

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 609 479**

51 Int. Cl.:

G06F 19/22 (2011.01)

G06F 19/28 (2011.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **09.11.2005 PCT/CH2005/000664**

87 Fecha y número de publicación internacional: **18.05.2007 WO07053962**

96 Fecha de presentación y número de la solicitud europea: **09.11.2005 E 05798706 (7)**

97 Fecha y número de publicación de la concesión europea: **19.10.2016 EP 1946230**

54 Título: **Procedimiento implementado por ordenador y sistema informático para la identificación de organismos**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
20.04.2017

73 Titular/es:
**SMARTGENE GMBH (100.0%)
INDUSTRIESTRASSE 16
6300 ZUG, CH**

72 Inventor/es:
EMLER, STEFAN

74 Agente/Representante:
DE ELZABURU MÁRQUEZ, Alberto

ES 2 609 479 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

Procedimiento implementado por ordenador y sistema informático para la identificación de organismos

5 La presente invención se refiere a un procedimiento implementado por ordenador y un sistema informático para la identificación de organismos. Específicamente, la presente invención se refiere a un procedimiento implementado por ordenador y un sistema informático para la identificación de tipos de organismos a partir de una secuencia genética diana. La presente invención se refiere también a un producto de programa de ordenador para controlar el sistema basado en ordenador de manera que el sistema ejecute el procedimiento de identificación de tipos de organismos a partir de la
10 secuencia genética diana.

Los diagnósticos médicos dependen cada vez más del análisis de dianas genéticas de seres humanos o microorganismos. Típicamente, este análisis se basa en la comparación de una secuencia genética diana individual con
15 secuencias de referencia obtenidas de una base de datos de referencia. La secuencia de referencia con una coincidencia más estrecha es recuperada desde la base de datos de referencia. De esta manera, para la identificación de tipos de organismos a partir de una secuencia genética diana, los procedimientos y sistemas convencionales comparan y recuperan las secuencias de referencia con respecto a su similitud con la secuencia diana. Convencionalmente, la similitud se determina a partir de coincidencias globales a lo largo del segmento común más largo de las secuencias diana y de referencia. Se tiene en cuenta el número de errores de coincidencia en un segmento independientemente de
20 sus posiciones. Por ejemplo, en microbiología, el proyecto de base de datos ribosómica (Ribosomal Database Project, RDP-II), accesible en <http://rdp.cme.msu.edu/>, la diferenciación ribosómica de microorganismos (Ribosomal Differentiation of Microorganisms, RIDOM™), accesible en <http://www.ridom.de/>, y la base de datos europea de ARN ribosómico, accesible en <http://www.psb.ugent.be/rRNA/>, ofrecen servicios de clasificación basados en secuencias rADN 16S (ácido desoxirribonucleico ribosómico) bacterianas. En estos sistemas, se usan procedimientos que incluyen las etapas siguientes: (i) comparar la secuencia diana con una base de datos de secuencias de referencia de ADNr 16S, produciendo un conjunto de secuencias estrechamente coincidentes; (ii) construir un alineamiento múltiple de secuencias (Multiple Sequence Alignment, MSA) a partir de la secuencia diana y este conjunto de secuencias de referencia; (iii) extraer una matriz de distancias a partir de la MSA, y usar esta matriz para construir un árbol evolutivo; (iv) colocar la secuencia de consulta en el árbol, y visualizar las relaciones resultantes. Sin embargo, estos sistemas no discriminan
30 entre diferencias entre secuencias que podrían ser triviales en origen, por ejemplo, debido a errores de secuenciación o variaciones biológicamente no importantes, y diferencias encontradas en posiciones que se sabe que son diagnósticas de diferencias entre cepas o entre especies. Por consiguiente, en estos sistemas, los errores de secuenciación o las mutaciones aleatorias en regiones que no son significativas tienen el mismo impacto sobre la clasificación de secuencias que los errores de coincidencia en ciertas regiones variables relacionadas con el género o la especie, por ejemplo. Debido a que las posiciones de estas regiones variables no se conocen antes de la identificación del tipo de organismo (por ejemplo, género, especie, sub-tipo, variante o clado) de una muestra determinada, la metodología basada en la comparación de secuencias depende mucho del usuario y requiere un nivel de experiencia que no se encuentra fácilmente en los laboratorios de diagnóstico.

40 Un objeto de la presente invención es proporcionar un procedimiento implementado por ordenador y un sistema informático para la identificación de tipos de organismos a partir de una secuencia genética diana, en el que el tipo de organismo incluye entidad, género, especie, subtipo, variante y/o clado asociados con el organismo, cuyos sistema y procedimiento no tengan las desventajas de la técnica anterior. En particular, un objeto de la presente invención es proporcionar un procedimiento implementado por ordenador y un sistema informático para la identificación de tipos de
45 organismos a partir de una secuencia genética diana, cuyos sistema y procedimiento discriminen entre diferencias triviales y significativas entre secuencias. Las referencias realizadas en la presente memoria al término "organismo" o "tipo de organismo" se entenderá que incluyen los términos "entidad", "género", "especie", "sub-tipo", "grupo", "cepa", "variante" y/o "clado" en el contexto de la clasificación taxonómica.

50 Según la presente invención, los objetos anteriores se consiguen particularmente mediante las características de las reivindicaciones independientes. Además, a partir de las reivindicaciones dependientes y la descripción se desprenden realizaciones ventajosas adicionales.

Según la presente invención, los objetos indicados anteriormente se consiguen particularmente en el sentido de que, para
55 identificar los tipos de organismos a partir de una secuencia genética diana, seleccionado automáticamente desde una base de datos hay un perfil de secuencias seleccionado que tiene una correlación más alta con la secuencia genética diana en base a información específica de la posición del perfil específico del tipo de organismo relacionado. El perfil de secuencias se selecciona de entre una pluralidad de perfiles específicos del tipo de organismo en la base de datos, en el que cada perfil define regiones de secuencias informativas para diferenciar los tipos de organismos individuales e incluye dicha información específica de la posición relacionada con más de una secuencia de referencia asociada de un tipo de organismo específico. Preferiblemente, los perfiles específicos del tipo incluyen perfiles específicos del género o específicos del grupo; además, los perfiles específicos del tipo pueden incluir perfiles específicos de la especie, específicos
60

