



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 311 958**

51 Int. Cl.:
G06F 19/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **05700367 .5**

96 Fecha de presentación : **11.02.2005**

97 Número de publicación de la solicitud: **1846853**

97 Fecha de publicación de la solicitud: **24.10.2007**

54

Título: **Procedimiento y sistema informático para validar datos de secuenciación de ADN.**

45

Fecha de publicación de la mención BOPI:
16.02.2009

45

Fecha de la publicación del folleto de la patente:
16.02.2009

73

Titular/es: **Smartgene GmbH**
Industriestrasse 16
6300 Zug, CH

72

Inventor/es: **Emler, Stefan**

74

Agente: **Carpintero López, Mario**

ES 2 311 958 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 y sistema informático para validar datos de secuenciación de ADN.

5 Campo de la invención

La presente invención se refiere a un procedimiento y un sistema informáticos para validar datos de secuenciación de ADN. De forma específica, la presente invención se refiere a un procedimiento y un sistema informático para validar los datos de secuenciación de ADN procedentes de los datos de secuencia de uno o más fragmentos de ADN (datos de secuencia de fragmentos). La presente invención se refiere también a un producto de programa informático para controlar el sistema informático, de forma que el sistema ejecuta el procedimiento de validar los datos de secuenciación de ADN.

Antecedentes de la invención

La secuenciación del ADN (ácido desoxirribonucleico) es la determinación de la secuencia de nucleótidos precisa en una muestra de ADN. El procedimiento más habitual para la secuenciación de ADN fue desarrollado por Frederick Sanger y se denomina el procedimiento dideoxi o secuenciación de Sanger. El procedimiento dideoxi permite la secuenciación de ADN basándose en la secuenciación de fragmentos de ADN. A día de hoy, se emplean secuenciadores automáticos para generar datos de secuencia informáticos a partir de fragmentos de ADN. En su forma bruta, los datos de secuencia incluyen electroferogramas. Un electroferograma incluye una señal electroferográfica para cada uno de los cuatro tipos de nucleótidos (A: adenina, C: citosina, G: guanina y T: timina). A partir de los picos de amplitud en las señales electroferográficas pueden obtenerse códigos (A, C, G, T) para los tipos de nucleótidos. Además de las señales electroferográficas, los datos de secuencia obtenidos de un secuenciador también pueden incluir la secuencia codificada del fragmento de ADN, es decir, una secuencia de códigos de los tipos de nucleótidos derivados. De forma típica, las secuencias se validan mediante la intervención humana, por un técnico de laboratorio experto, por ejemplo. Para la validación, la secuencia de un fragmento de ADN se compara con una secuencia de referencia adecuada. Para este fin, el operario humano primero debe buscar y extraer “de forma manual” una secuencia de referencia a partir de una base de datos. Después el operario humano compara de modo visual la secuencia de un fragmento de ADN con la secuencia de referencia comprobando, nucleótido a nucleótido, la correspondencia entre los respectivos códigos de nucleótidos. La búsqueda, la selección y la extracción manuales de las secuencias de referencia consume tiempo y no proporciona garantías de que se seleccione una secuencia de referencia de modo óptimo. Es posible que exista una secuencia de referencia más adecuada que proporcione una correspondencia mejor con las múltiples secuencias de fragmentos de ADN que hay que validar y, por tanto, que ayuda a ahorrar tiempo y a reducir los errores. Además, la búsqueda y la selección de una secuencia de referencia por un operario humano son propensas a errores a medida que se desarrollan las intervenciones humana y manual.

Sumario de la invención

Un objeto de esta invención es proporcionar un procedimiento y un sistema informáticos para validar datos de secuenciación de ADN procedentes de los datos de secuencia de uno o más fragmentos de ADN (también denominados en la presente “datos de secuencia de fragmentos”) que no tienen las desventajas de la técnica anterior. En particular, un objeto de la presente invención es proporcionar un procedimiento y un sistema informáticos para validar datos de secuenciación de ADN procedentes de los datos de secuencia de fragmentos, no requiriendo dicho procedimiento y dicho sistema la intervención humana para buscar, seleccionar y extraer una secuencia de referencia para validar los datos de secuencia. Otro objeto de la presente invención es proporcionar un procedimiento y un sistema informáticos para validar datos de secuenciación de ADN procedentes de los datos de secuencia de fragmentos no requiriendo dicho procedimiento y dicho sistema la intervención humana para identificar una codificación de nucleótidos ambigua en los datos de secuencia de los fragmentos de ADN.

Según la presente invención, estos objetos se logran, en particular, a través de las características de las reivindicaciones independientes. Además, a partir de las reivindicaciones dependientes y de la descripción se seguirán otras realizaciones ventajosas.

