70246	CTAGGTGTCGGGCG	767-780
<i>Picrophilus oshimae</i>	X84901	CTAGCTGTAAACTC	742-755

Tabla III

Especie	N.º de registro del GenBank	Secuencia peptídica	P.M.	Posición
<i>Desulfurobacterium thermolithotrophum</i>	AJ001049	RAQPLSLVASG*	1097,40	1079-1136
Aquificales OPS132 no cultivada	AF027104	RAQPLSCVTSG*	1117,40	1074-1131
<i>Bacteroides caccae</i>	X83951	RAQPLSSVTNRSC*	1417,70	1069-1126
<i>Actinomyces bovis</i>	X81061	RAQPLSRVASTLWWGLAGD	2083,60	1088-1145
<i>Actinomyces meyeri</i>	X82451	RAQPLPYVASTLWWGLVGD	2128,60	1082-1139
<i>Denitrobacterium detoxifcans</i>	AF079507	RAQPLPHVASIRLGTHGG	1866,50	1039-1094
Bacteria BPCI10 GNS no cultivada	AF154084	RAQPLLYVIRVIPD	1652,10	1074-1116
Bacteria GCA004 GNS no cultivada	AF154104	RAQPSLYVTRIIRD	1687,10	1080-1122
Bacteria GCA112 GNS no cultivada	AF154100	RAQPSPYVIRVIRD	1669,00	1082-1124
<i>Acetobacter aceti</i>	AF127399	RAQPLSLVASMFGWAL*	1746,30	1038-1095
<i>Glnconobacter asati</i>	AB024492	RAQPLSLVASTFRWAL*	1815,30	1034-1092
<i>Burkholderia</i> sp. 1B1	X92188	RAQPLSLVATQEHSRET	1922,20	1094-1144
<i>Denitrobacter permamns</i>	Y12639	RAQPLPLVATFSWAL*	1669,10	1077-1131
<i>Desulfobacter curvatus</i>	M34413	RAQPLSLVASTLCGNSNET	1960,40	1116-1172
<i>Desulfobulbites</i> sp. BG25	U85473	RAQPLPLVASSAGHSGKT	1863,40	1114-1170
<i>Legionella anisa</i>	X73394	RAQPLSLVAST*	1141,40	1078-1135
Clon SB-1 mineralizante de benceno	AF029039	RAQPLPLVANRSSWGL*	1764,20	1077-1134
<i>Escherichia coli</i>	E05133	RAQPLSFVASGPAGNSKET	1916,30	1103-1159
Acidobacterium Sub. Div-1 no cultivada	Z73363	RAQPLSLVASGSSRAL*	1612,10	775-832
Acidobacterium Sub. Div-1 no cultivada	Z73365	RAQPLSSVAIGSSRATLAK	1912,50	777-835
Acidobacterium Sub. Div-1 no cultivada	Z73368	RAQPLFASCHH*	1933,50	779-835
Acidobacterium Sub. Div-1 no cultivada	Z73364	RAQPLFAQLPSFSWALCRN	2204,80	778-835
Acidobacterium Sub. Div-1 no cultivada	U68659	RAQPLLPXAI*	1218,70	573-630
Acidobacterium Sub. Div-1 no cultivada	D26171	RAQPLLPVATI*	1177,50	1035-1090
Acidobacterium Sub. Div-1 no cultivada	X97101	RAQPLSPVAII*	1163,50	943-998
Acidobacterium Sub. Div-1 no cultivada	X97098	RAQPLSSVATI*	1141,40	1054-1109
Acidobacterium Sub. Div-1 no cultivada	AP047646	RAQPLFLVATI*	1227,60	1035-1090
Acidobacterium Sub. Div-1 no cultivada	AF050548	RAQPSSLVANTLW*	1441,70	572-629
Acidobacterium Sub. Div-2 no cultivada	U68612	RAQPLUVVATRKRELYVD	2150,60	577-630
Acidobacterium Sub. Div-2 no cultivada	Y07646	RAQPLHWATPQGGTLRG	1857,30	1085-1140
Acidobacterium Sub. Div-3 no cultivada	X97097	RAQPSSLVANPQGKHPKGT	1972,40	1060-1116
Acidobacterium Sub. Div-3 no cultivada	U68648	RARPLSCVAII*	1197,70	574-629
Acidobacterium Sub. Div-3 no cultivada	AF013515	RAQPLSCVANPQGCTLRR	1969,50	1057-1112
Acidobacterium Sub. Div-3 no cultivada	AF027004	RAQPSPCVATPPRAGALSGD	1950,40	1036-1096

* Indica que en la secuencia polimórfica se encontró un codón de terminación en marco

LISTADO DE SECUENCIAS

<110> Matthew, Ashby N.

5 <120> Métodos para el estudio y análisis genético de poblaciones

<130> ASHBY-1

10 <140> PCT/US01/11609
<141> 10-04-2010

<150> US 60/196063
<151> 10-04-2010

15 <150> US 60/196258
<151> 11-04-2010

<160> 244

20 <170> PatentIn versión 3.1

<210> 1
<211> 16
<212> ADN
25 <213> desconocido

<220>
<223> organismo del suelo no identificado

30 <400> 1
acgatgagca ctagct 16

<210> 2
<211> 16
35 <212> ADN
<213> desconocido

<220>
<223> organismo del suelo no identificado

40 <400> 2
acgatgagta ctagct 16

<210> 3
<211> 16
45 <212> ADN
<213> desconocido

<220>
50 <223> organismo del suelo no identificado

<400> 3
acgatgatga ctagct 16

55 <210> 4
<211> 16
<212> ADN
<213> desconocido

60 <220>
<223> organismo del suelo no identificado

<400> 4
65 acgatggatg ctagct 16

<210> 5

<211> 16
 <212> ADN
 <213> desconocido
 5 <220>
 <223> organismo del suelo no identificado
 <400> 5
 atgctagtct ggagct 16
 10 <210> 6
 <211> 16
 <212> ADN
 <213> desconocido
 15 <220>
 <223> organismo del suelo no identificado
 <400> 6
 atggctgtcg tcagct 16
 20 <210> 7
 <211> 16
 <212> ADN
 <213> desconocido
 25 <220>
 <223> organismo del suelo no identificado
 30 <400> 7
 atggttgcg tcagct 16
 <210> 8
 <211> 16
 <212> ADN
 <213> desconocido
 35 <220>
 <223> organismo del suelo no identificado
 40 <400> 8
 attccgtgcc gtagct 16
 <210> 9
 <211> 16
 <212> ADN
 <213> desconocido
 45 <220>
 <223> organismo del suelo no identificado
 50 <400> 9
 cactagtggc gcagct 16
 <210> 10
 <211> 16
 <212> ADN
 <213> desconocido
 55 <220>
 <223> organismo del suelo no identificado
 60 <400> 10
 cccccgtgcc gaagct 16
 65 <210> 11

ES 2 632 538 T3

<211> 16
<212> ADN
<213> desconocido

5 <220>
<223> organismo del suelo no identificado

<400> 11
cccccggtgcc gcagct 16

10 <210> 12
<211> 16
<212> ADN
<213> desconocido

15 <220>
<223> organismo del suelo no identificado

<400> 12
cccccttcct ccagct 16

20 <210> 13
<211> 16
<212> ADN
<213> desconocido

25 <220>
<223> organismo del suelo no identificado

30 <400> 13
ccccggtgcc gcagct 16

<210> 14
<211> 16
<212> ADN
<213> desconocido

35 <220>
<223> organismo del suelo no identificado

40 <400> 14
ccgggtagtc ccagct 16

<210> 15
<211> 16
<212> ADN
<213> desconocido

45 <220>
<223> organismo del suelo no identificado

50 <400> 15
cctccgtgcc gaagct 16

<210> 16
<211> 16
<212> ADN
<213> desconocido

55 <220>
<223> organismo del suelo no identificado

60 <400> 16
cctccgtgcc gcagct 16

65 <210> 17

<211> 16
 <212> ADN
 <213> desconocido

5

<220>
 <223> organismo del suelo no identificado

<400> 17
 cctccgtgct gcagct 16

10

<210> 18
 <211> 16
 <212> ADN
 <213> desconocido

15

<220>
 <223> organismo del suelo no identificado

<400> 18
 cctcggcgcc gcagct 16

20

<210> 19
 <211> 16
 <212> ADN
 <213> desconocido

25

<220>
 <223> organismo del suelo no identificado

30

<400> 19
 cctcggtgcc gcagct 16

<210> 20
 <211> 16
 <212> ADN
 <213> desconocido

35

<220>
 <223> organismo del suelo no identificado

40

<400> 20
 cctcgggtgc gcagct 16

45

<210> 21
 <211> 16
 <212> ADN
 <213> desconocido

50

<220>
 <223> organismo del suelo no identificado

<400> 21
 cctgggtgcc gcagct 16

55

<210> 22
 <211> 16
 <212> ADN
 <213> desconocido

60

<220>
 <223> organismo del suelo no identificado

<400> 22
 cctgtgtgac gaagct 16

65

<210> 23

<211> 16
 <212> ADN
 <213> desconocido

5

<220>
 <223> organismo del suelo no identificado

<400> 23
 ccttgtaac gaagct 16

10

<210> 24
 <211> 16
 <212> ADN
 <213> desconocido

15

<220>
 <223> organismo del suelo no identificado

<400> 24
 ccttggtacc gaagct 16

20

<210> 25
 <211> 16
 <212> ADN
 <213> desconocido

25

<220>
 <223> organismo del suelo no identificado

30

<400> 25
 cgccagtgcc gtagct 16

<210> 26
 <211> 16
 <212> ADN
 <213> desconocido

35

<220>
 <223> organismo del suelo no identificado

40

<400> 26
 cgcctgtgcc gtagct 16

45

<210> 27
 <211> 16
 <212> ADN
 <213> Desconocido

50

<220>
 <223> organismo del suelo no identificado

<400> 27
 cgtccgtgcc gaagct 16

55

<210> 28
 <211> 16
 <212> ADN
 <213> desconocido

60

<220>
 <223> organismo del suelo no identificado

<400> 28
 cgtccgtgcc gcagct 16

65

<210> 29

<211> 16
 <212> ADN
 <213> desconocido
 5 <220>
 <223> organismo del suelo no identificado
 10 <400> 29
 cgtcgggtgcc gcagct 16
 <210> 30
 <211> 16
 <212> ADN
 15 <213> desconocido
 <220>
 <223> organismo del suelo no identificado
 20 <400> 30
 ctcccgtgcc gcagct 16
 <210> 31
 <211> 16
 25 <212> ADN
 <213> desconocido
 <220>
 <223> organismo del suelo no identificado
 30 <400> 31
 ctcccgtgcc ggagct 16
 <210> 32
 <211> 16
 35 <212> ADN
 <213> desconocido
 <220>
 <223> organismo del suelo no identificado
 40 <400> 32
 ctccgggtgcc gcagct 16
 <210> 33
 <211> 16
 <212> ADN
 <213> desconocido
 45 <220>
 <223> organismo del suelo no identificado
 50 <400> 33
 ctctgtgcc gaagct 16
 <210> 34
 <211> 16
 <212> ADN
 <213> desconocido
 55 <220>
 <223> organismo del suelo no identificado
 60 <400> 34
 ctctgtgcc gcagct 16

<210> 35
 <211> 16
 <212> ADN
 <213> desconocido
 5
 <220>
 <223> organismo del suelo no identificado
 <400> 35
 ctgccgtgcc gaagct 16
 10
 <210> 36
 <211> 16
 <212> ADN
 <213> desconocido
 15
 <220>
 <223> organismo del suelo no identificado
 20
 <400> 36
 ctgctgtgcc gaagct 16
 <210> 37
 <211> 16
 <212> ADN
 <213> desconocido
 25
 <220>
 <223> organismo del suelo no identificado
 30
 <400> 37
 ctgtcgtgcc gaagct 16
 <210> 38
 <211> 16
 <212> ADN
 <213> desconocido
 35
 <220>
 <223> organismo del suelo no identificado
 40
 <400> 38
 cttcagtatc gaagct 16
 <210> 39
 <211> 16
 <212> ADN
 <213> desconocido
 45
 <220>
 <223> organismo del suelo no identificado
 50
 <400> 39
 cttccgcgcc ggagct 16
 55
 <210> 40
 <211> 16
 <212> ADN
 <213> desconocido
 60
 <220>
 <223> organismo del suelo no identificado
 <400> 40
 cttccgtgcc gcagct 16
 65

<210> 41
 <211> 16
 <212> ADN
 <213> desconocido
 5
 <220>
 <223> organismo del suelo no identificado
 <400> 41
 10 cttccgtgcc ggagct 16
 <210> 42
 <211> 16
 <212> ADN
 15 <213> desconocido
 <220>
 <223> organismo del suelo no identificado
 20 <400> 42
 cttcgtgcc gcagct 16
 <210> 43
 <211> 16
 25 <212> ADN
 <213> desconocido
 <220>
 <223> organismo del suelo no identificado
 30 <400> 43
 cttctgtgcc gaagct 16
 <210> 44
 <211> 16
 35 <212> ADN
 <213> desconocido
 <220>
 <223> organismo del suelo no identificado
 40 <400> 44
 gatccgtgcc gtagct 16
 45 <210> 45
 <211> 16
 <212> ADN
 <213> desconocido
 50 <220>
 <223> organismo del suelo no identificado
 <400> 45
 55 gctctgtgcc gaagct 16
 <210> 46
 <211> 16
 <212> ADN
 <213> desconocido
 60 <220>
 <223> organismo del suelo no identificado
 <400> 46
 65 gctgggtgcc caagct 16