- del sub-tipo, específicos de la variante y/o específicos del clado. Las secuencias de referencia, asociadas con el perfil de secuencias seleccionado, son recuperadas automáticamente desde la base de datos. La secuencia genética diana es comparada automáticamente con las secuencias de referencia y los resultados de la comparación, relacionados con las regiones de secuencias informativas, son ponderados de manera automática usando factores de ponderación asociados con las regiones de secuencias informativas definidas en el perfil de secuencias seleccionado. Posteriormente, a partir de las secuencias de referencia, se determina una secuencia de referencia específica del tipo del organismo que tiene una mejor coincidencia con la secuencia genética diana, en el que la mejor coincidencia se determina en base a los resultados de la comparación ponderados para las regiones de secuencias informativas. La secuencia de referencia específica del tipo que tiene la mejor coincidencia con la secuencia genética diana, considerando los resultados ponderados de la comparación, se selecciona de manera automática o se establece como una entrada superior en una lista ordenada. La ponderación de los resultados de comparación para las regiones de secuencias informativas hace que sea posible identificar el tipo de organismo a partir de la secuencia genética diana, mientras se discrimina entre diferencias triviales e importantes entre secuencias. Los resultados obtenidos mediante una búsqueda de perfiles y el alineamiento ponderado proporcionarán una medida que refleja la asignación correcta del tipo de organismo en bacteriología, micología y virología. Por consiguiente, se mejora la asignación de tipos de organismos, por ejemplo especies bacterianas y fúngicas o subtipos virales. Los tipos de organismos se asignan no sólo en base a criterios estadísticos, sino también en base a perfiles biológicamente relevantes. Por consiguiente, se derivan resultados más fiables para el análisis de secuencias en una disposición rutinaria de fácil uso. Generalmente, el tiempo necesario para producir resultados se acorta y el tratamiento de los pacientes se beneficiará de unos resultados más rápidos y precisos.
- En una realización, los resultados de la comparación incluyen un número de diferencias y/o correspondencias en los códigos de nucleótidos de cada una de las secuencias de referencia en comparación con la secuencia genética diana (códigos de nucleótidos que incluyen también códigos IUPAC (International Union of Pure and Applied Chemistry, unión internacional de códigos de química pura y aplicada)). La ponderación los resultados de la comparación incluye determinar para cada secuencia de referencia un número ponderado de diferencias y/o correspondencias multiplicando con un factor de ponderación el número de diferencias y/o correspondencias relacionadas con las regiones de secuencias informativas. Además, hay almacenada una lista de las secuencias de referencia, ordenada por el número ponderado de diferencias y/o correspondencias de la secuencia de referencia respectiva.
- En una realización preferida, la secuencia genética diana y las secuencias de referencia asociadas con el perfil de secuencias seleccionado se evalúan para nuevas regiones de secuencias informativas para el perfil seleccionado. Además, el perfil seleccionado es adaptado mediante el almacenamiento de una nueva región de secuencia informativa como una parte del perfil de secuencias seleccionado. El refinamiento del perfil de secuencias con regiones de secuencias informativas recientemente identificadas hace que sea posible considerar aspectos evolutivos de los organismos, por ejemplo, relaciones evolutivas entre especies y cepas. La adaptación continua de los perfiles de secuencias ayuda a ajustar las anotaciones filogenéticas y, en última instancia, taxonómicas y, de esta manera, proporcionará información importante para los microbiólogos y los médicos con relación a la patogenicidad y la epidemiología de microorganismos desconocidos o clasificados erróneamente.
- En una realización, la evaluación de la secuencia genética diana y las secuencias de referencia incluye el alineamiento de la secuencia genética diana y las secuencias de referencia asociadas con el perfil de secuencias seleccionado, y la identificación de nuevas regiones de secuencias informativas. Las nuevas regiones de secuencias informativas se determinan mediante la identificación de los códigos de nucleótidos correspondientes a una misma posición secuencial en al menos un número definido de la secuencia genética diana y las secuencias de referencia.
- Preferiblemente, los perfiles de secuencias específicas del tipo de organismo almacenados en la base de datos se determinan mediante el alineamiento de las secuencias genéticas específicas del tipo, mediante la creación de secuencias de consenso para cada tipo de organismo, mediante la identificación de las regiones informativas que permiten la diferenciación de los tipos de organismos individuales, mediante la adición de información de anotación de secuencia, y mediante la definición de perfiles de secuencias específicas del tipo en base a las regiones informativas.
- En una realización, los perfiles de secuencias específicas del tipo almacenados en la base de datos incluyen perfiles de secuencias específicas del género o específicas del grupo. Los perfiles específicos del género o específicos del grupo se determinan mediante el alineamiento de secuencias genéticas específicas del género o específicas del grupo, mediante la creación de secuencias de consenso para cada tipo de organismo, mediante la identificación de las regiones informativas que permiten la diferenciación de los organismos individuales, y mediante la definición de perfiles de secuencias específicas del género o específicas del grupo en base a las regiones informativas.
- En una realización adicional, la secuencia genética diana es corregida en base al perfil de secuencias seleccionado mediante la comparación de la secuencia genética diana con las secuencias de referencia asociadas con el perfil de secuencias seleccionado. Para las diferencias de códigos de nucleótidos situadas en las regiones de secuencias informativas, se evalúa si las diferencias indican o no otro tipo de organismo. La adaptación del perfil de secuencias

seleccionado se inicia para aquellas diferencias evaluadas como indicativas de otro tipo de organismo. La corrección automática basada en el perfil de secuencias seleccionado hace que sea posible corregir la secuencia genética diana mientras se discrimina entre diferencias triviales e importantes entre secuencias.

5 Preferiblemente, la secuencia genética diana es recibida por un servidor desde un usuario a través de una red de telecomunicaciones. Además, el tipo organismo de la secuencia genética diana, definido por la secuencia de referencia específica del tipo de organismo, es transmitido por el servidor a través de la red de telecomunicaciones a una interfaz de usuario. La implementación del procedimiento en un servidor basado en la red hace que sea posible proporcionar de manera eficiente (en términos de rendimiento y costes financieros) la identificación automática de los tipos de organismos a partir de una secuencia genética diana como un servicio centralizado, disponible para una pluralidad de usuarios conectados a la red de telecomunicaciones. El uso de una tecnología basada en servidor para identificar tipos de organismos a partir de una secuencia genética diana hace posible que un usuario use su propio equipo informático sin tener que instalar ningún software o hardware. En la base de datos en red, los perfiles específicos del tipo pueden ser añadidos y mejorados continuamente en base a secuencias diana suministradas a través de la red por los usuarios. 10 Además, la base de datos de secuencias de referencia, la aplicación de software, así como cualquier herramienta de software, pueden ser actualizadas en línea sin ninguna molestia para los usuarios. Además, el servidor basado en red puede permitir el intercambio y la compartición de datos entre institutos expertos alejados, así como la evaluación de entradas de base de datos que representan tipos de organismos, por ejemplo, especies bacterianas y fúngicas o subtipos víricos, con respecto a su clasificación taxonómica. De esta manera, el servidor basado en la red hace posible que los expertos re-evalúen y validen conjuntos de datos de referencia para bacterias, micobacterias, hongos y virus. 15 20

Además de un procedimiento implementado por ordenador y un sistema informático para la identificación de tipos de organismos a partir de una secuencia genética diana, la presente invención se refiere también a un producto de programa de ordenador que incluye medios de código de programa de ordenador para controlar uno o más procesadores del sistema informático de manera que el sistema ejecute el procedimiento de identificación de los tipos de organismos a partir de una secuencia genética diana. En particular, un producto de programa de ordenador que incluye unos medios legibles por ordenador que contienen los medios de código de programa de ordenador (por ejemplo, módulos de software programados, tal como se describe más detalladamente más adelante). 25

30 La presente invención se explicará más detalladamente, a modo de ejemplo, con referencia a los dibujos, en los que:

La Figura 1 muestra un diagrama de bloques que ilustra esquemáticamente una configuración ejemplar de un sistema informático para poner en práctica las realizaciones de la presente invención, en el que dicha configuración comprende un servidor con una base de datos, y dicha configuración está conectada a un terminal de entrada de datos a través de una red de telecomunicaciones. 35

La Figura 2 muestra un diagrama de flujo que ilustra un ejemplo de una secuencia de etapas para el establecimiento de un perfil de secuencias para un organismo diana.

La Figura 3 muestra un diagrama de flujo que ilustra un ejemplo de una secuencia de etapas ejecutadas según la presente invención para la identificación de un tipo de organismo a partir de una secuencia genética diana. 40

La Figura 4 muestra un diagrama de flujo que ilustra un ejemplo de una secuencia de etapas ejecutadas según la presente invención para la adaptación de los perfiles de secuencias en base a una secuencia genética diana.

En la Figura 1, el número de referencia 1 se refiere a un terminal de entrada de datos. Tal como se ilustra en la Figura 1, el terminal 1 de entrada de datos incluye un ordenador 11 personal con un teclado 12 y un monitor 13 de visualización. Tal como se ilustra esquemáticamente, en una realización, el ordenador 11 personal incluye un módulo 14 de usuario implementado como un módulo de software programado, por ejemplo una miniaplicación ("applet") de programa ejecutable que se descarga desde el servidor 3 a través de la red 2 de telecomunicaciones. 45

Conectado al ordenador 11 personal, hay un secuenciador 5 convencional, que proporciona al ordenador 11 personal datos de secuencia de fragmentos de ADN (ácido desoxirribonucleico). Por ejemplo, los datos de secuencia de fragmentos incluyen señales de secuencia e información asociada (por ejemplo, valores de pico) de los fragmentos de ADN, en el que cada señal de secuencia incluye señales de los cuatro tipos de nucleótidos adenina, citosina, guanina y timina (A, C, G, T). En general, las expresiones "secuencia genética", "secuencia diana" o "secuencia de referencia" se usan en la presente memoria para hacer referencia a una secuencia de códigos de nucleótidos, es decir, una secuencia de 55 códigos, cada uno de los cuales representa uno de los tipos de nucleótidos. Las secuencias están relacionadas con bacterias, hongos, hongos microscópicos y virus, por ejemplo.