Según la presente invención, los objetos mencionados anteriormente se logran, en particular, porque, para validar los datos de secuenciación de ADN a partir de los datos de secuencia de fragmentos de uno o más fragmentos de ADN, es decir, para validar las secuencias que resultan de una “llamada de bases” de un secuenciador, se obtiene una especificación diana de un usuario. Se identifica una secuencia de referencia seleccionada, que tenga la mayor correlación con los datos de secuencia de uno o más fragmentos secuenciados, y se selecciona de modo automático a partir de un conjunto de una o más posibles secuencias de referencia, se relaciona con la especificación diana, y se almacena en una base de datos. Los datos de secuencia de fragmentos se alinean de modo automático con la secuencia de referencia seleccionada. Se ajusta una orientación inversa al complemento con relación a la secuencia de referencia seleccionada. Se identifican de modo automático las posiciones en la secuencia en que los códigos de nucleótidos de los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados no se corresponden. La validación de los datos de secuencia de uno o más fragmentos de ADN con la selección automática de la secuencia de referencia, basándose en evaluar el nivel de correlación (es decir, el grado de correspondencia del patrón) entre una secuencia de referencia y los datos de secuencia de fragmentos de uno o más fragmentos de ADN tiene la ventaja de que no se requiere la intervención humana en el proceso de selección. Esto aumenta la calidad de la selección porque

ES 2 311 958 T3

no hay errores operativos y porque se selecciona una secuencia de referencia que se corresponde mejor, mediante la maximización de la correlación entre la secuencia de referencia y los datos de secuencia de los fragmentos de ADN. La utilización de un ordenador para seleccionar la secuencia de referencia permite utilizar un mayor número de secuencias de referencia disponibles y, con ello, se aumenta la probabilidad de correspondencias correctas. Además, basándose en la secuencia de referencia seleccionada, es posible detectar y localizar, sin intervenciones humanas, los códigos de nucleótidos que no se corresponden en los datos de secuencia de fragmentos. Comparado con la selección “manual” y la validación realizada por un operario, el tiempo de procesamiento para validar (o comprobar) datos de secuenciación de ADN se reduce significativamente, mientras que la calidad aumenta sustancialmente.

En una realización preferida, un servidor obtiene la especificación diana del usuario a través de una red de telecomunicaciones, y el servidor selecciona la secuencia de referencia seleccionada a partir de una base de datos. Por ejemplo, la especificación diana identifica una secuencia génica, y la secuencia de referencia seleccionada es seleccionada por el servidor a partir de la base de datos a partir de un conjunto de uno o más variantes de la secuencia génica. La ejecución del proceso de selección en un servidor basado en una red permite proporcionar una selección automática y eficaz (en términos de rendimiento y costes financieros) de secuencias de referencia como un servicio centralizado, disponible a una pluralidad de usuarios conectados a la red de telecomunicaciones.

En otra realización preferida, se genera un cóntigo como una secuencia consenso a partir de todos los datos de secuencia de fragmentos alineados con la secuencia de referencia seleccionada. En las posiciones de la secuencia que tengan códigos de nucleótidos que no se corresponden en los datos de secuencia de fragmentos, se inserta un código especial que indica ambigüedad (por ejemplo, un código IUPAC) en la secuencia consenso. En una realización, se genera un cóntigo como una secuencia consenso a partir de la secuencia de referencia seleccionada y a partir de los datos de secuencia de fragmentos alineados con la secuencia de referencia seleccionada. En las posiciones de la secuencia que tengan códigos de nucleótidos correspondientes o ausentes en los datos de secuencia de fragmentos, se copia un código de nucleótido de la secuencia de referencia seleccionada en la secuencia consenso. La generación de un cóntigo a partir de los datos de secuencia de fragmentos de ADN y la secuencia de referencia permite proporcionar una secuencia continua incluso cuando los datos de secuencia de fragmentos dejen secciones de la secuencia sin definir. Marcando de modo automático las posiciones en la secuencia en que secuencias solapantes de fragmentos de ADN tienen códigos de nucleótidos que no se corresponden permite reducir significativamente el tiempo necesario para validar los datos de secuencia. Un operario humano, es decir, el usuario, puede navegar con rapidez y solamente a las posiciones en la secuencia que tienen códigos de nucleótidos que no se corresponden en las secuencias alineadas.

Preferiblemente se muestran secciones, una al lado de otra, de los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados. Los datos de secuencia de cada fragmento de ADN se muestran a lo largo de una línea separada. Las posiciones en la secuencia con códigos de nucleótidos que no se corresponden se indican de forma visual en las secciones. Se obtienen instrucciones del usuario para modificar un código de nucleótido en las posiciones en la secuencia que tienen códigos de nucleótidos que no se corresponden. Los códigos de nucleótidos se modifican según las instrucciones obtenidas del usuario. Mostrando las secuencias de fragmentos de ADN y la secuencia de referencia alineadas unas al lado de la otra y a lo largo de líneas separadas permite una comparación visual muy sencilla y eficaz de los datos de secuencia de fragmentos y la secuencia de referencia. El marcaje visual de las posiciones en la secuencia con códigos de nucleótidos que no se corresponden facilita aún más la localización eficaz de posiciones ambiguas en la secuencia y la posterior edición (alteración) de los códigos de nucleótidos.