ES 2 632 538 T3

5
<210> 47
<211> 16
<212> ADN
<213> desconocido

<220>
<223> organismo del suelo no identificado

10
<400> 47
ggtccgtgcc gcagct 16

<210> 48
<211> 16
<212> ADN
15
<213> desconocido

<220>
<223> organismo del suelo no identificado

20
<400> 48
ggtgctcttc ggagct 16

<210> 49
<211> 16
25
<212> ADN
<213> desconocido

<220>
<223> organismo del suelo no identificado
30
<400> 49
gtaaacgatg gaagct 16

<210> 50
<211> 16
35
<212> ADN
<213> desconocido

<220>
<223> organismo del suelo no identificado
40
<400> 50
gtggctgtcg tcagct 16

45
<210> 51
<211> 16
<212> ADN
<213> desconocido

50
<220>
<223> organismo del suelo no identificado

<400> 51
55
gttccgtgcc gaagct 16

<210> 52
<211> 16
<212> ADN
60
<213> desconocido

<220>
<223> organismo del suelo no identificado

65
<400> 52
gttccgtgcc gcagct 16

<210> 53
 <211> 16
 <212> ADN
 <213> desconocido
 5
 <220>
 <223> organismo del suelo no identificado
 <400> 53
 10 tatcagtggc gcagct 16
 <210> 54
 <211> 16
 <212> ADN
 15 <213> desconocido
 <220>
 <223> organismo del suelo no identificado
 20 <400> 54
 tctccgtgcc gcagct 16
 <210> 55
 <211> 16
 25 <212> ADN
 <213> desconocido
 <220>
 <223> organismo del suelo no identificado
 30 <400> 55
 tctctgtgcc gcagct 16
 <210> 56
 35 <211> 16
 <212> ADN
 <213> desconocido
 40 <220>
 <223> organismo del suelo no identificado
 <400> 56
 45 tctctgtgcc gtagct 16
 <210> 57
 <211> 16
 <212> ADN
 50 <213> desconocido
 <220>
 <223> organismo del suelo no identificado
 <400> 57
 55 tgtccgtgcc gtagct 16
 <210> 58
 <211> 16
 <212> ADN
 60 <213> desconocido
 <220>
 <223> organismo del suelo no identificado
 65 <400> 58
 ttccgtgcc gcagct 16

5 <210> 59
 <211> 16
 <212> ADN
 <213> desconocido

<220>
 <223> organismo del suelo no identificado

10 <400> 59
 acgatgataa ctagct 16

<210> 60
 <211> 16
 <212> ADN
 <213> desconocido

15 <220>
 <223> organismo del suelo no identificado

20 <400> 60
 acgatgatga caagct 16

<210> 61
 <211> 16
 <212> ADN
 <213> desconocido

25 <220>
 <223> organismo del suelo no identificado

30 <400> 61
 acgatgggca ctagct 16

35 <210> 62
 <211> 16
 <212> ADN
 <213> desconocido

40 <220>
 <223> organismo del suelo no identificado

45 <400> 62
 acggctgtcg tcagct 16

<210> 63
 <211> 16
 <212> ADN
 <213> desconocido

50 <220>
 <223> organismo del suelo no identificado

55 <400> 63
 actacgagcg caagct 16

<210> 64
 <211> 16
 <212> ADN
 <213> desconocido

60 <220>
 <223> organismo del suelo no identificado

65 <400> 64
 acttaatgcg ttagct 16

5 <210> 65
 <211> 16
 <212> ADN
 <213> desconocido

<220>
 <223> organismo del suelo no identificado

10 <400> 65
 atgctagtct ggagct 16

<210> 66
 <211> 16
 <212> ADN
 <213> desconocido

15 <220>
 <223> organismo del suelo no identificado

20 <400> 66
 atggctctcg tcagct 16

25 <210> 67
 <211> 16
 <212> ADN
 <213> desconocido

30 <220>
 <223> organismo del suelo no identificado

35 <400> 67
 atggctgtcg ccagct 16

<210> 68
 <211> 16
 <212> ADN
 <213> desconocido

40 <220>
 <223> organismo del suelo no identificado

45 <400> 68
 atggctgtcg tcagct 16

<210> 69
 <211> 16
 <212> ADN
 <213> desconocido

50 <220>
 <223> organismo del suelo no identificado

55 <400> 69
 atggtgtcg tcagct 16

<210> 70
 <211> 16
 <212> ADN
 <213> desconocido

60 <220>
 <223> organismo del suelo no identificado

65 <400> 70

ES 2 632 538 T3

atgtagacct ggagct 16

5 <210> 71
<211> 16
<212> ADN
<213> desconocido

10 <220>
<223> organismo del suelo no identificado

15 <400> 71
attccgtgcc gcagct 16

20 <210> 72
<211> 16
<212> ADN
<213> desconocido

25 <220>
<223> organismo del suelo no identificado

30 <400> 72
attccgtgcc gtagct 16

35 <210> 73
<211> 16
<212> ADN
<213> desconocido

40 <220>
<223> organismo del suelo no identificado

45 <400> 73
cacaagcggg ggagct 16

50 <210> 74
<211> 16
<212> ADN
<213> desconocido

55 <220>
<223> organismo del suelo no identificado

60 <400> 74
cactagtggc gcagct 16

65 <210> 75
<211> 16
<212> ADN
<213> desconocido

<220>
<223> organismo del suelo no identificado

<400> 75
catccgtgcc gaagct 16

<210> 76
<211> 16
<212> ADN
<213> desconocido

<220>
<223> organismo del suelo no identificado

<400> 76

ES 2 632 538 T3

5 ccccagggcc caagct 16
 <210> 77
 <211> 16
 <212> ADN
 <213> Desconocido
 <220>
 10 <223> organismo del suelo no identificado
 <400> 77
 ccccggtgcc gcagct 16
 <210> 78
 15 <211> 16
 <212> ADN
 <213> desconocido
 <220>
 20 <223> organismo del suelo no identificado
 <400> 78
 cccgctgcc ggagct 16
 25 <210> 79
 <211> 16
 <212> ADN
 <213> desconocido
 30 <220>
 <223> organismo del suelo no identificado
 <400> 79
 35 ccgctgtgcc gtagct 16
 <210> 80
 <211> 16
 <212> ADN
 40 <213> desconocido
 <220>
 <223> organismo del suelo no identificado
 <400> 80
 45 ccggtagtc ccagct 16
 <210> 81
 <211> 16
 50 <212> ADN
 <213> desconocido
 <220>
 <223> organismo del suelo no identificado
 55 <400> 81
 ccggtagtc ctagct 16
 <210> 82
 60 <211> 16
 <212> ADN
 <213> desconocido
 <220>
 65 <223> organismo del suelo no identificado
 <400> 82

ES 2 632 538 T3

cctccgtgcc gaagct 16

5 <210> 83
<211> 16
<212> ADN
<213> desconocido

10 <220>
<223> organismo del suelo no identificado

<400> 83
cctccgtgcc gcagct 16

15 <210> 84
<211> 16
<212> ADN
<213> desconocido

20 <220>
<223> organismo del suelo no identificado

<400> 84
cctccgtgcc ggagct 16

25 <210> 85
<211> 16
<212> ADN
<213> desconocido

30 <220>
<223> organismo del suelo no identificado

<400> 85
cctcgtaagg ggagct 16

35 <210> 86
<211> 16
<212> ADN
<213> desconocido

40 <220>
<223> organismo del suelo no identificado

45 <400> 86
cctgggtgcc gcagct 16

50 <210> 87
<211> 16
<212> ADN
<213> desconocido

<220>
<223> organismo del suelo no identificado

55 <400> 87
cctgtagtc ccagct 16

60 <210> 88
<211> 16
<212> ADN
<213> desconocido

65 <220>
<223> organismo del suelo no identificado

<400> 88

ES 2 632 538 T3

cctgtagtc ctagct 16

5 <210> 89
<211> 16
<212> ADN
<213> desconocido

10 <220>
<223> organismo del suelo no identificado

<400> 89
ccttagtaac gcagct 16

15 <210> 90
<211> 16
<212> ADN
<213> desconocido

20 <220>
<223> organismo del suelo no identificado

<400> 90
ccttgtaac gaagct 16

25 <210> 91
<211> 16
<212> ADN
<213> desconocido

30 <220>
<223> organismo del suelo no identificado

<400> 91
cgccagtgcc gaagct 16

35 <210> 92
<211> 16
<212> ADN
<213> desconocido

40 <220>
<223> organismo del suelo no identificado

45 <400> 92
cgccgtgcc gcagct 16

<210> 93

50 <211> 16
<212> ADN
<213> desconocido

<220>
<223> organismo del suelo no identificado

55 <400> 93
cgccgtgcc gtagct 16

60 <210> 94
<211> 16
<212> ADN
<213> desconocido

65 <220>
<223> organismo del suelo no identificado

<400> 94
 cgctcgtggc gaagct 16

5 <210> 95
 <211> 16
 <212> ADN
 <213> desconocido

10 <220>
 <223> organismo del suelo no identificado

 <400> 95
 cggaggcgtc gtagct 16

15 <210> 96
 <211> 16
 <212> ADN
 <213> desconocido

20 <220>
 <223> organismo del suelo no identificado

 <400> 96
 cgtcagtgtc gcagct 16

25 <210> 97
 <211> 16
 <212> ADN
 <213> desconocido

30 <220>
 <223> organismo del suelo no identificado

 <400> 97
 cgtccgtgcc gaagct 16

35 <210> 98
 <211> 16
 <212> ADN
 <213> desconocido

40 <220>
 <223> organismo del suelo no identificado

45 <400> 98
 cgtccgtgcc gcagct 16

 <210> 99
 <211> 16
 <212> ADN
 <213> desconocido

50 <220>
 <223> organismo del suelo no identificado

55 <400> 99
 cgtccgtgcc ggagct 16

 <210> 100
 <211> 16
 <212> ADN
 <213> desconocido

60 <220>
 <223> organismo del suelo no identificado

65 <220>
 <223> organismo del suelo no identificado

ES 2 632 538 T3

<400> 100
cgtcggtgcc gcagct 16

5 <210> 101
<211> 16
<212> ADN
<213> desconocido

10 <220>
<223> organismo del suelo no identificado

<400> 101
ctccagtgcc gcagct 16

15 <210> 102
<211> 16
<212> ADN
<213> desconocido

20 <220>
<223> organismo del suelo no identificado

<400> 102
ctcccgtgcc acagct 16

25 <210> 103
<211> 16
<212> ADN
<213> desconocido

30 <220>
<223> organismo del suelo no identificado

35 <400> 103
ctcccgtgcc gcagct 16

40 <210> 104
<211> 16
<212> ADN
<213> desconocido

45 <220>
<223> organismo del suelo no identificado

<400> 104
ctcccgtgcc ggagct 16

50 <210> 105
<211> 16
<212> ADN
<213> desconocido

55 <220>
<223> organismo del suelo no identificado

<400> 105
ctccggtgcc gcagct 16

60 <210> 106
<211> 16
<212> ADN
<213> desconocido

65 <220>
<223> organismo del suelo no identificado

<400> 106
 ctctgtgcc gcagct 16
 5 <210> 107
 <211> 16
 <212> ADN
 <213> desconocido
 10 <220>
 <223> organismo del suelo no identificado
 <400> 107
 ctgccgcgcc ggagct 16
 15 <210> 108
 <211> 16
 <212> ADN
 <213> desconocido
 20 <220>
 <223> organismo del suelo no identificado
 <400> 108
 ctgccgtgcc gaagct 16
 25 <210> 109
 <211> 16
 <212> ADN
 <213> desconocido
 30 <220>
 <223> organismo del suelo no identificado
 <400> 109
 ctgccgtgcc taagct 16
 35 <210> 110
 <211> 16
 <212> ADN
 <213> desconocido
 40 <220>
 <223> organismo del suelo no identificado
 <400> 110
 ctgcggtgcc gcagct 16
 45 <210> 111
 <211> 16
 <212> ADN
 <213> desconocido
 50 <220>
 <223> organismo del suelo no identificado
 <400> 111
 ctgctgtgcc gaagct 16
 55 <210> 112
 <211> 16
 <212> ADN
 <213> desconocido
 60 <220>
 <223> organismo del suelo no identificado
 65

<400> 112
 ctgtcgtgcc gaagct 16
 5 <210> 113
 <211> 16
 <212> ADN
 <213> desconocido
 10 <220>
 <223> organismo del suelo no identificado
 <400> 113
 15 ctcccgcgcc ggagct 16
 <210> 114
 <211> 16
 <212> ADN
 20 <213> desconocido
 <220>
 <223> organismo del suelo no identificado
 <400> 114
 25 ctcccgtgcc gaagct 16
 <210> 115
 <211> 16
 <212> ADN
 30 <213> desconocido
 <220>
 <223> organismo del suelo no identificado
 35 <400> 115
 ctcccgtgcc gcagct 16
 <210> 116
 <211> 16
 40 <212> ADN
 <213> desconocido
 <220>
 <223> organismo del suelo no identificado
 45 <400> 116
 ctcccgtgcc ggagct 16
 <210> 117
 <211> 16
 50 <212> ADN
 <213> desconocido
 <220>
 55 <223> organismo del suelo no identificado
 <400> 117
 ctccggtgcc gcagct 16
 60 <210> 118
 <211> 16
 <212> ADN
 <213> desconocido
 65 <220>
 <223> organismo del suelo no identificado