Tal como se ilustra en la Figura 1, el terminal 1 de entrada de datos está conectado al servidor 3 a través de una red 2 de telecomunicaciones. Preferiblemente, la red 2 de telecomunicaciones incluye Internet y/o una Intranet, haciendo que el servidor 3 sea accesible como un servidor web a través de la World Wide Web o dentro de una red IP diferente, respectivamente. La red 2 de telecomunicaciones puede incluir también otra red fija, tal como una red de área local (Local Area Network, LAN) o una red digital de servicios integrados (Integrated Services Digital Network, ISDN) y/o una red 60

inalámbrica, tal como una red de radio móvil (por ejemplo, sistema global para comunicaciones móviles (Global System for Mobile communication, GSM) o sistema universal de telefonía móvil (Universal Mobile Telephone System, UMTS)), o una red de área local inalámbrica (Wireless Local Area Network, WLAN). En una realización local, el terminal 1 de entrada de datos está conectado directamente al servidor 3, o el terminal 1 de entrada de datos y el servidor 3 están implementados en el mismo ordenador, por ejemplo, un PC.

Tal como se ilustra esquemáticamente en la Figura 1, el servidor 3 está conectado a la base de datos 4. El servidor 3 puede incluir uno o más ordenadores, cada uno con uno o más procesadores. La base de datos 4 puede estar implementada en un ordenador compartido con el servidor 3 o en un ordenador separado.

El servidor 3 incluye diferentes módulos funcionales, concretamente, un módulo 30 de comunicaciones, un módulo 31 de aplicación, un módulo 32 de selección de perfil, un módulo 33 de recuperación, un módulo 34 de comparación, un módulo 35 de determinación de tipo, un módulo 36 de adaptación de perfil, un módulo 37 de creación de perfil, y un módulo 38 de corrección. El módulo 30 de comunicaciones incluye elementos de hardware y software convencionales configurados para intercambiar datos a través de la red 2 de telecomunicaciones con una pluralidad de terminales 1 de entrada de datos. El módulo 31 de aplicación es un módulo de software programado configurado para proporcionar una interfaz de usuario a los usuarios del terminal 1 de entrada de datos. Preferiblemente, la interfaz de usuario es proporcionada a través de un navegador de Internet convencional tal como Microsoft Explorer o Mozilla. De manera alternativa, la interfaz de usuario es generada por el módulo 14 de usuario. Por ejemplo, el módulo 31 de aplicación transmite una copia del módulo 14 de usuario a través de la red 2 de telecomunicaciones al ordenador 11 personal del usuario. El módulo 14 de usuario es instalado y activado en el ordenador 11 personal. Cuando está activo, el módulo 14 de usuario controla un procesador del ordenador 11 personal de manera que genera la interfaz de usuario en pantalla 13. El módulo 32 de selección de perfil, el módulo 33 de recuperación, el módulo 34 de comparación, el módulo 35 de determinación de tipo, el módulo 36 de adaptación de perfil, el módulo 37 de creación de perfil y el módulo 38 de corrección son módulos de software programado que se ejecutan en un ordenador del servidor 3.

Tal como se ilustra esquemáticamente en la Figura 1, la base de datos 4 incluye secuencias 41 de referencia y perfiles 42 de secuencia asociados para diferentes tipos de organismos. Los perfiles 42 se almacenan y se asignan a las secuencias 41 de referencia. Típicamente, más de una secuencia 41 de referencia es asignada a un perfil 42 de secuencias. Los perfiles 42 definen regiones de secuencias informativas para la diferenciación de organismos individuales, es decir, los perfiles 42 incluyen información específica de la posición relacionada con una o más secuencias 41 de referencia asociadas de un tipo específico de organismo, es decir, un género, especie, sub-tipo, variante y/o clado específicos. La información específica de la posición es asignada a una posición específica o a un intervalo de múltiples posiciones en una secuencia. El intervalo está definido, por ejemplo, por una posición inicial y una posición final en la secuencia de referencia o por una posición inicial y una longitud (es decir, número de posiciones de la secuencia). De manera alternativa, la información específica de la posición puede ser representada también por medio de modelos ocultos de Markov (Hidden Markov Models, HMM). Aunque no son idénticos, los perfiles 42 y HMM son suficientemente similares de manera que pueden ser usados de manera intercambiable en el presente contexto. Los perfiles 42 de secuencia resumen la información contenida en un alineamiento múltiple de secuencias: el documento de anotaciones específicas de la posición, para cada posición (columna) en el alineamiento, la probabilidad de encontrar un residuo determinado (patrón) en esa posición, el coste de abrir un hueco en el alineamiento antes o después de esta posición ("coste de hueco abierto"), y las probabilidades de encontrar residuos particulares inmediatamente antes o después de la posición actual. Tal como se ilustra en la Tabla 1, en una realización, los perfiles 42 están provistos de información adicional, por ejemplo, la ponderación asignada a las posiciones individuales en el cálculo de las puntuaciones de alineamiento, o los parámetros estructurales asociados con intervalos de posiciones.

Perfil de secuencia				Secuencia(s) genética(s) de referencia asignada(s)
Identificador de perfil	Información específica de la posición			
	Región/ especificación de posición	Anotación	Factor de ponderación	

Tabla 1

Los perfiles 42 iniciales son establecidos y almacenados por el módulo 37 de creación de perfil, tal como se describirá a continuación con referencia a la Figura 2. Un perfeccionamiento y una adaptación adicionales de los perfiles son proporcionados por el módulo 36 de adaptación de perfil, tal como se describirá más adelante con referencia a la Figura 4.

Tal como se ilustra en la Figura 2, en la etapa S20, se especifica un organismo diana o grupo de organismos diana. Preferiblemente, el organismo diana está definido por un género o especie específicos. Sin embargo, una persona con conocimientos en la materia entenderá que el organismo diana puede estar definido además por un sub-tipo, variante y/o

clado. La diana es especificada por un usuario a través de una interfaz de usuario visualizada en la pantalla 13 o en otro terminal de entrada de datos conectado al servidor 3. Por ejemplo, la diana se selecciona de una lista desplegable o es introducida en un campo de entrada de datos. De manera alternativa, la diana puede ser seleccionada también automáticamente por el módulo 37 de creación de perfil desde una lista de posibles dianas.

5 En la etapa S21, se seleccionan secuencias de tipo validadas para la diana especificada en la etapa S20. Por ejemplo, las secuencias genéticas validadas son seleccionadas automáticamente por el módulo 37 de creación de perfil a partir de las secuencias 41 de referencia en la base de datos 4. Sin embargo, al menos inicialmente, las secuencias genéticas validadas son recuperadas desde una base de datos de referencia validada o son seleccionadas e introducidas por una
10 persona experta usando el terminal 1 de entrada de datos. Las secuencias de tipo validadas para la diana especificada cubren todas las posiciones variables conocidas, es decir, todas las regiones de secuencias informativas (posiciones) que se sabe que son diagnósticas de diferencias entre cepas o entre especies y, de esta manera, indicativas de tipos de organismos (incluyendo género, especie, sub-tipo, variante y/o clado).

15 En la etapa S22, usando el módulo 37 de creación de perfil, se genera una semilla MSA (alineamiento múltiple de secuencias) a partir de las secuencias de tipo validadas seleccionadas en la etapa S21. En particular, usando el módulo 37 de creación de perfil, se alinean las secuencias de tipo y se crea una secuencia de consenso para el tipo de organismo respectivo.

20 En la etapa S23, se identifican regiones informativas que permiten la diferenciación de tipos de organismos individuales. Usando el módulo 37 de creación de perfil, el MSA generado en la etapa S22 es provisto de anotaciones para estructuras secundarias (regiones de emparejamiento 3' y 5'), y para posiciones que se sabe que son diagnósticas de diferencias entre cepas o entre especies en el organismo diana, es decir, posiciones que se sabe que son indicativas de los tipos de organismos (incluyendo género, especie, sub-tipo, variante y/o clado).

25 En la etapa S24, el módulo 37 de creación de perfil convierte el MSA anotado en uno o más perfiles anotados y almacena estos perfiles 42 en la base de datos 4.

30 En la etapa S25, usando el módulo 37 de creación de perfil, los perfiles 42 almacenados en la etapa S24 son calibrados de manera iterativa. La calibración iterativa se consigue usando el perfil o los perfiles para buscar una colección de secuencias de referencia (validadas por expertos) que se sabe que pertenecen o que no pertenecen al tipo de organismo respectivo, es decir, género, especie, sub-tipo, variante y/o clado.