En una variante, se almacena la información acerca de los códigos de nucleótidos modificados por el usuario. De manera selectiva, las secciones modificadas de los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados, que contienen los códigos de nucleótidos modificados por el usuario, se muestran una al lado de otra. Los códigos de nucleótidos modificados por el usuario se indican de modo visual en las secciones modificadas. El almacenamiento de la información, tal como el identificador del fragmento de ADN, la posición en la secuencia, el valor previo y el identificador de usuario del operario humano que ha realizado la alteración, tiene la ventaja de que las modificaciones en los datos de secuencia de fragmentos (y/o en el cóntigo) pueden localizarse y analizarse más adelante.

En otra realización, las máscaras de secuencias se almacenan en la base de datos asignadas a las secuencias de referencia. Las máscaras de secuencias incluyen cada una información del perfil relacionada con una o más posiciones de la respectiva secuencia de referencia. Se obtiene información de interés del usuario. Se muestran secciones seleccionadas, una al lado de otra, de los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados. Las secciones seleccionadas se determinan basándose en la información de interés obtenida del usuario y de la información del perfil incluida en la máscara de secuencia asignada a la secuencia de referencia seleccionada. Las máscaras predefinidas específicas de las secuencias de referencia permiten localizar y navegar de modo automático hacia áreas de interés especificadas por el usuario en la presentación que muestra los datos de secuencia de fragmentos, la secuencia de referencia y el cóntigo alineados. Por tanto, pueden localizarse áreas de la secuencia interesantes y/o críticas conocidas de una secuencia de ADN, de una manera selectiva y eficaz.

En una variante, cada una de las máscaras de secuencias se almacena en una base de datos asignada a un identificador de usuario, y se determinan las secciones seleccionadas basándose en la máscara de secuencia asignada a un identificador de usuario obtenido del usuario. Las máscaras específicas del usuario permite que diferentes usuarios o grupos de usuario definan y asocien diferentes informaciones de perfiles con las secuencias de referencia.

En otra realización preferida, los datos de secuencia de fragmentos incluyen señales electroferográficas. Se muestran secciones, una al lado de otra, de los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados, mostrándose los datos de secuencia de cada fragmento de ADN a lo largo de líneas separadas, como una secuencia de códigos de nucleótidos y como una señal electroferográfica. Los niveles de señal de las señales electroferográficas se ajustan individualmente para los diferentes tipos de nucleótidos basándose en ajustes obtenidos del usuario. Mostrando los datos de secuencia de fragmentos alineados unos junto a otros como secuencias de códigos y como señales electroferográficas tiene la ventaja de que los códigos de nucleótidos pueden compararse directamente con las correspondientes señales electroferográficas. Mediante el ajuste de los niveles de señal de las señales electroferográficas, la comparación entre las señales electroferográficas y los correspondientes códigos de nucleótidos puede ser más sencilla y clara para el usuario.

En una realización, los datos de secuencia de fragmentos son generados por un secuenciador y cargados a través de una red de telecomunicaciones al servidor. Preferiblemente, el servidor lleva a cabo las etapas de alinear los datos de secuencia de fragmentos y la secuencia de referencia seleccionada, generando el cóntigo como una secuencia consenso, y almacenando el cóntigo en una base de datos asignado a los datos de secuencia de fragmentos, la secuencia de referencia seleccionada, un identificador de usuario obtenido del usuario, y la información acerca de los códigos de nucleótidos modificados por el usuario. Preferiblemente, los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados se muestran en un dispositivo visualizador colocado donde se encuentra el usuario. Además, a través de un terminal de entrada de datos localizado donde se encuentra el usuario se obtienen del usuario las instrucciones para ajustar un código de nucleótido en el cóntigo.

Además del procedimiento y el sistema informático para validar datos de secuenciación de ADN a partir de los datos de secuencia de uno o más fragmentos de ADN, la presente invención también se refiere a un producto de programa informático que incluye un medio de código de programa informático para controlar uno o más procesadores del sistema informático, de forma que el sistema ejecuta el procedimiento de validar datos de secuenciación de ADN basándose en los datos de secuencia de uno o más fragmentos de ADN. En particular, es un producto de programa informático que incluye un medio de lectura informática que contiene el medio de código de programa informático (por ejemplo, módulos de software programados, como se describirá con más detalle a continuación). La utilización de tecnología basada en servidores para validar datos de secuenciación de ADN permite que un usuario emplee su propio equipo informático sin tener que instalar ningún software o hardware. Además, pueden emplearse diferentes formatos de archivo de varios fabricantes de secuenciadores para los archivos de electroferogramas, permitiendo, con ello, archivar datos de secuencia de diferentes laboratorios procedentes de diferentes máquinas. La base de datos de la secuencia de referencia, la aplicación de software, así como cualquier herramienta de software, puede actualizarse en línea sin molestias para el usuario.