<400> 118
 cttcggtgct gcagct 16
 5 <210> 119
 <211> 16
 <212> ADN
 <213> desconocido
 10 <220>
 <223> organismo del suelo no identificado

 <400> 119
 cttgggtgcc gcagct 16
 15 <210> 120
 <211> 16
 <212> ADN
 <213> desconocido
 20 <220>
 <223> organismo del suelo no identificado

 <400> 120
 ctttagtaac gcagct 16
 25 <210> 121
 <211> 16
 <212> ADN
 <213> desconocido
 30 <220>
 <223> organismo del suelo no identificado

 <400> 121
 gacccgcaag ggagct 16
 35 <210> 122
 <211> 16
 <212> ADN
 <213> desconocido
 40 <220>
 <223> organismo del suelo no identificado

 <400> 122
 gatccgtgcc gcagct 16
 45 <210> 123
 <211> 16
 <212> ADN
 <213> desconocido
 50 <220>
 <223> organismo del suelo no identificado

 <400> 123
 gctccgtgcc gaagct 16
 55 <210> 124
 <211> 16
 <212> ADN
 <213> desconocido
 60 <220>
 <223> organismo del suelo no identificado

 <400> 123
 gctccgtgcc gaagct 16
 65 <210> 124
 <211> 16
 <212> ADN
 <213> desconocido
 70 <220>
 <223> organismo del suelo no identificado

<400> 124
 gctccgtgcc gtagct 16
 5 <210> 125
 <211> 16
 <212> ADN
 <213> desconocido
 10 <220>
 <223> organismo del suelo no identificado

 <400> 125
 gctctgtgcc gaagct 16
 15 <210> 126
 <211> 16
 <212> ADN
 <213> desconocido
 20 <220>
 <223> organismo del suelo no identificado

 <400> 126
 gctctgtgcc gtagct 16
 25 <210> 127
 <211> 16
 <212> ADN
 <213> desconocido
 30 <220>
 <223> organismo del suelo no identificado

 <400> 127
 gggcttgtcg tcagct 16
 35 <210> 128
 <211> 16
 <212> ADN
 <213> desconocido
 40 <220>
 <223> organismo del suelo no identificado

 <400> 128
 gtaaacgatg gaagct 16
 45 <210> 129
 <211> 16
 <212> ADN
 <213> desconocido
 50 <220>
 <223> organismo del suelo no identificado

 <400> 129
 gttccgtgcc gcagct 16
 55 <210> 130
 <211> 16
 <212> ADN
 <213> desconocido
 60 <220>
 <223> organismo del suelo no identificado

 <400> 129
 gttccgtgcc gcagct 16
 65 <210> 130
 <211> 16
 <212> ADN
 <213> desconocido
 70 <220>
 <223> organismo del suelo no identificado

ES 2 632 538 T3

<400> 130
 gttccgtgcc gtagct 16
 5 <210> 131
 <211> 16
 <212> ADN
 <213> desconocido
 10 <220>
 <223> organismo del suelo no identificado

 <400> 131
 gttctgtgcc gcagct 16
 15 <210> 132
 <211> 16
 <212> ADN
 <213> desconocido
 20 <220>
 <223> organismo del suelo no identificado

 <400> 132
 tctcagaca cgagct 16
 25 <210> 133
 <211> 16
 <212> ADN
 <213> desconocido
 30 <220>
 <223> organismo del suelo no identificado
 35 <400> 133
 tctcagtaac gtagct 16
 <210> 134
 <211> 16
 <212> ADN
 <213> desconocido
 40 <220>
 <223> organismo del suelo no identificado
 45 <400> 134
 tctccgtgcc gcagct 16
 <210> 135
 <211> 16
 <212> ADN
 <213> desconocido
 50 <220>
 <223> organismo del suelo no identificado

 <400> 135
 tctctgtgcc gcagct 16
 55 <210> 136
 <211> 16
 <212> ADN
 <213> Desconocido
 60 <220>
 65 <220>

<223> organismo del suelo no identificado
 <400> 136
 tggacgttgc ggagct 16
 5
 <210> 137
 <211> 16
 <212> ADN
 <213> desconocido
 10
 <220>
 <223> organismo del suelo no identificado
 <400> 137
 ttccgtgcc ggagct 16
 15
 <210> 138
 <211> 20
 <212> ADN
 <213> Secuencia Artificial
 20
 <220>
 <223> Oligonucleótido
 25
 <220>
 <221> modified_base
 <222> (1)..(1)
 <223> Biotina
 30
 <400> 138
 gtgtaghrgt gaaatdcodya 20
 <210> 139
 <211> 21
 <212> ADN
 <213> Secuencia Artificial
 35
 <220>
 <223> Oligonucleótido
 40
 <400> 139
 ytcacgrcay gagctgacga c 21
 <210> 140
 <211> 28
 <212> ADN
 <213> Secuencia Artificial
 45
 <220>
 <223> Oligonucleótido
 50
 <220>
 <221> modified_base
 <222> (1)..(1)
 <223> 5' Fosfato
 55
 <220>
 <221> modified_base
 <222> (28)..(28)
 <223> Didesoxi
 60
 <400> 140
 gctccaggtc tacatcctag tcaggacc 28
 65
 <210> 141
 <211> 24

<212> ADN
 <213> Secuencia Artificial
 <220>
 5 <223> Oligonucleótido
 <400> 141
 ataggtcctg actaggatgt agac 24
 10 <210> 142
 <211> 28
 <212> ADN
 <213> Secuencia Artificial
 15 <220>
 <223> Oligonucleótido
 <220>
 <221> modified_base
 20 <222> (1)..(1)
 <223> 5' Fosfato
 <220>
 <221> modified_base
 25 <222> (28)..(28)
 <223> Didesoxi
 <400> 142
 30 gctccagact agcatccgct gacttgac 28
 <210> 143
 <211> 24
 <212> ADN
 35 <213> Secuencia Artificial
 <220>
 <223> Oligonucleótido
 40 <400> 143
 aatgtcaagt cagcggatgc tagt 24
 <210> 144
 <211> 21
 <212> ADN
 <213> Secuencia Artificial
 45 <220>
 <223> Oligonucleótido
 50 <220>
 <221> modified_base
 <222> (1)..(1)
 <223> Biotina
 55 <400> 144
 ggtcctgact aggatgtaga c 21
 60 <210> 145
 <211> 21
 <212> ADN
 <213> Secuencia Artificial
 65 <220>
 <223> Oligonucleótido

5 <220>
 <221> modified_base
 <222> (1)..(1)
 <223> Biotina

10 <400> 145
 gtcaagtcag cggatgctag t 21
 <210> 146
 <211> 21
 <212> ADN
 <213> Secuencia Artificial

15 <220>
 <223> Oligonucleótido

20 <220>
 <221> modified_base
 <222> (1)..(1)
 <223> Biotina

25 <400> 146
 ggattagawa cccbgtagt c 21
 <210> 147
 <211> 30
 <212> ADN
 <213> Secuencia Artificial

30 <220>
 <223> Oligonucleótido

35 <400> 147
 ataggtcctg actaggatgt agacctggag 30
 <210> 148
 <211> 30
 <212> ADN
 <213> Secuencia Artificial

40 <220>
 <223> Oligonucleótido

45 <400> 148
 aatgtcaagt cagcggatgc tagtctggag 30
 <210> 149
 <211> 27
 <212> ADN
 <213> Secuencia Artificial

50 <220>
 <223> Oligonucleótido

55 <220>
 <221> modified_base
 <222> (1)..(1)
 <223> 5' Fosfato

60 <220>
 <221> modified_base
 <222> (27)..(27)

<223> Didesoxi

5 <400> 149
ctccaggtct acatcctagt caggacc 27

10 <210> 150
<211> 30
<212> ADN
<213> Secuencia Artificial

15 <220>
<223> Oligonucleótido

20 <400> 150
ataggtcctg actaggatgt agacctggag 30

25 <210> 151
<211> 27
<212> ADN
<213> Secuencia Artificial

30 <220>
<223> Oligonucleótido

35 <220>
<221> modified_base
<222> (1)..(1)
<223> 5' Fosfato

40 <220>
<221> modified_base
<222> (27)..(27)
<223> Didesoxi

45 <400> 151
ctccagacta gcatccgctg acttgac 27

50 <210> 152
<211> 30
<212> ADN
<213> Secuencia Artificial

55 <220>
<223> Oligonucleótido

60 <400> 152
aatgtcaagt cagcggatgc tagtctggag 30

65 <210> 153
<211> 33
<212> ADN
<213> Secuencia Artificial

70 <220>
<223> Oligonucleótido

75 <220>
<221> modified_base
<222> (1)..(1)
<223> Biotina

80 <400> 153
ctcgatagtc acggtcctga ctaggatgta gac 33

<210> 154
 <211> 33
 <212> ADN
 <213> Secuencia Artificial
 5
 <220>
 <223> Oligonucleótido
 <220>
 <221> modified_base
 <222> (1)..(1)
 <223> Biotina
 10
 <400> 154
 cagggtgcaa cggcaagtc agcggatgct agt 33
 <210> 155
 <211> 22
 <212> ADN
 <213> Secuencia Artificial
 20
 <220>
 <223> Oligonucleótido
 25
 <220>
 <221> modified_base
 <222> (1)..(1)
 <223> Biotina
 30
 <220>
 <221> misc_feature
 <222> (22)..(22)
 <223> A, G, C o T
 35
 <400> 155
 taytycccar gcggybsrct tn 22
 <210> 156
 <211> 25
 <212> ADN
 <213> Secuencia Artificial
 40
 <220>
 <223> Oligonucleótido
 45
 <220>
 <221> misc_feature
 <222> (25)..(25)
 <223> A, G, C o T
 50
 <400> 156
 ggtgbcascm gccgcgtaa yaccn 25
 <210> 157
 <211> 22
 <212> ADN
 <213> Secuencia Artificial
 55
 <220>
 <223> Oligonucleótido
 60
 <220>
 <221> misc_feature
 <222> (14)..(22)
 <223> A, G, C o T
 65

<400> 157
 aacgagcgca accnnnnnnn nn 22

5

<210> 158
 <211> 16
 <212> ADN
 <213> *Desulfurobacterium thermolithotrophum*

10

<400> 158
 gtcagttgcc gaagct 16

15

<210> 159
 <211> 16
 <212> ADN
 <213> Desconocido

20

<220>
 <223> Acuificales OPS132 no cultivado

<400> 159
 gtcctgccc taagct 16

25

<210> 160
 <211> 16
 <212> ADN
 <213> *Bacteroides caccae*

30

<400> 160
 atggtgtcgc tcagct 16

35

<210> 161
 <211> 16
 <212> ADN
 <213> *Actinomyces bovis*

<400> 161
 tttccgccc gtagct 16

40

<210> 162
 <211> 16
 <212> ADN
 <213> *Actinomyces meyeri*

45

<400> 162
 tttctgccc gtagct 16

50

<210> 163
 <211> 16
 <212> ADN
 <213> *Denitrobacterium detoxificans*

55

<400> 163
 cctccgccc gcagct 16

60

<210> 164
 <211> 16
 <212> ADN
 <213> Desconocido

<220>
 <223> Bacteria BPC110 GNS no cultivada

65

<400> 164
 cccgtagtc ctgact 16

<210> 165
 <211> 16
 <212> ADN
 <213> Desconocido
 5
 <220>
 <223> Bacteria GCA004 GNS no cultivada
 <400> 165
 10 catcgggtgcc gcagct 16
 <210> 166
 <211> 16
 <212> ADN
 15 <213> Desconocido
 <220>
 <223> Bacteria GCA112 GNS no cultivada
 20 <400> 166
 cggcgggtgcc gtagct 16
 <210> 167
 <211> 16
 25 <212> ADN
 <213> *Acetobacter acetii*
 <400> 167
 30 actcagtgtc gtagct 16
 <210> 168
 <211> 16
 <212> ADN
 35 <213> *Gluconobacter asaii*
 <400> 168
 actcagtgtc gaagct 16
 <210> 169
 <211> 16
 40 <212> ADN
 <213> Desconocido
 <220>
 45 <223> *Burkholderia sp.* JB1
 <400> 169
 ccttagtaac gaagct 16
 <210> 170
 <211> 16
 <212> ADN
 50 <213> *Denitrobacter permanens*
 <400> 170
 55 acgatgtcaa ctagct 16
 <210> 171
 <211> 16
 <212> ADN
 60 <213> *Desulfobacter curvatus*
 <220>
 <221> misc_feature
 65 <222> (11)..(11)
 <223> A, G, C o T

ES 2 632 538 T3

<400> 171
 ctgctgtgcc naagct 16
 5 <210> 172
 <211> 16
 <212> ADN
 <213> Desconocido
 10 <220>
 <223> *Desulfobulbus* sp. BG25
 <400> 172
 cctctgtgct gcagct 16
 15 <210> 173
 <211> 16
 <212> ADN
 <213> *Legionella anisa*
 20 <400> 173
 acgatgtcaa ctagct 16
 25 <210> 174
 <211> 16
 <212> ADN
 <213> Desconocido
 30 <220>
 <223> Clon SB-1 mineralizante de benceno
 <400> 174
 atggtgtcg gcagct 16
 35 <210> 175
 <211> 16
 <212> ADN
 <213> *Escherichia coli*
 40 <400> 175
 cgtggcttcc ggagct 16
 45 <210> 176
 <211> 16
 <212> ADN
 <213> Desconocido
 50 <220>
 <223> *Acidobacterium* Sub.Div-1 no cultivada
 <400> 176
 ccgccgtgcc gaagct 16
 55 <210> 177
 <211> 16
 <212> ADN
 <213> Desconocido
 60 <220>
 <223> *Acidobacterium* Sub.Div-1 no cultivada
 <400> 177
 cggctgtgcc gaagct 16
 65