35 En la etapa S26, usando el módulo 37 de creación de perfil, la anotación de los perfiles 42 es enriquecida mediante la inclusión de posiciones que discriminan entre el organismo diana y otros géneros.

40 En la etapa S27, los perfiles anotados son validados por personas expertas en la materia y por medio de estadísticas de secuencias de muestra disponibles. Para fines de validación, los perfiles 42 almacenados en la base de datos 4 se ponen a disposición de los expertos a través del servidor 3 y la red 2 de telecomunicaciones. Por ejemplo, los perfiles validados son provistos de un indicador, o un certificado o firma electrónicos en la base de datos 4.

En los párrafos siguientes, la identificación de un tipo de organismo a partir de una secuencia genética diana se describe con referencia a la Figura 3.

45 En la etapa S1, una secuencia genética diana es recibida por el módulo 30 de comunicaciones a través de la red 2 de telecomunicaciones. La secuencia genética diana es proporcionada por el terminal 1 de entrada de datos. Por ejemplo, la secuencia genética diana está almacenada en el ordenador 11 personal o es generada a partir de los datos de secuencia de fragmentos de ADN proporcionados por un secuenciador 5. Preferiblemente, la secuencia genética diana es definida por un usuario a través de una interfaz de usuario visualizada en la pantalla 13 por el módulo 31 de aplicación o por el
50 módulo 14 de usuario. Posteriormente, el procedimiento para identificar el tipo de organismo se inicia automáticamente. De manera alternativa, el procedimiento es iniciado por el usuario al activar un elemento de control, tal como un pulsador gráfico, en la interfaz de usuario.

55 En la etapa S2, el módulo 32 de selección de perfil determina en la base de datos 4 el perfil 42 de secuencia que tiene la correlación más alta con la secuencia genética diana recibida en la etapa S1. El grado de correlación entre la secuencia genética diana y los perfiles 42 de secuencia se determina en base a la información específica de la posición contenida en los perfiles 42 de secuencias, es decir, el módulo 32 de selección de perfil usa las anotaciones de perfil en las regiones de secuencias informativas para seleccionar el perfil 42 que tiene la mejor coincidencia con la secuencia genética diana. Preferiblemente, el perfil de secuencias con la mejor coincidencia se determina aplicando para cada perfil sus factores de ponderación específicos de la posición a las desviaciones y/o las correspondencias de la secuencia diana con relación al
60 perfil.

En la etapa S3, el módulo 33 de recuperación carga desde la base de datos 4 las secuencias 41 genéticas de referencia asociadas con el perfil 42 de secuencia seleccionado en la etapa S2.

5 En la etapa S4, el módulo 34 de comparación compara la secuencia genética diana, recibida en la etapa S1, con una de las secuencias genéticas de referencia, recuperada en la etapa S3. Además, el módulo 34 de comparación pondera los resultados de comparación con los factores de ponderación asociados con el perfil de secuencias, particularmente, los factores de ponderación asociados con las regiones de secuencias informativas. Por consiguiente, los resultados de comparación relacionados con una primera región de secuencia pueden ser ponderados con otro factor de ponderación distinto al de los resultados de comparación relacionados con otra segunda región de secuencia. De esta manera, el
10 módulo 34 de comparación pondera el número de diferencias y/o el número de correspondencias, entre la secuencia genética diana y la secuencia genética de referencia respectiva, usando factores de ponderación asociados con las regiones de secuencias informativas descritas en el perfil 42.

15 En la etapa S6, el módulo 34 de comparación almacena una puntuación, indicativa del nivel de coincidencia, asignada a la secuencia genética de referencia respectiva. La puntuación se basa en los resultados de comparación ponderados. Por ejemplo, el módulo 34 de comparación almacena una puntuación basada en el número ponderado de diferencias y/o correspondencias.

20 En la etapa S7, el módulo 31 de aplicación comprueba si hay o no secuencias genéticas de referencia adicionales asignadas al perfil seleccionado que deben ser procesadas. Si hay más secuencias de referencia a ser procesadas, el procesamiento continúa en la etapa S4. De lo contrario, si se han procesado todas las secuencias de referencia asignadas al perfil seleccionado, el procesamiento continúa en la etapa S8.

25 En la etapa S8 opcional, el módulo 35 de determinación de tipo genera una lista completa o parcial de las secuencias de referencia asignadas al perfil seleccionado. La lista está ordenada por la puntuación asignada a las secuencias de referencia. Por ejemplo, la lista (con sus entradas y puntuaciones asignadas) es transmitida a través de la red 2 de telecomunicaciones al terminal 1 de entrada de datos, donde se muestra al usuario en la pantalla 13.

30 En la etapa S9, el módulo 35 de determinación de tipo determina la secuencia genética de referencia específica del tipo que tiene la mejor coincidencia con la secuencia genética diana. El módulo 35 de determinación de tipo determina la secuencia genética de referencia específica del tipo en base a las puntuaciones asignadas, es decir, los resultados de comparación ponderados, por ejemplo, el número ponderado de diferencias y/o correspondencias asignado a las secuencias de referencia recuperadas en la etapa S3. Por ejemplo, la secuencia genética de referencia específica del tipo se define por el número ponderado más bajo de diferencias y/o el número ponderado más alto de correspondencias. La
35 información del tipo asociada con la secuencia genética de referencia específica del tipo es seleccionada para definir el tipo de organismo de la secuencia genética diana. De esta manera, el tipo de organismo de la secuencia genética diana es definido por el género, especie, sub-tipo, variante y/o clado asociados con la secuencia genética de referencia específica del tipo. Preferiblemente, el tipo de organismo y su puntuación asignada son transmitidos por el módulo 30 de comunicaciones a través de la red 2 de telecomunicaciones al terminal 1 de entrada de datos, donde el tipo de organismo y su puntuación asignada se muestran al usuario en la pantalla 13.
40

Tal como se ilustra en la Figura 4, en una realización, después de las etapas S1 a S9, la secuencia genética diana recibida en la etapa S1 es usada para refinar el perfil 42 de secuencia seleccionado en la etapa S2. En la etapa S10
45 opcional, el módulo 36 de adaptación de perfil almacena la secuencia genética diana como una secuencia genética de referencia en la base de datos 4, asignada al perfil 42 seleccionado en la etapa S2. En una realización, la secuencia genética diana es almacenada como una secuencia genética de muestra en la base de datos 4. El refinamiento del perfil 42 de secuencia es desencadenado, por ejemplo, por un número definido de muestras y/o estadísticas almacenadas de muestras y secuencias de referencia almacenadas.

50 En la etapa S11, el módulo 36 de adaptación de perfil evalúa la nueva secuencia genética de referencia de la etapa S10, o una o más secuencias genéticas de muestra almacenadas en la etapa S10, para las nuevas regiones de secuencias informativas. En particular, el módulo 36 de adaptación de perfil determina si las secuencias genéticas de referencia (y posiblemente secuencias genéticas de muestra), asignadas al perfil 42 de secuencia respectivo tienen o no regiones de secuencias informativas que todavía no están incluidas en el perfil de secuencias pero indican un tipo de organismo, es
55 decir, un género, especie, sub-tipo, variante y/o clado. En esencia, el módulo 36 de adaptación de perfil determina las posiciones o regiones de secuencia que tienen una correlación a lo largo de las secuencias genéticas de referencia respectivas (y, posiblemente, las secuencias genéticas de muestra) que supera un umbral de correlación definido. Por ejemplo, el módulo 36 de adaptación de perfil alinea la nueva secuencia de referencia (y, posiblemente, las secuencias genéticas de muestra) y las secuencias de referencia relacionadas con el perfil seleccionado, e identifica nuevas regiones
60 de secuencias informativas mediante la identificación de los códigos de nucleótidos correspondientes en la misma posición secuencial en al menos un número definido de la nueva secuencia de referencia (y, posiblemente, las secuencias genéticas de muestra) y las secuencias de referencia.

En la etapa S12, el módulo 36 de adaptación de perfil determina si hay o no una o más regiones de secuencias informativas nuevas. Si hay regiones de secuencias informativas nuevas, el procesamiento continúa en la etapa S13. De lo contrario, el procesamiento de la secuencia genética diana termina en la etapa S14.