Breve descripción de los dibujos

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

la figura 1 muestra un diagrama de bloques que ilustra de modo esquemático un ejemplo de configuración de un sistema informático para poner en práctica las realizaciones de la presente invención, comprendiendo dicha configuración un servidor con una base de datos, y estando conectada dicha configuración con un terminal de entrada de datos a través de una red de telecomunicaciones;

la figura 2 muestra un ejemplo de una interfase del usuario gráfica para validar y editar los datos de secuencia alineados de múltiples fragmentos de ADN;

la figura 3 muestra un ejemplo de una sección de los datos de secuencia de fragmentos de ADN, el cóntigo y la secuencia de referencia alineados, en el que se ilustran los códigos de nucleótidos que no se corresponden y modificados por el usuario;

la figura 4 muestra un ejemplo de una ventana de navegación que ilustra los datos de secuencia de fragmentos de ADN y la secuencia de referencia alineados, indicándose una sección seleccionada mediante un marco rectangular.

Descripción detallada de las realizaciones preferidas

En la figura 1, el número de referencia 1 se refiere al terminal de entrada de datos. Tal como se ilustra en la figura 1, el terminal de entrada de datos 1 incluye un ordenador personal 11 con un teclado 12 y una pantalla visualizadora 13. Tal como se ilustra de forma esquemática, el ordenador personal 11 incluye un módulo de usuario 14 y un módulo de edición 15. El módulo de usuario 14 y el módulo de edición 15 se ejecutan como un módulo de software programado, por ejemplo, un programa aplique ejecutable que se descarga del servidor 3 a través de una red de telecomunicaciones 2.

Conectado al ordenador personal 11 está un secuenciador convencional 5, que proporciona al ordenador personal 11 los datos de secuencias de fragmentos de ADN. Preferiblemente, los datos de secuencia de fragmentos incluyen electroferogramas de los fragmentos de ADN, incluyendo cada electroferograma señales electroferográficas de los cuatro tipos de nucleótidos (A, C, G, T).

ES 2 311 958 T3

Tal como se ilustra en la figura 1, el terminal de entrada de datos 1 se conecta al servidor 3 a través de una red de telecomunicaciones 2. Preferiblemente, la red de telecomunicaciones 2 incluye el Internet y/o un Intranet, lo cual hace que el servidor 3 sea accesible como un servidor de red a través del World Wide Web o dentro una red IP diferente, respectivamente. La red de telecomunicaciones 2 también puede incluir otra red fija, tal como una red de área local (LAN) o una red digital de servicios integrada (ISDN) y/o una red inalámbrica, tal como una red de radio móvil (por ejemplo, el Sistema Global de Telecomunicaciones Móviles (GSM) o el Sistema Universal de Telecomunicaciones Móviles (UMTS)) o una red de área local inalámbrica (WLAN).

Tal como se ilustra esquemáticamente en la figura 1, el servidor 3 está conectado a una base de datos 4. El servidor 3 puede incluir uno o más ordenadores, y cada uno puede tener uno o varios procesadores. La base de datos 4 puede ejecutarse en un ordenador compartido con el servidor 3 o en otro ordenador.

El servidor 3 incluye diferentes módulos funcionales, a saber, un módulo de comunicación 34, un módulo de aplicación 35, un módulo de selección 30, un módulo de alineamiento 31, un módulo de ensamblaje 32 y un módulo de detección 33. El módulo de comunicación 35 incluye elementos de hardware y software convencionales configurados para el intercambio de datos a través de la red de telecomunicaciones 2 con una pluralidad de terminales de entrada de datos 1. El módulo de aplicación 35 es un módulo de software programado configurado para proporcionar a los usuarios del terminal de entrada de datos 1 una interfase del usuario. Preferiblemente, la interfase del usuario se proporciona a través de un navegador de Internet convencional, tal como Microsoft Explorer o Mozilla. El módulo de selección 30, el módulo de alineamiento 31, el módulo de ensamblaje 32 y el módulo de detección 33 son módulos de software programados que se ejecutan en un ordenador del servidor 3. Aunque no se ilustra en la figura 1, el servidor 3 también incluye copias del módulo de usuario 14 con un módulo de edición 15 para su descarga por el módulo de aplicación 35 hacia el terminal de entrada de datos 1 para su ejecución en un procesador de un ordenador personal 11.