<210> 178
 <211> 16
 <212> ADN
 <213> Desconocido
 5
 <220>
 <223> *Acidobacterium* Sub.Div-1 no cultivada
 <400> 178
 10 ccactgtgcc gtagct 16
 <210> 179
 <211> 16
 <212> ADN
 15 <213> Desconocido
 <220>
 <223> *Acidobacterium* Sub.Div-1 no cultivada
 20 <400> 179
 ctgctgtgcc gcagct 16
 <210> 180
 <211> 16
 25 <212> ADN
 <213> Desconocido
 <220>
 <223> *Acidobacterium* Sub.Div-1 no cultivada
 30 <400> 180
 ctgccgtgcc ggagct 16
 <210> 181
 <211> 16
 35 <212> ADN
 <213> Desconocido
 <220>
 <223> *Acidobacterium* Sub.Div-1 no cultivada
 40 <400> 181
 ccaatgtgcc ggagct 16
 <210> 182
 <211> 16
 45 <212> ADN
 <213> Desconocido
 <220>
 <223> *Acidobacterium* Sub.Div-1 no cultivada
 50 <400> 182
 ccgtcgtgcc gtagct 16
 <210> 183
 <211> 16
 55 <212> ADN
 <213> Desconocido
 60 <220>
 <223> *Acidobacterium* Sub.Div-1 no cultivada
 <400> 183
 65 ccgtcgtgct gtagct 16

ES 2 632 538 T3

5
<210> 184
<211> 16
<212> ADN
<213> Desconocido

<220>
<223> *Acidobacterium* Sub.Div-1 no cultivada

10
<400> 184
ctgccgtgtc gaagct 16

<210> 185
<211> 16
<212> ADN
15
<213> Desconocido

<220>
<223> *Acidobacterium* Sub.Div-1 no cultivada

20
<400> 185
ctcccgtgtc gaagct 16

<210> 186
<211> 16
25
<212> ADN
<213> Desconocido

<220>
<223> *Acidobacterium* Sub.Div-1 no cultivada

30
<400> 186
ccgccgtgcc ggagct 16

<210> 187
<211> 16
35
<212> ADN
<213> Desconocido

<220>
40
<223> *Acidobacterium* Sub.Div-2 no cultivada

<400> 187
ctgaggaacg aaagct 16

45
<210> 188
<211> 16
<212> ADN
<213> Desconocido

50
<220>
<223> *Acidobacterium* Sub.Div-2 no cultivada

<400> 188
55
gtgtcgtccc ggagct 16

<210> 189
<211> 16
<212> ADN
60
<213> Desconocido

<220>
<223> *Acidobacterium* Sub.Div-3 no cultivada

65
<400> 189
gggctgtgcc gaagct 16

<210> 190
 <211> 16
 <212> ADN
 <213> Desconocido
 5
 <220>
 <223> *Acidobacterium* Sub.Div-3 no cultivada
 <400> 190
 10 ggtcggtgcc ggagct 16
 <210> 191
 <211> 16
 <212> ADN
 15 <213> Desconocido
 <220>
 <223> *Acidobacterium* Sub.Div-3 no cultivada
 20 <400> 191
 ggtcggtgcc agagct 16
 <210> 192
 <211> 16
 25 <212> ADN
 <213> Desconocido
 <220>
 <223> *Acidobacterium* Sub.Div-3 no cultivada
 30 <400> 192
 gggtcgtgcc ggagct 16
 <210> 193
 <211> 16
 35 <212> ADN
 <213> Desconocido
 <220>
 <223> *Acidobacterium* Sub.Div-3 no cultivada
 40 <400> 193
 tgtctgtgcc ggagct 16
 <210> 194
 <211> 16
 45 <212> ADN
 <213> Desconocido
 <220>
 <223> *Acidobacterium* Sub.Div-3 no cultivada
 50 <400> 194
 tatccgtgcc ggagct 16
 <210> 195
 <211> 16
 55 <212> ADN
 <213> Desconocido
 60 <220>
 <223> *Acidobacterium* Sub.Div-3 no cultivada
 <400> 195
 65 ggtccgtgcc ggagct 16

<210> 196
 <211> 14
 <212> ADN
 5 <213> *Aeropyrum pernix*
 <400> 196
 ctagggggcg ggag 14
 10 <210> 197
 <211> 14
 <212> ADN
 <213> *Desulfurococcus mobilis*
 15 <400> 197
 ctaggtgttg ggtg 14
 <210> 198
 <211> 14
 <212> ADN
 <213> *Staphylothermus marinus*
 20 <400> 198
 ctaggtgttg ggcg 14
 <210> 199
 <211> 14
 <212> ADN
 30 <213> *Metallosphaera sedula*
 <400> 199
 ctaggtgtcg cgta 14
 35 <210> 200
 <211> 14
 <212> ADN
 <213> *Sulfolobus acidocaldarius*
 40 <400> 200
 ctaggtgtcg agta 14
 <210> 201
 <211> 14
 <212> ADN
 <213> *Sulfolobus metallicus*
 45 <400> 201
 ctaggtgtca cgtg 14
 <210> 202
 <211> 14
 <212> ADN
 <213> *Caldivirga maquilingensis*
 50 <400> 202
 ctagctgttg ggtg 14
 <210> 203
 <211> 14
 <212> ADN
 <213> *Pyrobaculum islandicum*
 60 <400> 203
 ctagctgtcg gccg 14
 65

ES 2 632 538 T3

<210> 204
 <211> 14
 <212> ADN
 <213> *Archaeoglobus fulgidus*
 5
 <400> 204
 ctaggtgtca ccga 14
 <210> 205
 <211> 14
 <212> ADN
 <213> *Archaeoglobus veneficus*
 10
 <400> 205
 ctaggtgtca ccgg 14
 <210> 206
 <211> 14
 <212> ADN
 <213> *Haloarcula japonica*
 20
 <400> 206
 ctaggtgtgg cgta 14
 <210> 207
 <211> 14
 <212> ADN
 <213> *Halococcus morrhuae*
 25
 <400> 207
 ctaggtgtgg cgtt 14
 <210> 208
 <211> 14
 <212> ADN
 <213> *Methanococcus jannaschii*
 30
 <400> 208
 ctaggtgtcg cgtc 14
 <210> 209
 <211> 14
 <212> ADN
 <213> *Pyrococcus abyssi*
 35
 <400> 209
 ctaggtgtcg ggcg 14
 <210> 210
 <211> 14
 <212> ADN
 <213> *Picrophilus oshimae*
 40
 <400> 210
 ctagctgtaa actc 14
 <210> 211
 <211> 11
 <212> PRT
 <213> *Desulfurobacterium thermolithotrophum*
 45
 <400> 211

Arg Ala Gln Pro Leu Ser Leu Val Ala Ser Gly
 1 5 10

ES 2 632 538 T3

<210> 212
 <211> 11
 <212> PRT
 <213> Desconocido
 5
 <220>
 <223> Acuificales OPS132 no cultivado
 <400> 212
 10
 Arg Ala Gln Pro Leu Ser Cys Val Thr Ser Gly
 1 5 10
 <210> 213
 <211> 13
 <212> PRT
 <213> *Bacteroides caccae*
 <400> 213
 20
 Arg Ala Gln Pro Leu Ser Ser Val Thr Asn Arg Ser Cys
 1 5 10
 <210> 214
 <211> 19
 <212> PRT
 <213> *Actinomyces bovis*
 <400> 214
 25
 Arg Ala Gln Pro Leu Ser Arg Val Ala Ser Thr Leu Trp Trp Gly Leu
 1 5 10 15
 Ala Gly Asp
 30
 <210> 215
 <211> 19
 <212> PRT
 <213> *Actinomyces meyeri*
 <400> 215
 35
 Arg Ala Gln Pro Leu Pro Tyr Val Ala Ser Thr Leu Trp Trp Gly Leu
 1 5 10 15
 Val Gly Asp
 40
 <210> 216
 <211> 18
 <212> PRT
 <213> *Denitrobacterium detoxificans*
 <400> 216
 45
 Arg Ala Gln Pro Leu Pro His Val Ala Ser Ile Arg Leu Gly Thr His
 1 5 10 15
 Gly Gly

ES 2 632 538 T3

<210> 217
 <211> 14
 <212> PRT
 <213> Desconocido
 5
 <220>
 <223> Bacteria BPC110 GNS no cultivada
 <400> 217
 10
 Arg Ala Gln Pro Leu Leu Tyr Val Ile Arg Val Ile Pro Asp
 1 5 10

 <210> 218
 <211> 14
 <212> PRT
 <213> Desconocido
 15
 <220>
 <223> Bacteria GCA004 GNS no cultivada
 20
 <400> 218
 Arg Ala Gln Pro Ser Leu Tyr Val Thr Arg Ile Ile Arg Asp
 1 5 10

 <210> 219
 <211> 14
 <212> PRT
 <213> Desconocido
 25
 <220>
 <223> Bacteria GCA112 GNS no cultivada
 30
 <900> 219
 Arg Ala Gln Pro Ser Pro Tyr Val Ile Arg Val Ile Arg Asp
 1 5 10
 35

 <210> 220
 <211> 16
 <212> PRT
 <213> *Acetobacter acetii*
 40
 <400> 220
 Arg Ala Gln Pro Leu Ser Leu Val Ala Ser Met Phe Gly Trp Ala Leu
 1 5 10 15
 45

 <210> 221
 <211> 16
 <212> PRT
 <213> *Gluconobacter asaii*
 50
 <400> 221
 Arg Ala Gln Pro Leu Ser Leu Val Ala Ser Thr Phe Arg Trp Ala Leu
 1 5 10 15
 55

 <210> 222
 <211> 17
 <212> PRT
 <213> Desconocido

ES 2 632 538 T3

<220>

<223> *Burkholderia* sp. JB1

<400> 222

5

Arg Ala Gln Pro Leu Ser Leu Val Ala Thr Gln Glu His Ser Arg Glu
1 5 10 15

Thr

<210> 223

<211> 15

10

<212> PRT

<213> *Denitrobacter permanens*

<400> 223

15

Arg Ala Gln Pro Leu Pro Leu Val Ala Thr Phe Ser Trp Ala Leu
1 5 10 15

<210> 224

<211> 19

20

<212> PRT

<213> *Desullobacter curvatus*

<400> 224

Arg Ala Gln Pro Leu Ser Leu Val Ala Ser Thr Leu Cys Gly Asn Ser
1 5 10 15

25

Asn Glu Thr

<210> 225

<211> 19

30

<212> PRT

<213> Desconocido

<220>

<223> *Desulfobulbus* sp. BG25

35

<400> 225

Arg Ala Gln Pro Leu Pro Leu Val Ala Ser Ser Ser Ala Gly His Ser
1 5 10 15

Lys Gly Thr

40

<210> 226

<211> 11

<212> PRT

<213> *Legionella anisa*

45

<400> 226

Arg Ala Gln Pro Leu Ser Leu Val Ala Ser Thr
1 5 10

<210> 227

ES 2 632 538 T3

<211> 16
 <212> PRT
 <213> Desconocido
 5 <220>
 <223> Clon SB-1 mineralizante de benceno
 <400> 227
 10 Arg Ala Gln Pro Leu Pro Leu Val Ala Asn Arg Ser Ser Trp Gly Leu
 1 5 10 15
 <210> 228
 <211> 19
 <212> PRT
 15 <213> Escherichia coli
 <400> 228
 Arg Ala Gln Pro Leu Ser Phe Val Ala Ser Gly Pro Ala Gly Asn Ser
 1 5 10 15
 Lys Glu Thr
 20 <210> 229
 <211> 16
 <212> PRT
 <213> Desconocido
 25 <220>
 <223> *Acidobacterium* Sub.Div-1 no cultivada
 <400> 229
 30 Arg Ala Gln Pro Leu Ser Leu Val Ala Ser Gly Ser Ser Arg Ala Leu
 1 5 10 15
 <210> 230
 <211> 19
 35 <212> PRT
 <213> Desconocido
 <220>
 <223> *Acidobacterium* Sub.Div-1 no cultivada
 40 <400> 230
 Arg Ala Gln Pro Leu Ser Ser Val Ala Ile Gly Ser Ser Arg Ala Thr
 1 5 10 15
 Leu Ala Lys
 45 <210> 231
 <211> 11
 <212> PRT
 <213> Desconocido
 50 <220>
 <223> *Acidobacterium* Sub.Div-1 no cultivada
 <400> 231

ES 2 632 538 T3

Arg Ala Gln Pro Leu Phe Ala Ser Cys His His
1 . 5 10

5 <210> 232
<211> 19
<212> PRT
<213> Desconocido

10 <220>
<223> *Acidobacterium* Sub.Div-1 no cultivada
<400> 232

Arg Ala Gln Pro Leu Phe Ala Gln Leu Pro Ser Phe Ser Trp Ala Leu
1 5 10 15

Cys Arg Asn

15 <210> 233
<211> 11
<212> PRT
<213> Desconocido

20 <220>
<223> *Acidobacterium* Sub.Div-1 no cultivada

25 <220>
<221> MISC_FEATURE
<222> (8)..(8)
<223> Imposible de determinar
<400> 233