5 En la etapa S13, en base a la región o las regiones de secuencias informativas nuevas determinadas en la etapa S11, el módulo 36 de adaptación de perfil adapta el perfil de secuencias seleccionado en la etapa S2. De esta manera, el perfil de secuencias seleccionado en la etapa S2 es refinado añadiendo la región o las regiones de secuencias informativas nuevas al perfil de secuencias. Posteriormente, el perfil de secuencias refinado puede ser sometido a una calibración y una validación iterativas, tal como se describe en el contexto de las etapas S25 y S27.

15 El módulo 38 de corrección está configurado para ayudar a corregir, o corregir automáticamente, la secuencia genética diana, recibida en la etapa S1, en base al perfil de secuencias, seleccionado en la etapa S2. Para ayudar a un usuario en la corrección de la secuencia diana, el módulo 38 de corrección indica en un interfaz de usuario cualquier región de secuencia informativa que permite la diferenciación del tipo de organismo. Preferiblemente, esta indicación es proporcionada en un alineamiento de secuencia diana, secuencia o secuencias de referencia y/o una secuencia de consenso, mostradas en la interfaz de usuario, poniendo de relieve visualmente las regiones de secuencias informativas. Para proporcionar corrección automática, para las diferencias que surgen de los códigos de nucleótidos situados en las regiones de secuencias informativas, indicativas de tipos de organismos, el módulo 38 de corrección calcula los valores de la probabilidad de que la diferencia sea debida a un error en la secuencia diana o que la diferencia indique otro tipo de organismo. Además, el módulo 38 de corrección aplica valores de umbral a los valores de probabilidad calculados y corrige automáticamente un error (por ejemplo, en la secuencia de consenso), o inserta (en la secuencia de consenso) un código especial que indica ambigüedad, por ejemplo un código IUPAC (unión internacional de química pura y aplicada), o desencadena una adaptación del perfil de secuencias, tal como se ha descrito anteriormente con referencia a las Figuras 2 y 4.

25 En la descripción de las realizaciones representativas de la invención, la memoria descriptiva puede haber presentado el procedimiento y/o el proceso como una secuencia particular de etapas. Sin embargo, en la medida en que el procedimiento o proceso no depende del orden particular de las etapas establecido en la presente memoria, el procedimiento o proceso no debería estar limitado a la secuencia de etapas particular descrita. Tal como apreciará una persona con conocimientos ordinarios en la materia, son posibles otras secuencias de etapas. Por lo tanto, el orden particular de las etapas expuestas en la memoria descriptiva no debería interpretarse como una limitación de las reivindicaciones.

30 La descripción anterior de las realizaciones de la invención ha sido presentada con fines ilustrativos y descriptivos. No pretende ser exhaustiva ni limitar la invención a las formas precisas descritas. Muchas variaciones y modificaciones de las realizaciones descritas en la presente memoria serán evidentes para una persona con conocimientos ordinarios en la materia a la luz de la descripción anterior. El alcance de la invención está definido solamente por las reivindicaciones adjuntas, y por sus equivalentes. En particular, además de los sectores de la bacteriología, micología y virología, la presente invención puede ser aplicada también en cualquier otro sector en el que estén implicadas búsquedas de similitud de secuencias, por ejemplo, en enfermedades humanas y veterinarias y predisposiciones a enfermedades (por ejemplo, cáncer), en enfermedades infecciosas tales como HBV, HIV, etc., así como en la tipificación de animales, seres humanos, plantas, microorganismos y virus.

45

REIVINDICACIONES

1. Un procedimiento implementado por ordenador para identificar tipos de organismos a partir de una secuencia genética diana, que comprende:

5 recibir la secuencia genética diana;
 seleccionar automáticamente en una base de datos (4) a partir de una pluralidad de perfiles (42) de secuencias específicas de tipos de organismos, en el que cada perfil (42) de secuencias define regiones de secuencias informativas para la diferenciación de los tipos de organismos individuales e incluye información
 10 específica de la posición relacionada con más de una secuencia (41) de referencia asociada de un tipo de organismo específico, en el que un perfil (42) de secuencias seleccionado tiene una correlación más alta con la secuencia genética diana en base a dicha información específica de la posición del perfil específico del tipo de organismo relacionado;
 recuperar automáticamente desde la base de datos (4) secuencias (41) de referencia asociadas con el perfil
 15 (42) de secuencias seleccionado;
 comparar automáticamente la secuencia genética diana con las secuencias (41) de referencia y ponderar automáticamente los resultados de comparación relacionados con las regiones de secuencias informativas usando factores de ponderación asociados con las regiones de secuencias informativas definidas en el perfil de secuencias seleccionado;
 20 determinar a partir de las secuencias (41) de referencia una secuencia (41) de referencia específica del tipo de organismo que tiene una mejor coincidencia con la secuencia genética diana, en el que la mejor coincidencia se determina en base a los resultados de comparación ponderados para las regiones de secuencias informativas; y
 asignar a la secuencia genética diana un tipo de organismo en base a la secuencia de referencia específica
 25 del tipo de organismo con la mejor coincidencia.

2. Procedimiento según la reivindicación 1, en el que los resultados de la comparación incluyen un número de diferencias en los códigos de nucleótidos de cada una de las secuencias (41) de referencia en comparación con la secuencia genética diana; en el que la ponderación de los resultados de la comparación incluye determinar para
 30 cada secuencia (41) de referencia un número ponderado de diferencias multiplicando con un factor de ponderación el número de diferencias relacionadas con las regiones de secuencias informativas; y en el que el procedimiento incluye además almacenar una lista de las secuencias (41) de referencia, en el que la lista está ordenada por el número ponderado de diferencias de la secuencia (41) de referencia correspondiente.

3. Procedimiento según una de las reivindicaciones 1 o 2, en el que los resultados de la comparación incluyen un número de correspondencias en los códigos de nucleótidos de cada una de las secuencias (41) de referencia en comparación con la secuencia genética diana; en el que la ponderación de los resultados de la comparación incluye determinar para cada secuencia (41) de referencia un número ponderado de correspondencias multiplicando con un factor de ponderación el número de correspondencias relacionadas con las regiones de
 40 secuencias informativas; y en el que el procedimiento incluye además almacenar una lista de las secuencias (41) de referencia, en el que la lista está ordenada por el número ponderado de correspondencias de la secuencia (41) de referencia correspondiente.

4. Procedimiento según una de las reivindicaciones 1 a 3, en el que la secuencia genética diana y las secuencias (41) de referencia asociadas con el perfil (42) de secuencias seleccionado se evalúan para nuevas regiones de secuencias informativas para el perfil (42) de secuencias seleccionado; y en el que el perfil (42) de secuencias seleccionado es adaptado mediante el almacenamiento de una nueva región de secuencia informativa como una parte del perfil (42) de secuencias seleccionado.

5. Procedimiento según la reivindicación 4, en el que la evaluación de la secuencia genética diana y las secuencias (41) de referencia incluye el alineamiento de la secuencia genética diana y las secuencias (41) de referencia asociadas con el perfil (42) de secuencias seleccionado, y la identificación de las nuevas regiones de secuencias informativas mediante la identificación de los códigos de nucleótidos correspondientes a una misma posición secuencial en al menos un número definido de la secuencia genética diana y las secuencias (41) de referencia.

6. Procedimiento según una de las reivindicaciones 1 a 5, en el que los perfiles (42) de secuencias específicas del tipo de organismo almacenados en la base de datos (4) se determinan mediante el alineamiento de las secuencias genéticas específicas del tipo de organismo, mediante la creación de secuencias de consenso para cada tipo de organismo, mediante la identificación de las regiones informativas que permiten la diferenciación de los tipos de organismos individuales, y mediante la definición de los perfiles (42) de secuencias específicas del tipo de organismo en base a las regiones informativas.