TABLA 1

ID de la secuencia de referencia	ID del usuario	Máscara de referencia				
		ID de la máscara de referencia	Información del perfil			
			ID del área de interés	Descripción/nombre del área de interés	Posición de inicio del intervalo	Posición de fin del intervalo

Tal como se ilustra esquemáticamente en la figura 1, la base de datos 4 incluye identificadores de usuario 41, secuencias de referencia 42, y máscaras de secuencias 43. Los identificadores de usuario 431 se asignan a los datos del usuario de los usuarios y/o grupos de usuarios registrados. Las secuencias de referencia 42 se almacenan como conjuntos diferentes de secuencias de referencia relacionadas. Cada conjunto incluye diferentes variantes de una secuencia génica específica. Las máscaras de secuencias 43 se almacenan asignadas a las secuencias de referencia 42. En una variante, las secuencias de referencia 42 y/o las máscaras de secuencias 43 son específicas del usuario y se almacenan asignadas a los identificadores de usuario 41. Las máscaras de secuencias 43 incluyen información del perfil relacionada con una o más posiciones de la respectiva secuencia de referencia. Preferiblemente, la información del perfil está relacionada con un intervalo en la respectiva secuencia de referencia. El intervalo se define, por ejemplo, por una posición de inicio y de fin en la secuencia de referencia, o por una posición de inicio y una longitud (es decir, un número de posiciones en la secuencia). Asignada a estos intervalos definidos, la información del perfil incluye descripciones y/o nombres de área de interés específicas en la respectiva secuencia de referencia. Por ejemplo, las áreas de interés incluyen posiciones que codifican resistencia, desapareamientos, ambigüedades u otras zonas especiales o críticas. Tal como se ilustra en la tabla 1, cada máscara de referencia puede incluir un identificador de máscara de referencia, y cada área de interés también puede incluir un identificador de área de interés.

A través de la interfase del usuario proporcionada por el módulo de aplicación 35, se pide al usuario del terminal de entrada de datos 1 que introduzca información de acceso que incluye el identificador (o cuenta) de usuario y una contraseña, por ejemplo. Basándose en los identificadores de usuario 41 almacenados en la base de datos 4, el módulo de aplicación 35 comprueba los derechos de acceso del usuario. Tras pasar el control de acceso, el usuario puede pedir la carga, desde el ordenador personal 11 al servidor 3, de los datos de secuencia de fragmentos de ADN procedentes de una muestra de ADN, por ejemplo, desde el secuenciador 5 o desde otra fuente.

Para la validación de los datos de secuenciación de ADN, la interfase del usuario proporcionada por el módulo de aplicación 35 está configurada para que el usuario seleccione, por ejemplo, de una lista, los datos de secuencia de uno o más fragmentos de ADN de una muestra de ADN, cargados y almacenados previamente en el servidor 3 o en la base de datos 4. También se pide al usuario, a través de la interfase del usuario, que introduzca una especificación diana que identifica a una secuencia génica diana. Después, el usuario inicia el proceso de validación activando un elemento de control, tal como un botón de gráfica, en la interfase del usuario proporcionado por el módulo de aplicación 35.

En respuesta al inicio recibido del usuario a través de la interfase del usuario, el módulo de aplicación 35 activa el módulo de selección 30. El módulo de selección 30 está configurado para seleccionar y extraer, desde la base

ES 2 311 958 T3

de datos 4, el conjunto de secuencias de referencia relacionado con la secuencia génica diana especificada por el usuario. Después, el módulo de selección 30 determina, para cada secuencia de referencia en el conjunto extraído, la correlación con los datos de secuencia previamente seleccionados del fragmento de ADN de la muestra de ADN. Para una secuencia de referencia particular del conjunto, se emplea una correspondencia del patrón convencional, adaptable y ajustable por el usuario con respecto a los requerimientos específicos de la diana, para determinar la correlación de la secuencia de referencia particular con los datos de secuencia seleccionados de cada fragmento de ADN. A partir del conjunto seleccionado de secuencias de referencia, el módulo de selección 30 selecciona la secuencia de referencia que tenga la mayor correlación con los datos de secuencia del fragmento como la secuencia de referencia seleccionada, por ejemplo, el variante de la secuencia génica que tenga la mayor correlación con las secuencias de los fragmentos de ADN, definidas por los datos de secuencia del fragmento.

Después de la selección de la secuencia de referencia con la mayor correlación, el módulo de aplicación 35 activa el módulo de alineamiento 31. El módulo de alineamiento 31 alinea de modo automático los datos de secuencia (es decir, la secuencia) de cada fragmento de ADN con la secuencia de referencia previamente seleccionada. El alineamiento se realiza con respecto a una correlación óptima entre la secuencia de referencia seleccionada y la secuencia del respectivo fragmento de ADN. En la figura 4 se muestra una ventana de navegación 8 que ilustra el alineamiento de los datos de secuencia de seis fragmentos de ADN, 16S-SMG1, 16S-SMG2, 16S-SMG3, 16S-SMG4, 16S-SMG5, y 16S-SMG6, con la secuencia de referencia seleccionada AY328725 (por ejemplo, una secuencia génica particular de una bacteria no cultivada). Como puede observarse en la figura 4, los datos (secuencia) de cada fragmento de ADN se muestran esquemáticamente en su línea individual 81, 82, 83, 84, 85, 86 junto a la representación esquemática de la secuencia de referencia (y alineados con ésta) en la línea 87. En la ventana de navegación 8, se indica la posición de inicio "1", así como la posición de fin "1402" de la secuencia de referencia. Además, las posiciones en la secuencia de inicio y de fin de la secuencia (datos) alineada de cada fragmento de ADN se indican en la ventana de navegación 8 (16S-SMG1: 732-1402; 16S-SMG2: 1-490; 16S-SMG3: 742-1402; 16S-SMG4: 243-931; 16S-SMG5: 1-660; y 16S-SMG6: 340-1055).