Arg Ala Gln Pro Leu Leu Pro Xaa Ala Ile Ile
1 . 5 10

30 <210> 234
<211> 11
<212> PRT
35 <213> Desconocido

<220>
<223> *Acidobacterium* Sub.Div-1 no cultivada

40 <400> 234

Arg Ala Gln Pro Leu Leu Pro Val Ala Thr Ile
1 5 10

45 <210> 235
<211> 11
<212> PRT
<213> Desconocido

50 <220>
<223> *Acidobacterium* Sub.Div-1 no cultivada

<400> 235

Arg Ala Gln Pro Leu Ser Pro Val Ala Ile Ile
1 5 10

55

ES 2 632 538 T3

<210> 236
 <211> 11
 <212> PRT
 <213> Desconocido
 5
 <220>
 <223> *Acidobacterium* Sub.Div-1 no cultivada
 <400> 236
 10
 Arg Ala Gln Pro Leu Ser Ser Val Ala Thr Ile
 1 5 10
 <210> 237
 <211> 11
 <212> PRT
 <213> Desconocido
 15
 <220>
 <223> *Acidobacterium* Sub.Div-1 no cultivada
 20
 <400> 237
 Arg Ala Gln Pro Leu Phe Leu Val Ala Thr Ile
 1 5 10
 25
 <210> 238
 <211> 13
 <212> PRT
 <213> Desconocido
 30
 <220>
 <223> *Acidobacterium* Sub.Div-1 no cultivada
 <400> 238
 Arg Ala Gln Pro Ser Ser Leu Val Ala Asn Thr Leu Trp
 1 5 10
 35
 <210> 239
 <211> 18
 <212> PRT
 <213> Desconocido
 40
 <220>
 <223> *Acidobacterium* Sub.Div-2 no cultivada
 45
 <400> 239
 Arg Ala Gln Pro Leu His Val Val Ala Thr Arg Lys Arg Glu Leu Tyr
 1 5 10 15
 Val Asp
 50
 <210> 240
 <211> 18
 <212> PRT
 <213> Desconocido
 55
 <220>
 <223> *Acidobacterium* Sub.Div-2 no cultivada

ES 2 632 538 T3

<400> 240

Arg Ala Gln Pro Leu His Val Val Ala Thr Pro Gln Gly Gly Thr Leu
1 5 10 15

Arg Gly

5 <210> 241
<211> 19
<212> PRT
<213> Desconocido

10 <220>
<223> *Acidobacterium* Sub.Div-3 no cultivada

<400> 241

Arg Ala Gln Pro Ser Ser Leu Val Ala Asn Pro Gln Gly Lys His Pro
1 5 10 15

Lys Gly Thr

15 <210> 242
<211> 11
<212> PRT
<213> Desconocido

20 <220>
<223> *Acidobacterium* Sub.Div-3 no cultivada

25 <400> 242

Arg Ala Arg Pro Leu Ser Cys Val Ala Ile Ile
1 5 10

30 <210> 243
<211> 18
<212> PRT
<213> Desconocido

35 <220>
<223> *Acidobacterium* Sub.Div-3 no cultivada

<400> 243

Arg Ala Gln Pro Leu Ser Cys Val Ala Asn Pro Gln Gly Cys Thr Leu
1 5 10 15

Arg Arg

40 <210> 244
<211> 20
<212> PRT
<213> Desconocido

45 <220>
<223> *Acidobacterium* Sub.Div-3 no cultivada

50 <400> 244

ES 2 632 538 T3

Arg Ala Gln Pro Ser Pro Cys Val Ala Thr Pro Pro Arg Ala Gly Ala
1 5 10 15

Leu Ser Gly Asp
20

REIVINDICACIONES

1. Un método para la construcción de una base de datos de perfiles de diversidad de marcadores (PDM) que comprende las etapas de:
- a) secuenciar una pluralidad de marcadores de ARNr de una muestra, en el que dicha muestra comprende una población microbiana, y en el que cada marcador de ARNr comprende una secuencia polimórfica del gen del ARNr microbiano y en el que la pluralidad de marcadores de ARNr proporciona una representación exacta de la población microbiana de dicha muestra;
 - b) determinar la abundancia, en dicha muestra, de cada una de dicha pluralidad de marcadores de ARNr;
 - c) transducir la abundancia de cada marcador de ARNr en una señal de salida eléctrica;
 - d) almacenar dichas señales de salida eléctrica producidas en la etapa c) en una estructura de datos matricial y asociar, en dicha estructura, cada señal de salida con la secuencia correspondiente del marcador de ARNr cuya abundancia se convirtió en una señal de salida;
 - e) proporcionar las abundancias de uno o más parámetros de muestra que están asociados a dicha muestra;
 - f) transducir las abundancias de cada parámetro de muestra en una señal de salida eléctrica;
 - g) almacenar dichas señales de salida eléctrica producidas en la etapa f) en una estructura de datos matricial y asociar en dicha estructura cada señal de salida con el parámetro de muestra cuya abundancia se convirtió en una señal de salida;
 - h) diseñar como PDM las señales de salida eléctrica producidas en la etapa c) correspondientes a la pluralidad de abundancias de marcadores de ARNr y las señales de salida eléctrica producidas en la etapa f) correspondientes a la una o más abundancias de parámetros de muestra; y
 - i) repetir las etapas a-h para una pluralidad de muestras distintas, y diseñar, como base de datos PDM, la pluralidad de PDM almacenados en dicha estructura de datos.
2. El método de acuerdo con la reivindicación 1, en el que dicha abundancia de marcadores de ARNr es una abundancia relativa a la abundancia total de la pluralidad de marcadores de ARNr en la muestra.
3. El método de acuerdo con la reivindicación 1 o 2, en el que dicho gen de ARNr microbiano es un gen de ARNr 16S.
4. El método de acuerdo con la reivindicación 1 o 2, en el que dicha secuencia polimórfica del gen de ARNr microbiano es de la región intergénica entre un gen de ARNr 16S y un gen de ARNr 23S.
5. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 4, en el que la muestra se selecciona del grupo constituido por: una muestra de suelo, una muestra de roca, una muestra de agua, una muestra de aire, una muestra de hidrocarburo, una muestra de petróleo y una muestra de biopelícula.
6. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 5, en el que la muestra se obtiene del grupo constituido por: petróleo, hidrocarburos, un depósito de aceite, un depósito de gas, suelo, roca, agua, aire, una construcción y una carretera.
7. El método de acuerdo con la reivindicación 6, en el que el parámetro de muestra se selecciona del grupo constituido por: petróleo, hidrocarburos, aceite, gas, un compuesto inorgánico, un mineral, pH, temperatura y salinidad.
8. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 4, en el que la muestra se obtiene del grupo constituido por: un ser humano, una planta, un animal, un producto alimentario, un cultivo celular y un cultivo tisular.
9. El método de acuerdo con la reivindicación 8, en el que el parámetro de muestra se selecciona del grupo constituido por: una patología aguda, una patología crónica, un estado fisiológico y una fase de desarrollo.
10. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 4, en el que el parámetro de muestra es el tiempo, y en el que la pluralidad de muestras comprende una evolución temporal.
11. El método de acuerdo con la reivindicación 10, en el que la pluralidad de muestras comprende una evolución temporal útil para supervisar esfuerzos de remediación ambiental o la producción de campos petrolíferos.
12. Una base de datos de PDM construida y almacenada en la memoria de un ordenador de acuerdo con el método de una cualquiera de las reivindicaciones 1 a 11.

Regiones polimórficas de ARNc 16s de bacterias

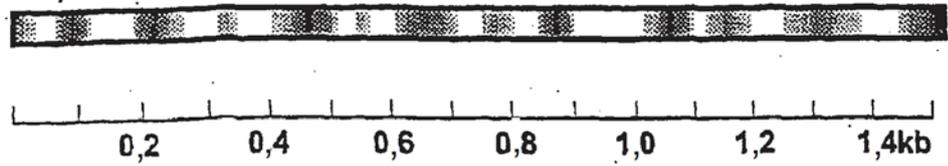


Figura 1.

Regiones polimórficas de ARNc 16s

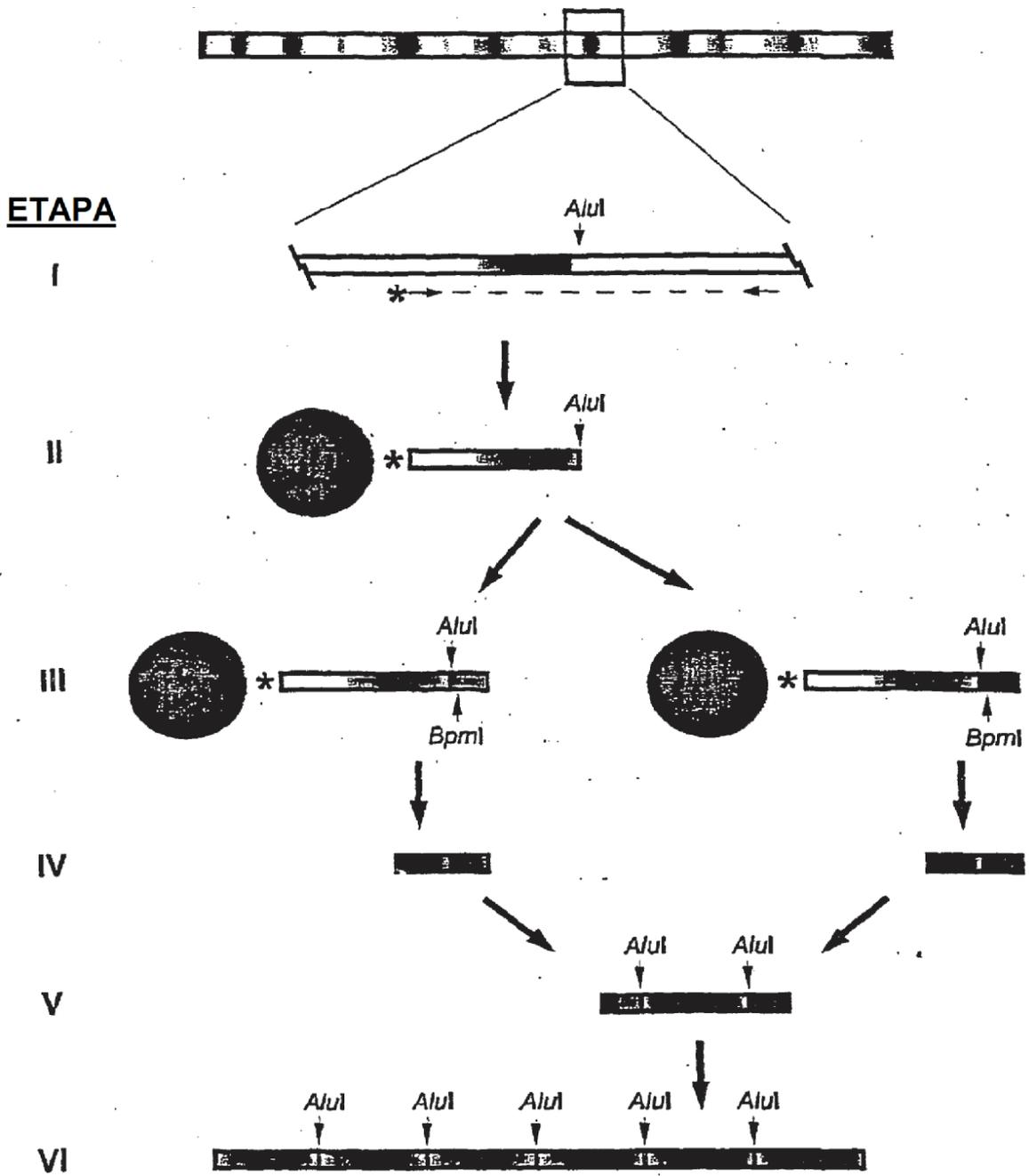


Figura 2

	Bacterias
Grupo Acidobacterias	Diversas acidobacterias ambientales no cultivadas
Acuificales	
Grupo Desulfurobacterium	
Desulfurobacterium	<i>Desulfurobacterium thermolithotrophum</i>
Muestras ambientales	Acuificales OPS132 no identificada
Grupo de bacterias verdes del azufre/CFB	
Bacteroidaceae	
Bacteroides	<i>Bacteroides caccae</i>
Firmicutes (bacterias gram positivas)	
Actinobacteria (bacterias gram positivas G+C alto)	
Actinomycetaceae	
Actinomyces	<i>Actinomyces bovis</i> <i>Actinomyces meyeri</i>
Denitrobacterium	<i>Denitrobacterium detoxificans</i>
Bacterias verdes no del azufre	
Muestras ambientales	Bacteria BPC110 de filtración HC no cultivada Bacteria GCA004 de filtración HC no cultivada Bacteria GCA112 de filtración HC no cultivada
Proteobacteria (bacteria púrpura no del azufre)	
Subdivisión alfa	
Acetobacteraceae	
Acetobacter	<i>Acetobacter aceti</i>
Gluconobacter	<i>Gluconobacter asaii</i>
Subdivisión beta	
Grupo de Burkholderia	
Burkholderia	<i>Burkholderia sp. JB1</i>
Denitrobacter	<i>Denitrobacter permanens</i>
Subdivisión delta	
Desulfobacter	<i>Desulfobacter curvatus</i>
Desulfobulbus	<i>Desulfobulbus sp. BG25</i>
Legionellaceae	
Legionella	<i>Legionella anisa</i>
Bacterias no clasificadas	Clon SB-1 mineralizante de benzeno Env.OPS 1 eubacteria no cultivada