7. Procedimiento según la reivindicación 6, en el que los perfiles (42) de secuencias específicas del tipo de organismo almacenados en la base de datos (4) incluyen perfiles (42) de secuencias específicas del género o específicas del grupo, y en el que los perfiles (42) de secuencias específicas del género o específicas del grupo se determinan mediante el alineamiento de las secuencias genéticas específicas del género o específicas del grupo, mediante la creación de secuencias de consenso para cada tipo de organismo, mediante la identificación de las regiones informativas que permiten la diferenciación de los tipos de organismos individuales, y mediante la definición de los perfiles (42) de secuencias específicas del género o específicas del grupo en base a las regiones informativas.
8. Procedimiento según una de las reivindicaciones 1 a 7, en el que la secuencia genética diana es corregida en base al perfil (42) de secuencias seleccionado mediante la comparación de la secuencia genética diana con las secuencias (41) de referencia asociadas con el perfil (42) de secuencias seleccionado, mediante la evaluación para las diferencias de los códigos de nucleótidos, situados en las regiones de secuencias informativas, de si las diferencias indican o no otro tipo de organismo, e iniciando la adaptación del perfil (42) de secuencias seleccionado para las diferencias evaluadas como indicativas de otro tipo de organismo.
9. Procedimiento según una de las reivindicaciones 1 a 8, en el que la secuencia genética diana es recibida por un servidor (3) desde un usuario a través de una red (2) de telecomunicaciones; y en el que el tipo de organismo de la secuencia genética diana, definido por la secuencia (41) de referencia específica del tipo de organismo, es transmitido por el servidor (3) a través de la red (2) de telecomunicaciones a una interfaz de usuario.
10. Un sistema informático para la identificación de tipos de organismos a partir de una secuencia genética diana, en el que el sistema comprende:
- una base de datos (4) que comprende una pluralidad de perfiles (42) de secuencias específicas del tipo de organismo, en el que cada perfil (42) de secuencias define regiones de secuencias informativas para diferenciar tipos de organismos individuales e incluye información específica de la posición relacionada con más de una secuencia (41) de referencia asociada de un tipo de organismo específico;
 - un módulo (32) de selección de perfil (42) configurado para seleccionar automáticamente, a partir de dichos perfiles (42) de secuencias, un perfil (42) de secuencias seleccionado que tiene una correlación más alta con la secuencia genética diana en base a dicha información específica de la posición del perfil específico del tipo de organismo relacionado;
 - un módulo (33) de recuperación configurado para recuperar automáticamente desde la base de datos (4) secuencias (41) de referencia asociadas con el perfil (42) de secuencias seleccionado;
 - un módulo (34) de comparación configurado para comparar automáticamente la secuencia genética diana con las secuencias (41) de referencia y ponderar automáticamente los resultados de la comparación relacionados con las regiones de secuencias informativas usando factores de ponderación asociados con las regiones de secuencias informativas definidas en el perfil de secuencias seleccionado;
 - un módulo (35) de determinación de tipo configurado para determinar, a partir de las secuencias (41) de referencia, una secuencia (41) de referencia específica del tipo de organismo que tiene una mejor coincidencia con la secuencia genética diana, en el que la mejor coincidencia se determina en base a los resultados de comparación ponderados para las regiones de secuencias informativas; y
 - asignar a la secuencia genética diana un tipo de organismo en base a la secuencia de referencia específica del tipo de organismo con la mejor coincidencia.
11. Sistema según la reivindicación 10, en el que los resultados de la comparación incluyen un número de diferencias en los códigos de nucleótidos de cada una de las secuencias (41) de referencia en comparación con la secuencia genética diana; en el que el módulo (34) de comparación está configurado para determinar, para cada secuencia (41) de referencia, un número ponderado de diferencias multiplicando con un factor de ponderación el número de diferencias relacionadas con las regiones de secuencias informativas; y en el que el módulo (35) de determinación de tipo está configurado para almacenar una lista de las secuencias (41) de referencia, en el que la lista está ordenada por el número ponderado de diferencias de la secuencia (41) de referencia respectiva.
12. Sistema según una de las reivindicaciones 10 u 11, en el que los resultados de la comparación incluyen un número de correspondencias en los códigos de nucleótidos de cada una de las secuencias (41) de referencia en comparación con la secuencia genética diana; en el que el módulo (34) de comparación está configurado para determinar, para cada secuencia (41) de referencia, un número ponderado de correspondencias multiplicando con un factor de ponderación el número de correspondencias relacionadas con las regiones de secuencias informativas; y en el que el módulo (35) de determinación de tipo está configurado para almacenar una lista de secuencias (41) de referencia, en el que la lista está ordenada por el número ponderado de correspondencias de la secuencia (41) de referencia respectiva.

- 5 13. Sistema según una de las reivindicaciones 10 a 12, que comprende además un módulo (36) de adaptación de perfil (42) configurado para evaluar la secuencia genética diana y las secuencias (41) de referencia asociadas con el perfil (42) de secuencias seleccionado para nuevas regiones de secuencias informativas para el perfil (42) de secuencias seleccionado, y para adaptar el perfil (42) de secuencias seleccionado mediante el almacenamiento de una nueva región de secuencia informativa como parte del perfil (42) de secuencias seleccionado.
- 10 14. Sistema según la reivindicación 13, en el que el módulo (36) de adaptación de perfil (42) está configurado para alinear la secuencia genética diana y las secuencias (41) de referencia asociadas con el perfil (42) de secuencias seleccionado, y para identificar las nuevas regiones de secuencias informativas mediante la identificación de los códigos de nucleótidos correspondientes a una misma posición secuencial en al menos un número definido de la secuencia genética diana y las secuencias (41) de referencia.
- 15 15. Sistema según una de las reivindicaciones 10 a 14, que comprende además un módulo (37) de creación de perfil configurado para determinar los perfiles (42) de secuencias específicas de tipo de organismo almacenados en la base de datos (4) mediante el alineamiento de las secuencias genéticas específicas del tipo de organismo, mediante la creación de secuencias de consenso para cada tipo de organismo, mediante la identificación de las regiones informativas que permiten la diferenciación de los tipos de organismos individuales, y mediante la definición de los perfiles (42) de secuencias específicas del tipo de organismo en base a las regiones informativas.
- 20 16. Sistema según la reivindicación 15, en el que los perfiles (42) de secuencias específicas del tipo de organismo almacenadas en la base de datos (4) incluyen perfiles (42) de secuencias específicas del género o específicas del grupo, y en el que el módulo (37) de creación de perfil está configurado para determinar los perfiles (42) de secuencias específicas del género o específicas del grupo mediante el alineamiento las secuencias genéticas específicas del género o específicas del grupo, mediante la creación de secuencias de consenso para cada tipo de organismo, mediante la identificación de las regiones informativas que permiten la diferenciación de los tipos de organismos individuales, y mediante la definición de los perfiles (42) de secuencias específicas del género o específicas del grupo en base a las regiones informativas.
- 25 17. Sistema según una de las reivindicaciones 10 a 16, que comprende además un módulo (38) de corrección configurado para corregir la secuencia genética diana en base al perfil (42) de secuencias seleccionado mediante la comparación de la secuencia genética diana con las secuencias (41) de referencia asociadas con el perfil (42) de secuencias seleccionado, para evaluar para las diferencias de los códigos de nucleótidos, situados en regiones de secuencias informativas, si las diferencias indican o no otro tipo organismo, y para iniciar la adaptación del perfil (42) de secuencias seleccionado para las diferencias evaluadas como indicativas de otro tipo de organismo.
- 30 35 18. Sistema según una de las reivindicaciones 10 a 17, que comprende además un módulo (30) de comunicaciones configurado para recibir la secuencia genética diana desde un usuario a través de una red (2) de telecomunicaciones, y para transmitir el tipo de organismo de la secuencia genética diana, definido por la secuencia (41) de referencia específica del tipo de organismo, a través de la red (2) de telecomunicaciones a una interfaz de usuario.
- 40 19. Un producto de programa de ordenador adaptado para realizar el procedimiento de la reivindicación 1.
- 45 20. Producto de programa de ordenador según la reivindicación 19, que comprende además medios de código de programa de ordenador para controlar los procesadores del sistema informático, de manera que el sistema incluye en los resultados de la comparación un número de diferencias en los códigos de nucleótidos de cada una de las secuencias (41) de referencia en comparación con la secuencia genética diana; determina para cada secuencia (41) de referencia un número ponderado de diferencias multiplicando con un factor de ponderación el número de diferencias relacionadas con las regiones de secuencias informativas; y almacena una lista de las secuencias (41) de referencia, en el que la lista está ordenada por el número ponderado de las diferencias de la secuencia (41) de referencia respectiva.
- 50 55 21. Producto de programa de ordenador según una de las reivindicaciones 19 o 20, que comprende además medios de código de programa de ordenador para controlar los procesadores del sistema informático, de manera que el sistema incluye en los resultados de la comparación un número de correspondencias en los códigos de nucleótidos de cada una de las secuencias (41) de referencia en comparación con la secuencia genética diana; determina para cada secuencia (41) de referencia un número ponderado de correspondencias multiplicando con un factor de ponderación el número de correspondencias relacionadas con las regiones de secuencias informativas; y almacena una lista de las secuencias (41) de referencia, en el que la lista está ordenada por el número ponderado de las correspondencias de la secuencia (41) de referencia respectiva.
- 60 22. Producto de programa de ordenador según una de las reivindicaciones 19 a 21, que comprende además