Después del alineamiento de los datos de secuencia del fragmento, el módulo de aplicación 35 activa el módulo de ensamblaje 32. El módulo de ensamblaje 32 está configurado para generar un cóntigo a partir de los datos de secuencia del fragmento alineados (en una variante, también a partir de la secuencia de referencia alineada). El cóntigo se genera como una secuencia consenso a partir de todos los datos de secuencia de fragmentos alineados con la secuencia de referencia seleccionada (en una variante, también a partir de la secuencia de referencia seleccionada). El módulo de detección 33 se invoca para que identifique las posiciones en la secuencia en que los códigos de nucleótidos de los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados no se corresponden. Si en una posición de secuencia específica, los códigos de nucleótidos en los datos de secuencia de los fragmentos de ADN alineados muestran códigos de nucleótidos que no se corresponden, o si en una posición específica en la secuencia los códigos de nucleótidos en los datos de secuencia alineados de cualquier fragmento de ADN tienen códigos de nucleótidos que no se corresponden con los códigos de nucleótidos en la secuencia de referencia, el módulo de detección 33 identifica esa posición específica en la secuencia que tiene códigos de nucleótidos que no se corresponden. Preferiblemente, los códigos de nucleótidos que no se corresponden y/o las posiciones en la secuencia que tienen códigos de nucleótidos que no se corresponden se marcan. Por ejemplo, para un código de nucleótido que no se corresponde, se almacena la posición en la secuencia y, si se determina, un identificador del fragmento de ADN asociado con el código de nucleótido que no se corresponde, asignados a los datos de secuencia del fragmento. Preferiblemente, en posiciones en la secuencia en que los códigos de nucleótidos en los datos de secuencia del fragmento se corresponden con el código de nucleótidos en la secuencia de referencia y en posiciones en la secuencia en que los códigos de nucleótidos no están presentes en los datos de secuencia del fragmento, el módulo de ensamblaje 32 copia, en la secuencia consenso, el código de nucleótidos de la secuencia de referencia seleccionada. En las posiciones en la secuencia identificadas por el módulo de detección 33 como que contienen códigos de nucleótidos que no se corresponden, el módulo de ensamblaje 32 inserta, en la secuencia consenso, un código especial que indica ambigüedad, por ejemplo, un código IUPAC (International Union of Pure and Applied Chemistry).

En el módulo de aplicación 35 y en el módulo de edición 15 se incluye una función de borrado. Cuando se determina que secciones de una secuencia de un fragmento de ADN tienen una correlación muy baja con la secuencia de referencia y/o las secuencias alineadas de otros fragmentos de ADN (un fenómeno que se observa a menudo en los dos extremos de la secuencia), la función de borrado permite que el usuario borre, de forma selectiva, áreas en los extremos de la secuencia de un fragmento de ADN. La información acerca de las secciones borradas de secuencias de fragmentos de ADN se almacena asignada a los datos de secuencia del fragmento. La función de borrado también es realizada automáticamente por el sistema para los datos "basura" evidentes en ambos extremos de un fragmento de secuencia, que tengan una correlación con la secuencia de referencia y/o las secuencias alineadas de otros fragmentos de ADN por debajo de un umbral definido. Esto facilita en gran medida la comprobación al usuario y también facilita los alineamientos automáticos de fragmentos.

Tras generarse el cóntigo, el módulo de aplicación 35 crea un conjunto de datos para el usuario. El conjunto de datos incluye la especificación diana y los datos de secuencia de fragmentos especificados por el usuario, la secuencia de referencia seleccionada por el módulo de selección 30, las máscaras de secuencias asignadas a la secuencia de referencia seleccionada y al usuario, el cóntigo generado por el módulo de ensamblaje 32, y cualquier información concerniente a los códigos de nucleótidos que no se corresponden y/o posiciones en la secuencia que tengan códigos de nucleótidos que no se corresponden, según fueron identificadas por el módulo de detección 33. El módulo de aplicación 35 transmite el conjunto de datos y las copias del módulo de usuario 14 y el módulo de edición 15 a través de la red de

ES 2 311 958 T3

telecomunicaciones 2 al ordenador personal 11 del usuario. Tal como se explicará a continuación con más detalle, el conjunto de datos también puede incluir información acerca de los códigos de nucleótidos modificados por el usuario.

El módulo de usuario 14 con el módulo de edición 15 se instalan y se activan en el ordenador personal 11. Cuando se activa, el módulo de usuario 14 controla un procesador del ordenador personal 11, de forma que genera la interfase del usuario gráfica 7 en el dispositivo visualizador 13.