Fig. 3

	Arqueas (arqueobacterias)
Crenarcaeota (arqueobacterias extremadamente termófilas)	
Desulfurococcales	
Desulfurococcaceae	
Aeropyrum	<i>Aeropyrum pernix</i>
Desulfurococcus	<i>Desulfurococcus mobilis</i>
Staphylothermus	<i>Staphylothermus marinus</i>
Sulfolobales	
Metallosphaera	<i>Metallosphaera sedula</i>
Sulfolobus	<i>Sulfolobus acidocaldarius</i> <i>Sulfolobus metallicus</i>
Thermoproteales	
Thermoproteaceae	
Candidivirga	<i>Candidivirga maquilingsensis</i>
Pyrobaculum	<i>Pyrobaculum islandicum</i>
Euriarcaeota	
Archaeoglobales	
Archaeoglobaceae	
Archaeoglobus	<i>Archaeoglobus fulgidus</i> <i>Archaeoglobus veneficus</i>
Halobacteriales	
Halobacteriaceae	
Haloarcula	<i>Haloarcula japonica</i>
Halococcus	<i>Halococcus morrhuae</i>
Methanococcales	
Methanococcaceae	
Methanococcus	<i>Methanococcus jannaschii</i>
Methanobacteriales	
Methanobacteriaceae	
Methanobacterium	<i>Methanobacterium bryantii</i> <i>Methanobacterium subterraneum</i>
Thennococcales	
Thermococcaceae	
Pyrococcus	<i>Pyrococcus abyssi</i>
Thermoplasmatales	
Picophilaceae	
Picophilus	<i>Picophilus oshimae</i>