- 5 medios de código de programa de ordenador para controlar los procesadores del sistema informático, de manera que el sistema evalúa la secuencia genética diana y las secuencias (41) de referencia asociadas con el perfil (42) de secuencias seleccionado para las nuevas regiones de secuencias informativas para el perfil (42) de secuencias seleccionado; y adapta el perfil (42) de secuencias seleccionado mediante el almacenamiento de una nueva región de secuencia informativa como una parte del perfil (42) de secuencias seleccionado.
- 10 23. Producto de programa de ordenador según la reivindicación 22, que comprende además medios de código de programa de ordenador para controlar los procesadores del sistema informático, de manera que el sistema, en la evaluación de la secuencia genética diana y las secuencias (41) de referencia, alinea la secuencia genética diana y las secuencias (41) de referencia asociadas con el perfil (42) de secuencias seleccionado; e identifica las nuevas regiones de secuencias informativas mediante la identificación de los códigos de nucleótidos correspondientes a una misma posición secuencial en al menos un número definido de la secuencia genética diana y las secuencias (41) de referencia.
- 15 24. Producto de programa de ordenador según una de las reivindicaciones 19 a 23, que comprende además medios de código de programa de ordenador para controlar los procesadores del sistema informático, de manera que el sistema determina los perfiles (42) de secuencias específicas del tipo de organismo almacenados en la base de datos (4) mediante el alineamiento de las secuencias genéticas específicas del tipo de organismo, mediante la creación de secuencias de consenso para cada tipo de organismo, mediante la identificación de las regiones informativas que permiten la diferenciación de los tipos de organismos individuales, y mediante la definición de los perfiles (42) de secuencias específicas del tipo de organismo en base a las regiones informativas.
- 20 25. Producto de programa de ordenador según la reivindicación 24, que comprende además medios de código de programa de ordenador para controlar los procesadores del sistema informático, de manera que el sistema incluye en los perfiles (42) de secuencias específicas del tipo de organismo, almacenados en la base de datos (4), perfiles (42) de secuencias específicas del género o específicas del grupo; y determina los perfiles (42) de secuencias específicas del género o específicas del grupo mediante el alineamiento de las secuencias genéticas específicas del género o específicas del grupo, mediante la creación de secuencias de consenso para cada tipo de organismo, mediante la identificación de las regiones informativas que permiten la diferenciación de los tipos de organismos individuales, y mediante la definición de los perfiles (42) de secuencias específicas del género o específicas del grupo en base a las regiones informativas.
- 25 26. Producto de programa de ordenador según una de las reivindicaciones 19 a 25, que comprende además medios de código de programa de ordenador para controlar los procesadores del sistema informático, de manera que el sistema corrige la secuencia genética diana en base al perfil (42) de secuencias seleccionado mediante la comparación de la secuencia genética diana con las secuencias (41) de referencia asociadas con el perfil (42) de secuencias seleccionado; evalúa para las diferencias de los códigos de nucleótidos, situados en regiones de secuencias informativas, si las diferencias indican o no otro tipo de organismo; e inicia la adaptación del perfil (42) de secuencias seleccionado para las diferencias evaluadas como indicativas de otro tipo de organismo.
- 30 27. Producto de programa de ordenador según una de las reivindicaciones 19 a 26, que comprende además medios de código de programa de ordenador para controlar los procesadores del sistema informático, de manera que el sistema recibe la secuencia genética diana desde un usuario a través de una red (2) de telecomunicaciones; y transmite a través de la red (2) de telecomunicaciones a una interfaz de usuario el tipo de organismo de la secuencia genética diana definido por la secuencia (41) de referencia específica del tipo de organismo.
- 35 40 45