Tal como se ilustra en la figura 2, a partir del conjunto de datos transmitido al ordenador personal 11, el módulo de usuario 14 muestra en la interfase del usuario gráfica 7 las secciones alineadas (por ejemplo, desde “793”, como la posición más baja en la secuencia mostrada, hasta “832”, como la posición más alta en la secuencia mostrada) de los datos de secuencia de los fragmentos de ADN, de la secuencia de referencia y del cóntigo. Los datos de secuencia de cada fragmento de ADN se muestran a lo largo de líneas separadas, como una secuencia de códigos de nucleótidos y como una señal electroferográfica. La secuencia de referencia y el cóntigo (secuencia consenso) se muestran una al lado del otro a lo largo de líneas separadas como una secuencia de códigos de nucleótidos. La interfase del usuario gráfica 7 también incluye elementos de control 71, 72, 73, 74 para ajustar los niveles de señal de las señales electroferográficas. Los elementos de control 71, 72, 73, 74 se asocian con las señales electroferográficas de cada fragmento de ADN para ajustar los niveles de señal de las señales electroferográficas de cada tipo de nucleótido para cada fragmento de ADN.

Tal como se ilustra en la figura 2, la interfase del usuario gráfica 7 incluye una barra de desplazamiento horizontal para seleccionar la sección de los datos de secuencia de fragmentos, la secuencia de referencia y el cóntigo alineados que se van a mostrar. La ventana de navegación 8 incluye un marco 88, que muestra la sección seleccionada que se muestra en la interfase del usuario gráfica 7. Desplazando la barra de desplazamiento horizontal se puede mover la sección seleccionada a lo largo de las posiciones en la secuencia.

Además, la interfase del usuario gráfica 7 incluye un menú desplegable para seleccionar áreas de interés. Los elementos del menú se rellenan según la información del perfil incluida en la máscara de referencia asociada con la secuencia de referencia. Cada descripción o nombre de un área de interés incluida en la información del perfil se lista como un elemento del menú en el menú desplegable. Cuando el usuario selecciona uno de los elementos del menú desplegable, la sección seleccionada mostrada de los datos de secuencia de fragmentos, la secuencia de referencia y el cóntigo se ajustan para incluir el intervalo de secuencia asociado en la información del perfil con la descripción seleccionada o el nombre del área de interés. Si el intervalo es mayor que el número de posiciones en la secuencia que pueden mostrarse en la interfase del usuario gráfica 7, la posición de inicio del intervalo se selecciona como la posición más baja en la secuencia mostrada.

Basándose en cualquier información, incluida en el conjunto de datos recibido, acerca de los códigos de nucleótidos que no se corresponden y/o las posiciones en la secuencia que tienen códigos de nucleótidos que no se corresponden, el módulo de usuario 14 indica de modo visual en la sección mostrada cualquier posición en la secuencia con códigos de nucleótidos que no se corresponden. Tal como se ilustra en la figura 3, las posiciones en la secuencia que tienen códigos de nucleótidos que no se corresponden se destacan mediante un color o mediante símbolos gráficos, por ejemplo. Tal como puede observarse en la figura 3, la sección mostrada 61a de una secuencia de un fragmento de ADN incluye un nucleótido diferente 661a que la sección mostrada 62a de una secuencia alineada de un fragmento de ADN. Esta ambigüedad puede indicarse en la sección mostrada 63a del cóntigo destacando o coloreando y/o ajustando un código de nucleótido 631a diferente del código del correspondiente nucleótido de la sección mostrada 64a de la secuencia de referencia.

Para la modificación, el módulo de edición 15 está configurado para aceptar del usuario la selección de un nucleótido particular 611b en la sección mostrada 61b o 62b de una secuencia de un fragmento de ADN. Para el nucleótido seleccionado 611b, el módulo de edición 15 recibe del usuario un código de nucleótido alternativo. El módulo de edición 15 modifica, en consecuencia, el código del nucleótido seleccionado 611b. Además, el módulo de edición 15 ajusta de modo automático el nuevo código de nucleótido para el correspondiente nucleótido 631b en la sección mostrada 63b del cóntigo, con la condición de que no existan más códigos de nucleótidos que no se corresponden en los datos de secuencia del fragmento en esta posición particular en la secuencia.

El módulo de edición 15 también está configurado para que almacene la información acerca de las modificaciones de códigos de nucleótidos realizadas por el usuario, tal como la posición en la secuencia, el identificador del fragmento de ADN, el código de nucleótido previo, el identificador de usuario del operario responsable de la modificación, y la fecha y momento de la modificación. La información acerca de las modificaciones incluye la misma información también para las secciones borradas de las secuencias de los fragmentos de ADN.

Basándose en la información almacenada acerca de las modificaciones, el módulo de usuario 14 indica de modo visual los códigos de nucleótidos modificados por el usuario en la interfase del usuario gráfica 7, por ejemplo, destacando un color de fondo o mediante un atributo parpadeante. Además, es posible que el usuario indique al módulo de usuario 14 que muestre, en la interfase del usuario gráfica 7, las secciones modificadas por el usuario (y la correspondiente secuencia de referencia y cóntigo), es decir, las secciones con datos de secuencia de fragmentos que incluyen al menos un código de nucleótido modificado por el usuario. Con los botones de comando “siguiente” y “anterior” la sección que se va a mostrar se mueve hasta la siguiente o la anterior sección modificada por el usuario.