Fig. 4

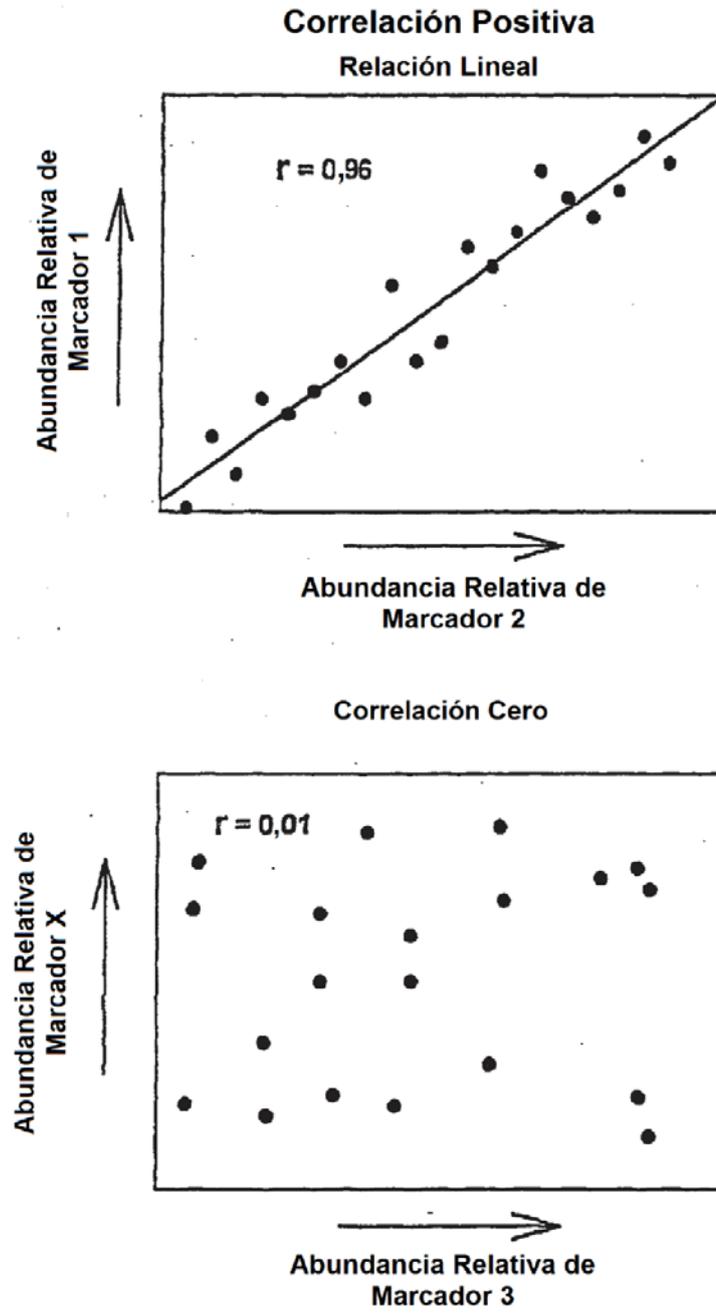


Fig. 5

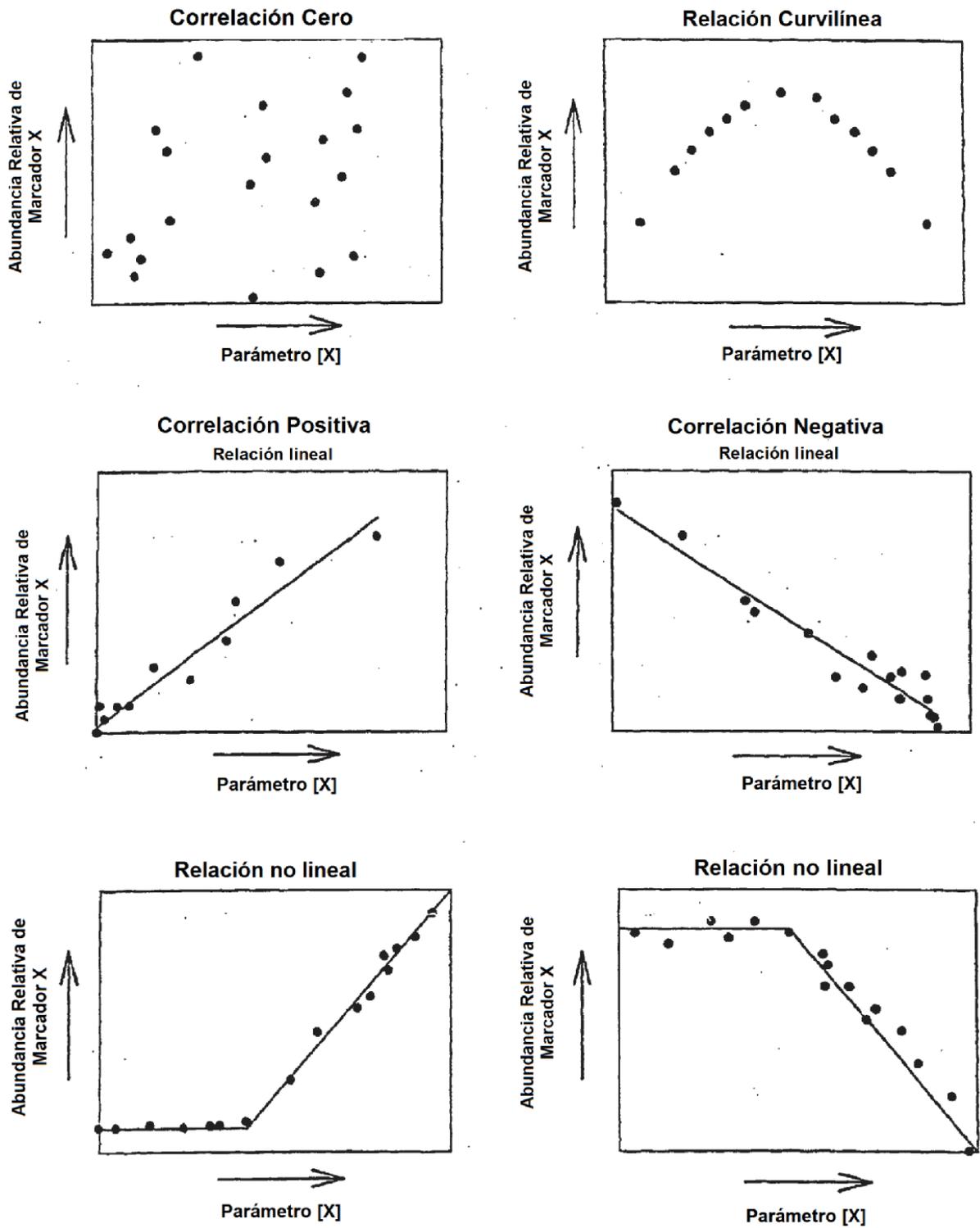


Fig. 6

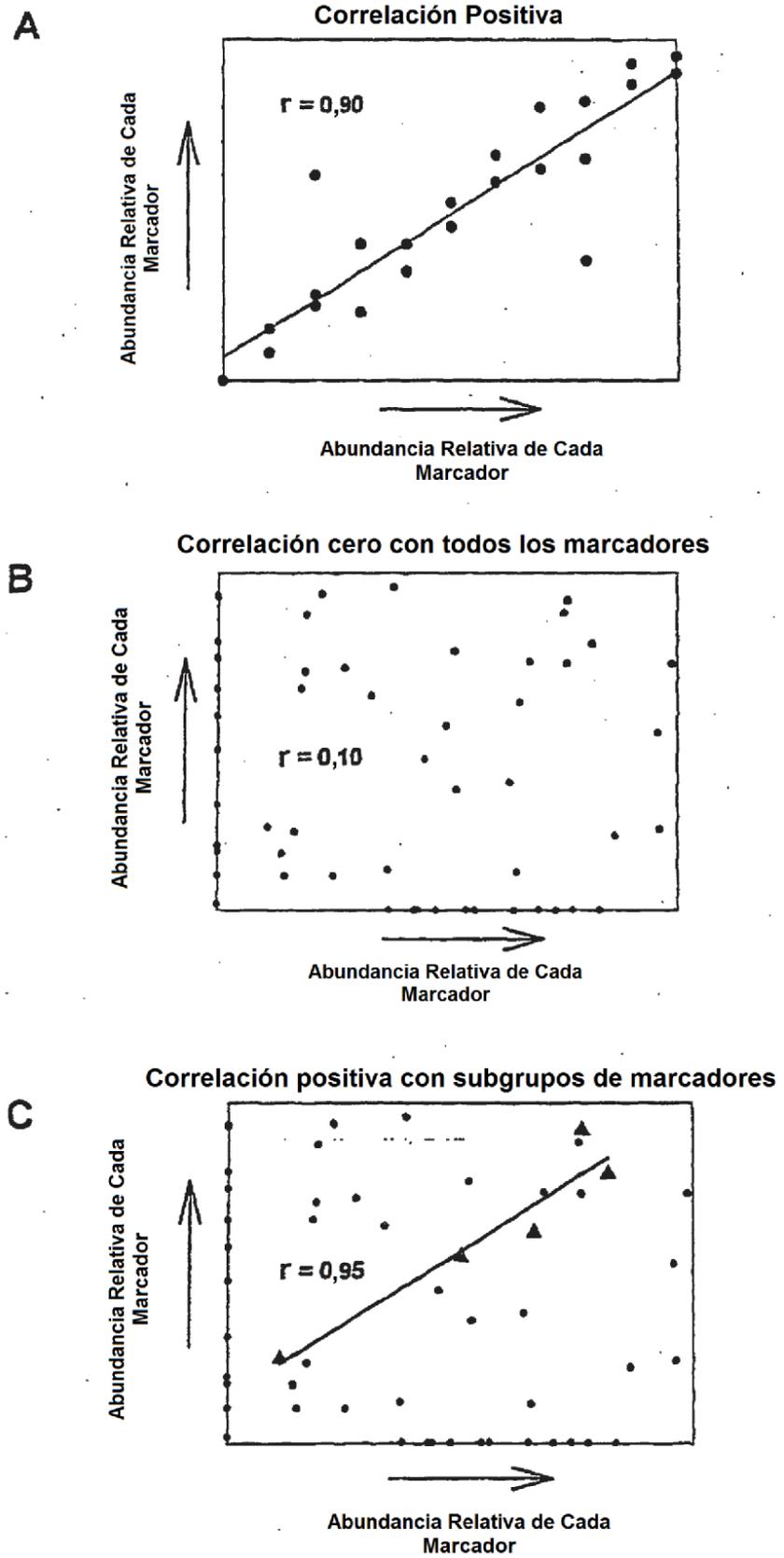


Fig. 7

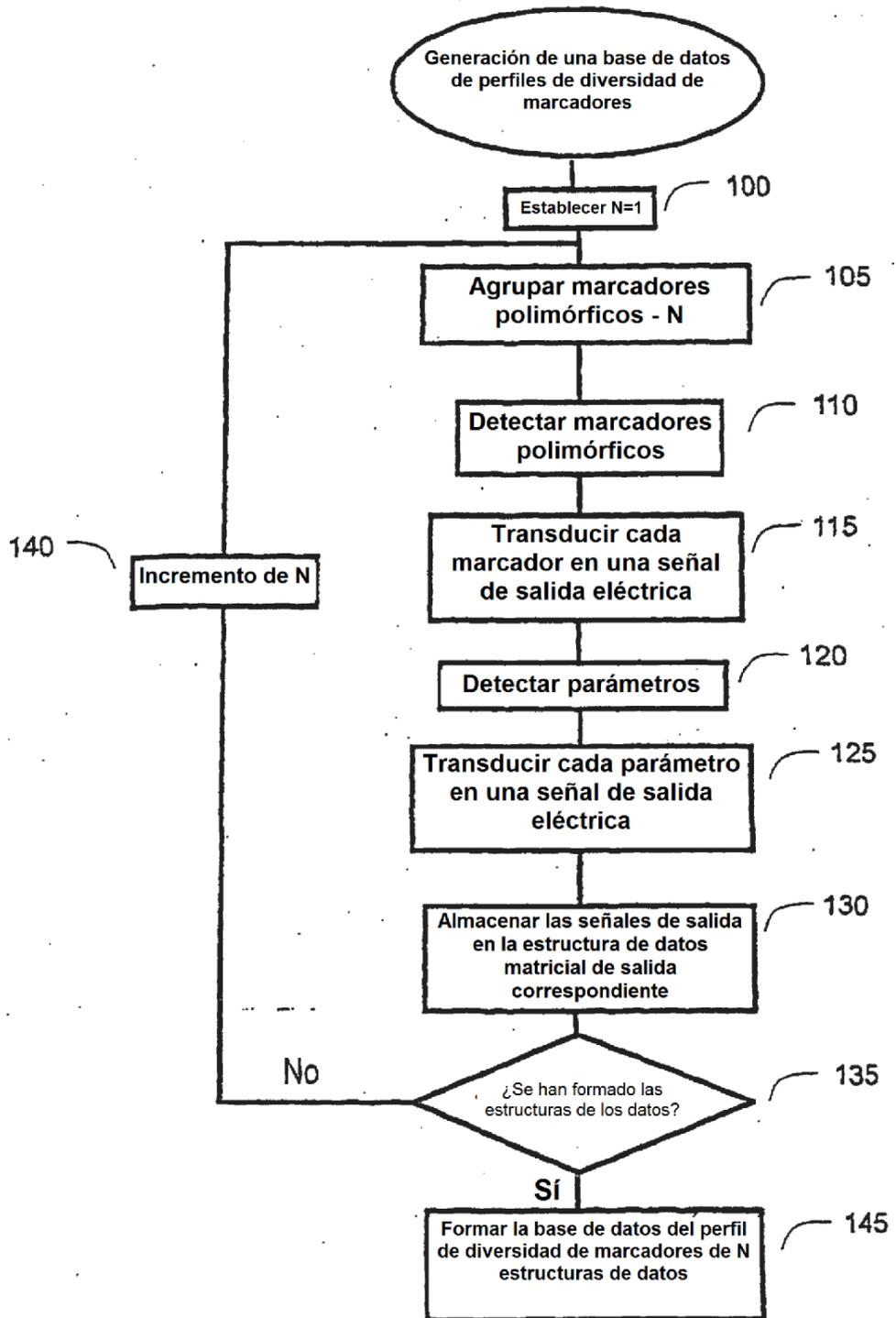


Figura 8

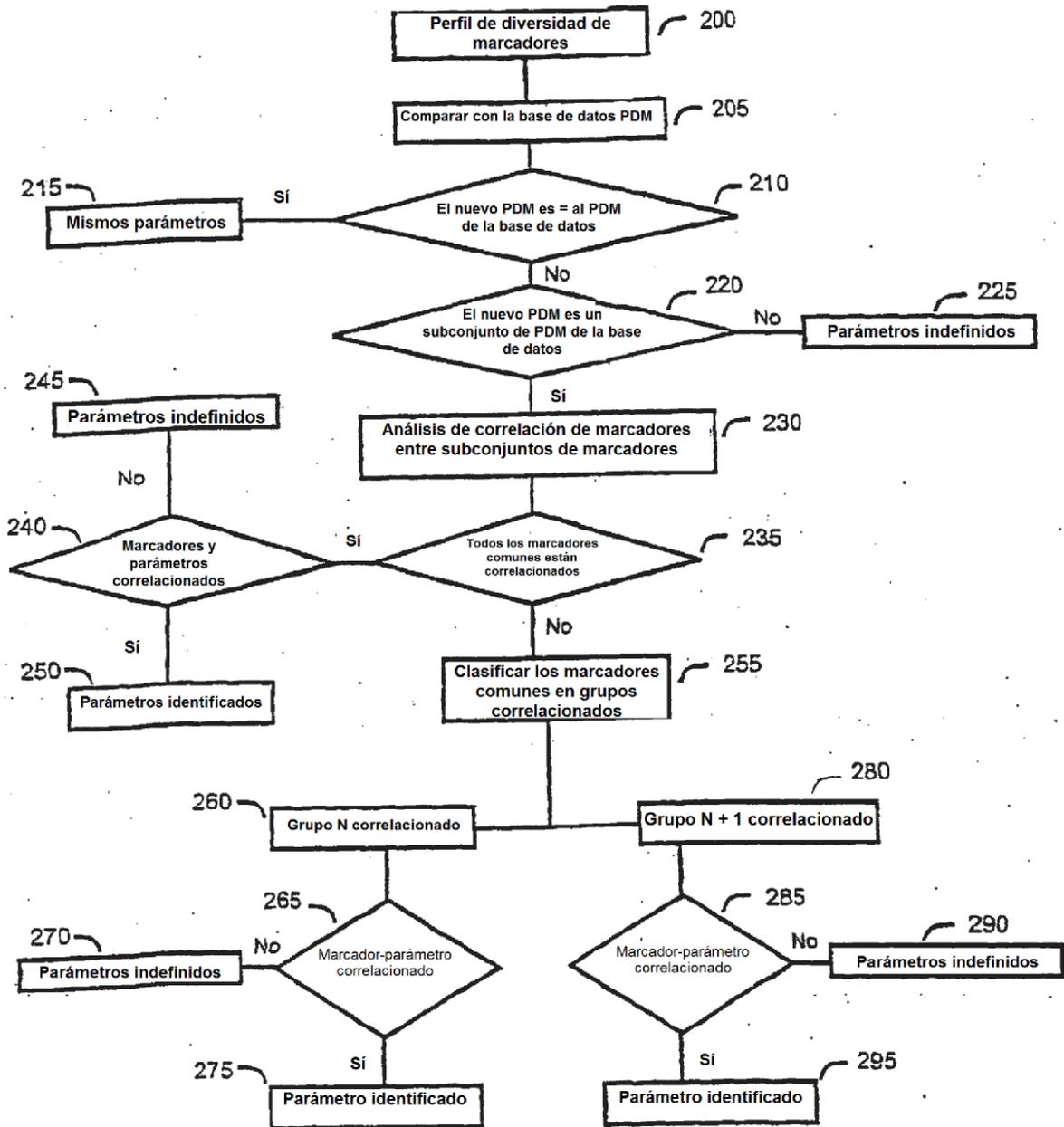


Figura 9

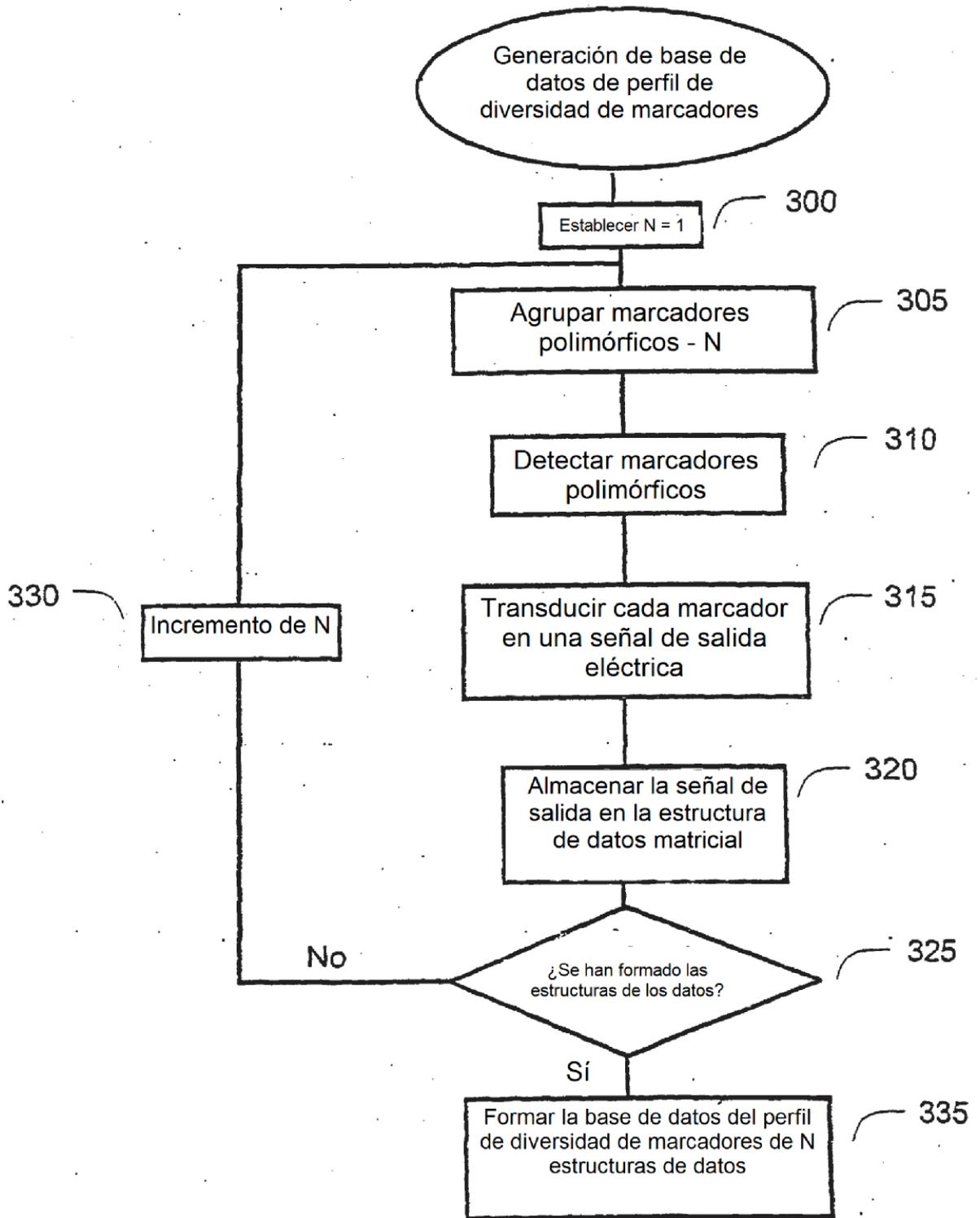


Figura 10

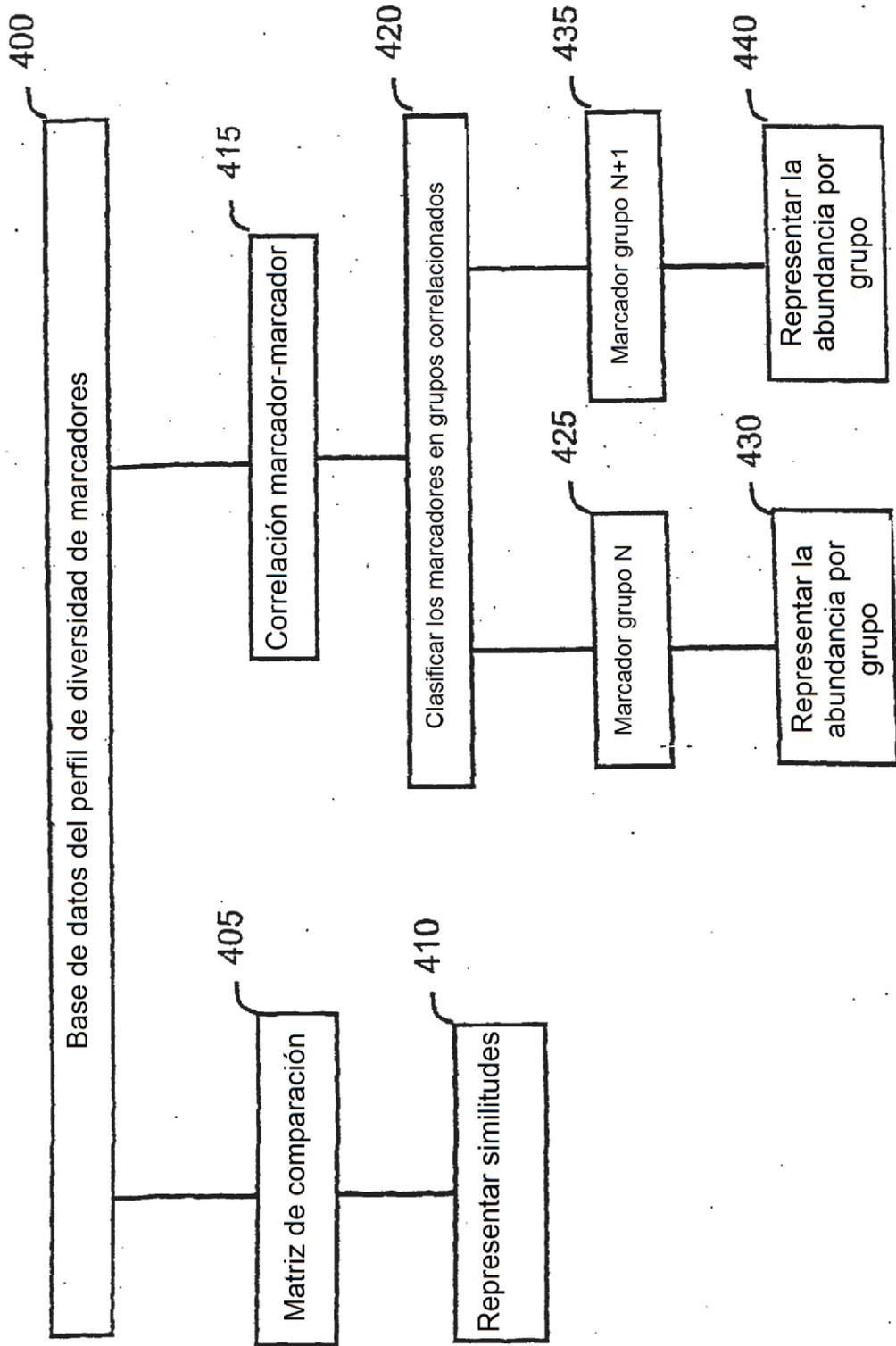


Figura 11

OLIGONUCLEÓTIDOS SARD

TX-003	5' -	Fosfato-GCTCCAGGTCTACATCCTAGTCAGGACddc-3'
TX-012	5' -	Fosfato- CTCCAGGTCTACATCCTAGTCAGGACddc
TX-010R	3'	GAGGTCCAGATGTAGGATCAGTCCTGGATA-5'
TX-013	3'	GAGGTCCAGATGTAGGATCAGTCCTGGATA-5'
TX-004R	3'	CAGATGTAGGATCAGTCCTGGATA-5'
TX-007R	3' -	CAGATGTAGGATCAGTCCTGG -Biotina-5'
TX-111	3' -	CAGATGTAGGATCAGTCCTGGCACTGATAGCTC -Biotina-5'
TX-005	5' -	Fosfato-GCTCCAGACTAGCATCCGCTGACTTGaddc
TX-014	5' -	Fosfato- CTCCAGACTAGCATCCGCTGACTTGaddc
TX-011R	3'	GAGGTCTGATCGTAGGCGACTGAACTGTAA
TX-015	3'	GAGGTCTGATCGTAGGCGACTGAACTGTAA
TX-006R	3'	TGATCGTAGGCGACTGAACTGTAA
TX-008R	3'	TGATCGTAGGCGACTGAACTG -Biotina-5'
TX-121	3'	TGATCGTAGGCGACTGAACTGGCAACGTTGGAC -Biotina-5'
TX-001	5' -	Biotina-GTG TAG HRG TGA AAT DCD YA
TX-002	5' -	YTC ACG RCA YGA GCT GAC GAC
TX-003	5' -	Fosfato-GCT CCA GCT CTA CAT CCT AGT CAG GACddc
TX-004	5' -	ATA GGT CCT GAC TAG GAT GTA GAC
TX-005	5' -	Fosfato-GCT CCA GAC TAG CAT CCG CTG ACT TGaddc
TX-006	5' -	AAT GTC AAG TCA GCG GAT GCT AGT
TX-007	5' -	Biotina-GGT CCT GAC TAG GAT GTA GAC
TX-008	5' -	Biotina-GTC AAG TCA GCG GAT GCT AGT
TX-009	5' -	Biotina-GGA TTA GAW ACC CBG GTA GTC
TX-010	5' -	ATA GGT CCT GAC TAG GAT GTA GAC CTG GAG
TX-011	5'	AAT GTC AAG TCA GCG GAT GCT AGT CTG GAG
TX-012	5' -	Fosfato-CTC CAG GTC TAC ATC CTA GTC AGG ACddc
TX-013	5' -	ATA GGT CCT GAC TAG GAT GTA GAC CTG GAG
TX-014	5' -	Fosfato-CTC CAG ACT AGC ATC CGC TGA CTT Gaddc
TX-015	5'	AAT GTC AAG TCA GCG GAT GCT AGT CTG GAG
TX-111	5'	Biotina-CTC GAT AGT CAC GGT CCT GAC TAG GAT GTA GAC
TX-121	5'	Biotina-CAG GTT GCA ACG GTC AAG TCA GCG GAT GCT AGT

Fig. 12

Análisis SARD de Población Definida

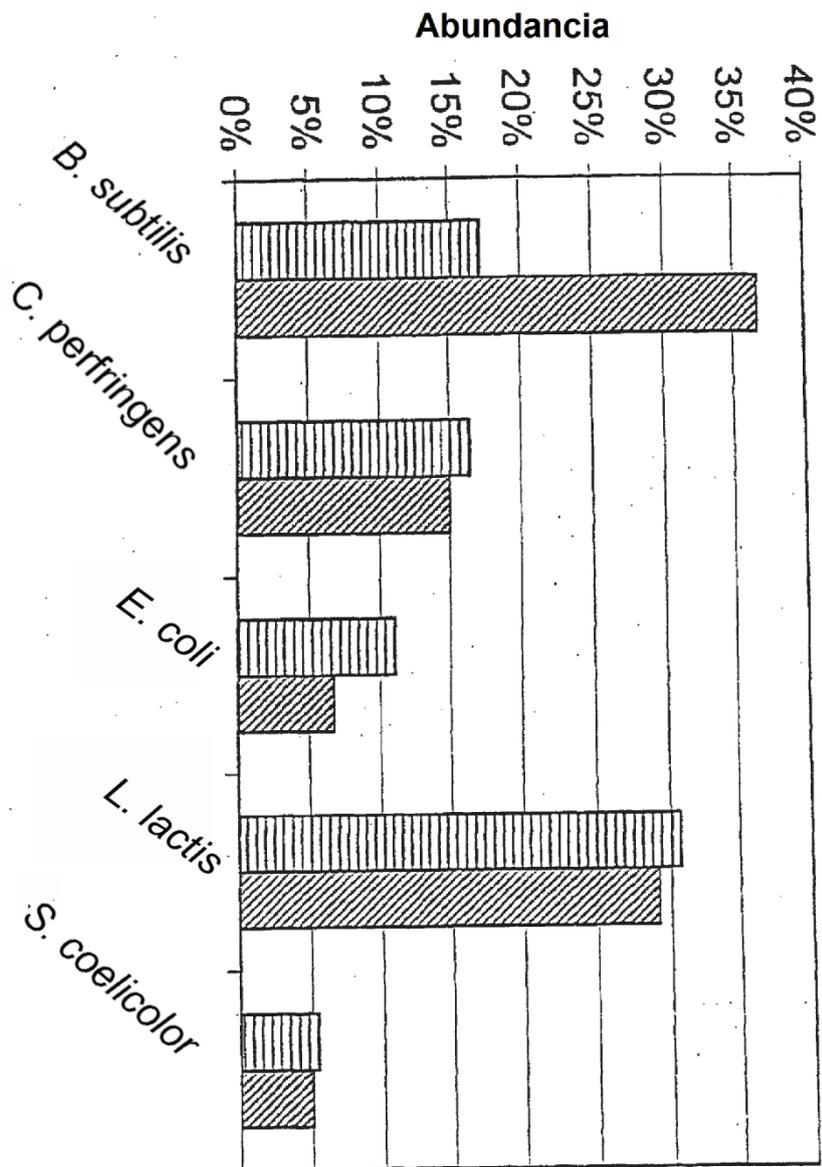


Figura 14

ACGATGAGCACTAGCT (1)	CCTGTGTGACGAAGCT (1)	CTTCTGTGGCGAAGCT (1)
ACGATGAGTACTAGCT (1)	CCTTGGTAACGAAGCT (1)	GATCCGTGCCGTAGCT (1)
ACGATGATGACTAGCT (2)	CCTTGGTACCGAAGCT (1)	GCTCTGTGCCGAAGCT (1)
ACGATGGATGCTAGCT (1)	CGCCAGTGCCGTAGCT (1)	GCTGGGTGCCCAAGCT (1)
ATGCTAGTCTGGAGCT (6)	CGCCTGTGCCGTAGCT (2)	GGTCCGTGCCGCAGCT (1)
ATGGCTGTCGTCAGCT (50)	CGTCCGTGCCGAAGCT (2)	GGTGCTCTTCGGAGCT (2)
ATGGTTGTCGTCAGCT (1)	CGTCCGTGCCGCAGCT (1)	GTAAACGATGGAAGCT (1)