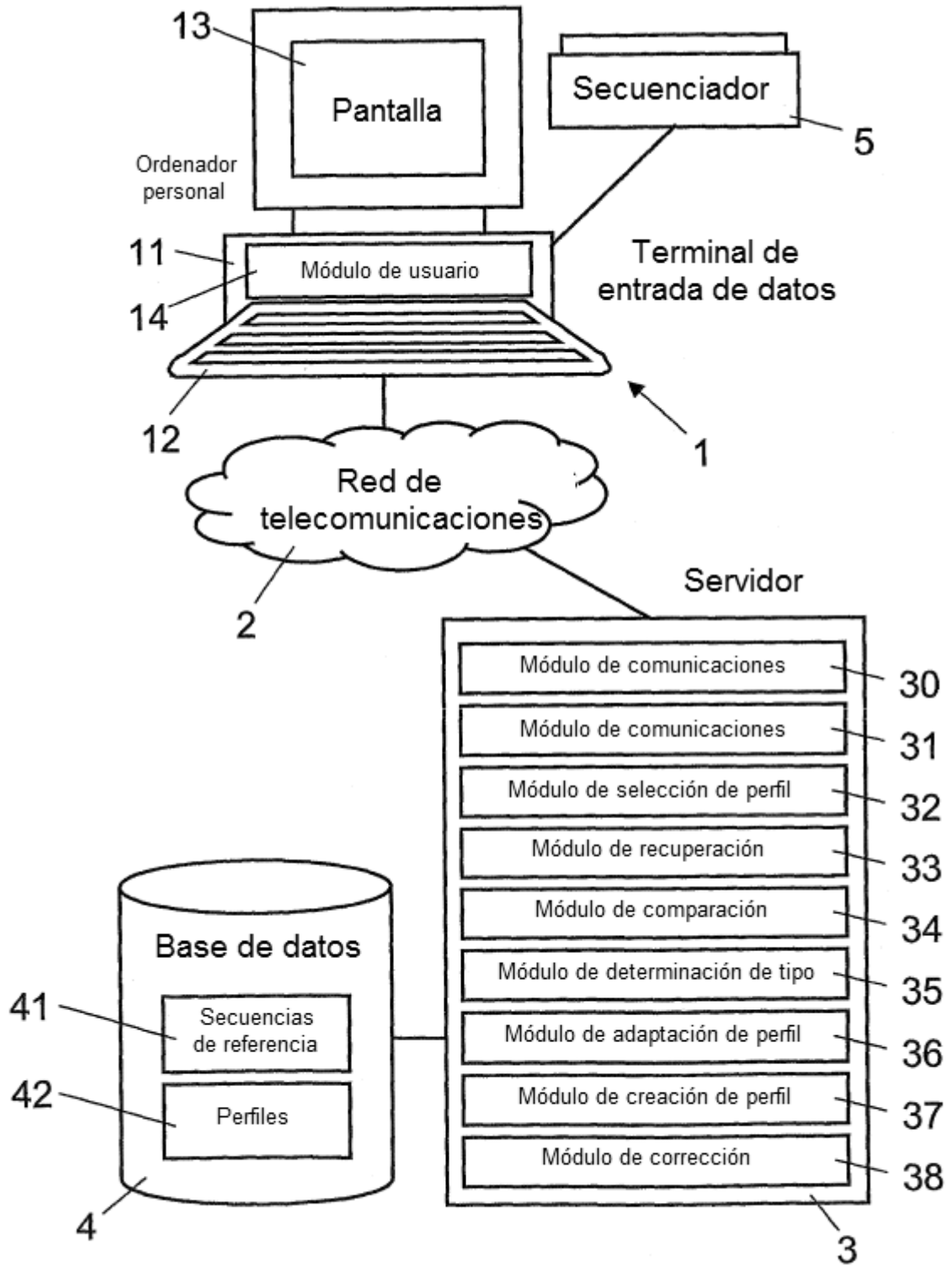


Fig. 1

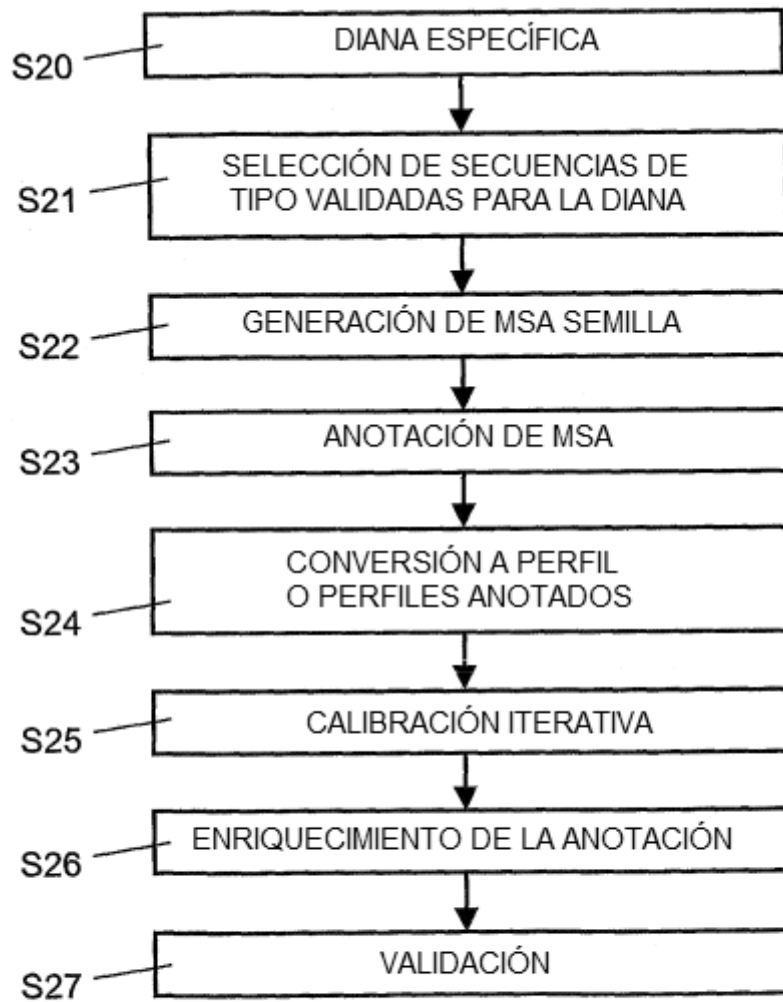


Fig. 2

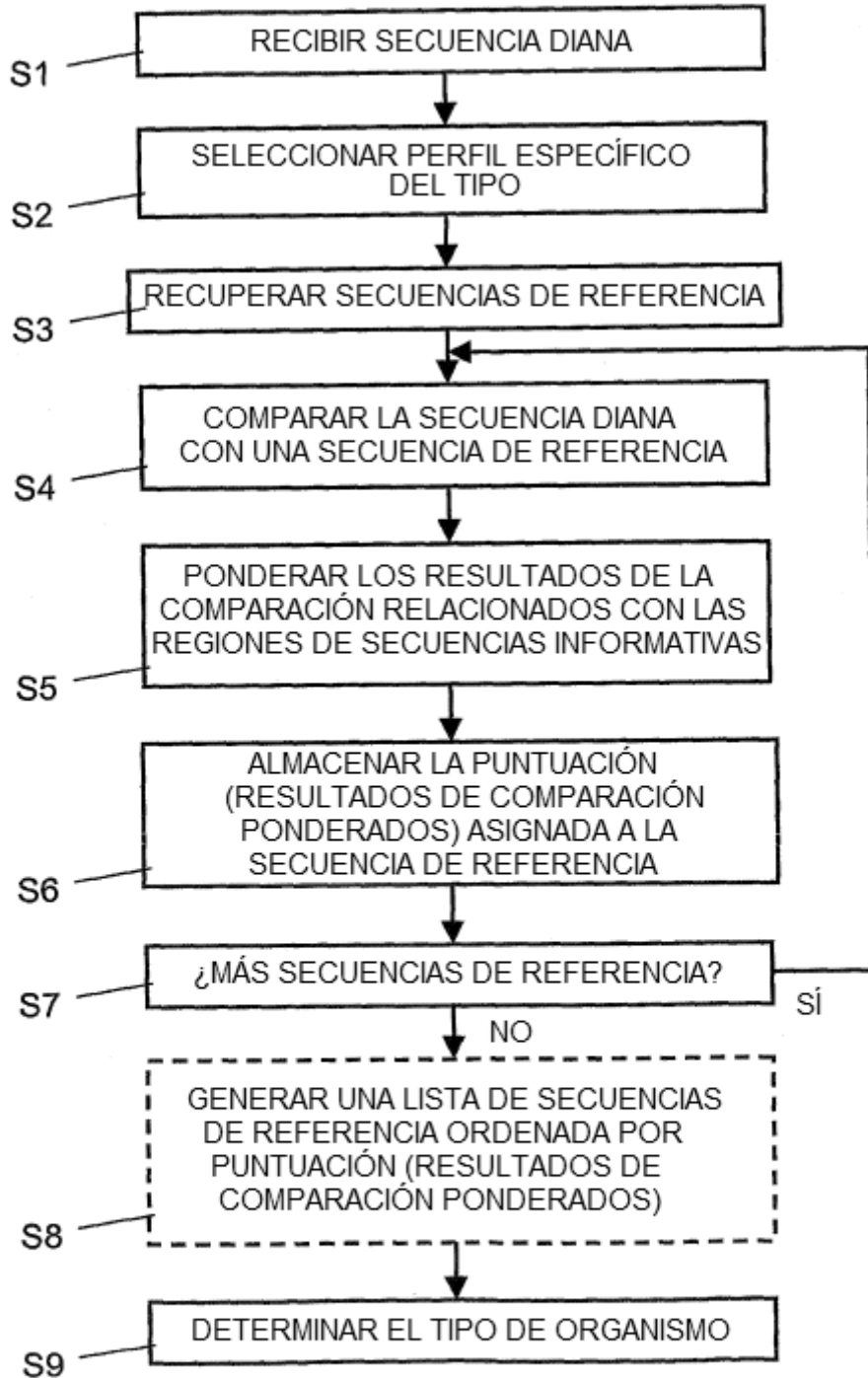


Fig. 3

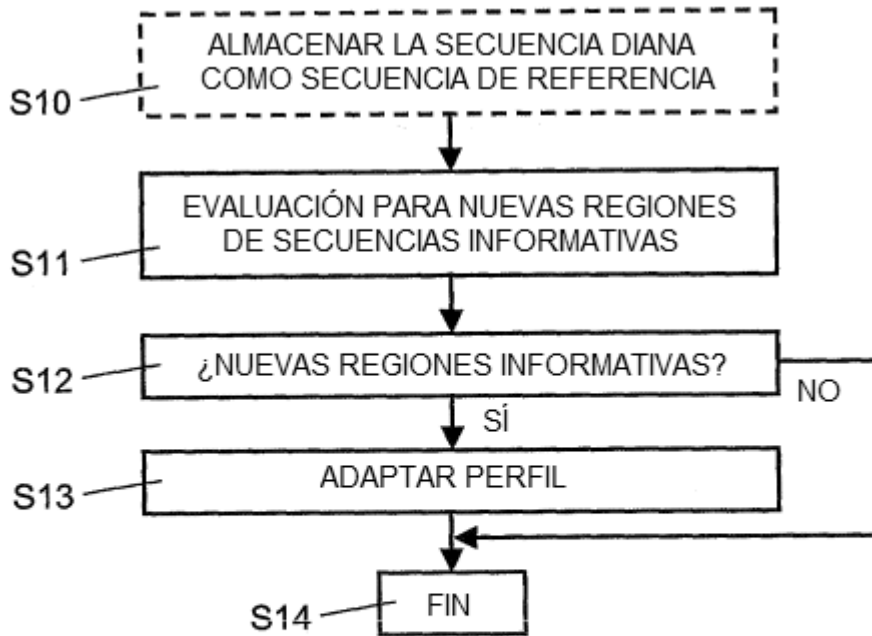


Fig. 4