ATTCCGTGCCGTAGCT (1)	CGTCCGTGCCGCAGCT (3)	GTGGCTGTCGTCAGCT (2)
CACTAGTGGCCGAGCT (2)	CTCCCGTGCCGCAGCT (1)	GTTCCGTGCCGAAGCT (2)
CCCCCGTGCCGAAGCT (1)	CTCCCGTGCCGGAGCT (1)	GTTCCGTGCCGCAGCT (3)
CCCCCGTGCCGCAGCT (1)	CTCCGGTGCCGCAGCT (1)	TATCAGTGGCCGAGCT (1)
CCCCCTTCTCCAGCT (1)	CTCCTGTGCCGAAGCT (1)	TCTCCGTGCCGCAGCT (1)
CCCCGGTGCCGCAGCT (1)	CTCCTGTGCCGCAGCT (2)	TCTCTGTGCCGCAGCT (2)
CCGGGTAGTCCAGCT (5)	CTGCCGTGCCGAAGCT (9)	TCTCTGTGCCGTAGCT (1)
CCTCCGTGCCGAAGCT (3)	CTGCTGTGCCGAAGCT (2)	TGTCGGTGCCGTAGCT (1)
CCTCCGTGCCGCAGCT (2)	CTGTGCTGCCGAAGCT (1)	TTTCCGTGCCGCAGCT (2)
CCTCCGTGCTGCAGCT (1)	CTTCAGTATCGAAGCT (1)	
CCTCGGCGCCGCAGCT (2)	CTTCCGCGCCGGAGCT (2)	
CCTCGGTGCCGCAGCT (1)	CTTCCGTGCCGCAGCT (3)	
CCTCGGTGTGCGAGCT (2)	CTTCCGTGCCGGAGCT (1)	
CCTGGGTGCCGCAGCT (2)	CTTCCGTGCCGCAGCT (1)	

Fig. 15

ACGATGATAACTAGCT (2)	CCGGGTAGTCCCAGCT (4)	CTCCAGTGCCGCAGCT (1)	GATCCGTGCCGCAGCT (2)
ACGATGATGACTAGCT (1)	CCGGGTAGTCCTAGCT (1)	CTCCCGTGCCACAGCT (1)	GCTCCGTGCCGAAGCT (1)
ACGATGGGCACTAGCT (1)	CCTCCGTGCCGAAGCT (2)	CTCCCGTGCCGCAGCT (2)	GCTCCGTGCCGTAGCT (1)
ACGGCTGTCGTCAGCT (2)	CCTCCGTGCCGCAGCT (2)	CTCCCGTGCCGGAGCT (3)	GCTCTGTGCCGAAGCT (2)
ACTACGAGCGCAAGCT (1)	CCTCCGTGCCGGAGCT (1)	CTCCCGTGCCGCAGCT (1)	GCTCTGTGCCGTAGCT (1)
ACTTAATGCGTTAGCT (1)	CCTCGTAAGGGGAGCT (1)	CTCCTGTGCCGCAGCT (1)	GGGCTTGTGTCGTCAGCT (1)
ATGCTAGTCTGGAGCT (8)	CCTGGGTGCCGCAGCT (1)	CTGCCGCGCCGGAGCT (1)	GTAAACGATGGAAGCT (1)
ATGGCTCTCGTCAGCT (1)	CCTGGTAGTCCCAGCT (2)	CTGCCGTGCCGAAGCT (5)	GTTCCGTGCCGCAGCT (2)
ATGGCTGTGCCAGCT (3)	CCTGGTAGTCCTAGCT (1)	CTGCCGTGCCTAAGCT (1)	GTTCCGTGCCGTAGCT (2)
ATGGCTGTCGTCAGCT (70)	CCTTAGTAACGCAGCT (1)	CTGCCGTGCCGCAGCT (1)	GTTCTGTGCCGCAGCT (2)
ATGETTGTGTCGTCAGCT (2)	CCTTGGTAACGAAGCT (1)	CTGCTGTGCCGAAGCT (7)	TCTCAGCACACGAGCT (2)
ATGTAGACCTGGAGCT (12)	CGCCAGTGCCGAAGCT (1)	CTGTGCTGCCGAAGCT (1)	TCTCAGTAACGTAGCT (1)
ATTCCGTGCCGCAGCT (6)	CGCCGCTGCCGCAGCT (2)	CTTCCGCGCCGGAGCT (7)	TCTCCGTGCCGCAGCT (2)
ATTCCGTGCCGTAGCT (1)	CGCCTGTGCCGTAGCT (2)	CTTCCGTGCCGAAGCT (2)	TCTCTGTGCCGCAGCT (4)
CACAAGCGGTGGAGCT (1)	CGCTCGTGSCGAAGCT (1)	CTTCCGTGCCGCAGCT (6)	TGGACGTTGCCGGAGCT (2)
CACTAGTGGCGCAGCT (4)	CGGAGGCGTCGTAGCT (1)	CTTCCGTGCCGGAGCT (2)	TTTCCGTGCCGGAGCT (1)
CATCCGTGCCGAAGCT (1)	CGTCAGTGTCCGCAGCT (2)	CTTCCGTGCCGCAGCT (1)	
CCCCAGGGCCCAAGCT (1)	CGTCCGTGCCGAAGCT (7)	CTTCCGTGTCCGCAGCT (1)	
CCCCGGTGCCGCAGCT (1)	CGTCCGTGCCGCAGCT (1)	CTTGGGTGCCGCAGCT (2)	
CCCCGGTGCCGGAGCT (1)	CGTCCGTGCCGGAGCT (1)	CTTTAGTAACGCAGCT (2)	
CCGCGGTGCCGTAGCT (1)	CGTCCGTGCCGCAGCT (2)	GACCCGCAAGGGAGCT (1)	

Fig. 16

Perfil de diversidad de etiquetas SARD Wy-1