ES 2 311 958 T3

5 Cuando el usuario finaliza la validación (comprobación y probablemente corrección) de los datos de secuenciación de ADN, el módulo de usuario 14 transmite los datos de validación a través de la red de telecomunicaciones 2 al servidor, en donde son almacenados por el módulo de aplicación 35 en la base de datos 4. Los datos de validación incluyen los datos de secuencia de fragmentos (incluyendo cualquier modificación), cualquier información acerca de las modificaciones, el cóntigo (incluyendo cualquier modificación), la secuencia de referencia (o al menos un identificador de la secuencia de referencia), así como la especificación diana. El módulo de aplicación 35 permite también que un usuario seleccione y se descargue los datos de validación almacenados, es decir, los datos de validación pueden descargarse desde la base de datos 4 hacia el ordenador personal 11 para analizarse y procesarse mediante el módulo de usuario 14 y el módulo de edición 15. Sin embargo, cualquier otra modificación dará como resultado el almacenamiento de otra versión más de los datos de validación. Los datos de validación almacenados también sirven como un rastro de auditoría.

10

15

20

25

30

35

40

45

50

55

60

65

ES 2 311 958 T3

REIVINDICACIONES

1. Un procedimiento informático para validar datos de secuenciación de ADN a partir de los datos de secuencia de fragmentos de uno o más fragmentos de ADN, que comprende:

obtener de un usuario una especificación diana que identifica una secuencia génica diana;

seleccionar de modo automático, a partir de un grupo de una o más posibles secuencias de referencia (42) relacionadas con la especificación diana y almacenadas en una base de datos (4), una secuencia de referencia seleccionada que tenga la mayor correlación con los datos de secuencia de fragmentos;

alinear de modo automático los datos de secuencia de fragmentos con la secuencia de referencia seleccionada; e

identificar de forma automática las posiciones en la secuencia en que los códigos de nucleótidos de los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados no se corresponden.

2. El procedimiento según la reivindicación 1, en el que la especificación diana es obtenida del usuario por un servidor (3) a través de una red de telecomunicaciones (2); en el que la especificación diana identifica una secuencia génica diana; y en el que la secuencia de referencia seleccionada es seleccionada por el servidor (3) a partir de la base de datos (4) a partir de un conjunto de uno o más variantes de la secuencia génica diana.

3. El procedimiento según una cualquiera de las reivindicaciones 1 ó 2, en el que se genera un cóntigo como una secuencia consenso a partir de los datos de secuencia de fragmentos alineados con la secuencia de referencia seleccionada, insertando en la secuencia consenso, en las posiciones en la secuencia que tienen códigos de nucleótidos que no se corresponden en los datos de secuencia de fragmentos, un código especial que indica ambigüedad.

4. El procedimiento según una cualquiera de las reivindicaciones 1 a 3, que comprende además mostrar secciones, una al lado de otra, de los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados, mostrándose los datos de secuencia de fragmentos de cada fragmento de ADN a lo largo de una línea separada; indicar en las secciones, de modo visual, las posiciones en la secuencia con códigos de nucleótidos que no se corresponden; obtener del usuario instrucciones para modificar un código de nucleótido en las posiciones en la secuencia que tienen códigos de nucleótidos que no se corresponden; y modificar los códigos de nucleótidos según las instrucciones obtenidas del usuario.

5. El procedimiento según la reivindicación 4, que comprende además almacenar la información acerca de los códigos de nucleótidos modificados por el usuario; mostrar de forma selectiva, una al lado de otra, las secciones modificadas de los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados que contienen los códigos de nucleótidos modificados por el usuario; e indicar en las secciones modificadas, de modo visual, los códigos de nucleótidos modificados por el usuario.

6. El procedimiento según una cualquiera de las reivindicaciones 1 a 5, que comprende además almacenar en la base de datos (4) las máscaras de secuencias (43) asignadas a las secuencias de referencia (42), incluyendo cada una de las máscaras de secuencias (43) información del perfil relacionada con una o más posiciones de la respectiva secuencia de referencia; obtener del usuario información de interés; y mostrar, una al lado de otra, secciones seleccionadas de los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados, determinándose las secciones seleccionadas basándose en la información de interés obtenida del usuario y la información del perfil incluida en la máscara de secuencia asignada a la secuencia de referencia seleccionada.

7. El procedimiento según la reivindicación 6, en el que cada una de las máscaras de secuencias (43) se almacena en la base de datos (4) asignada a un identificador de usuario; y en el que las secciones seleccionadas se determinan basándose en la máscara de secuencia asignada a un identificador de usuario obtenido del usuario.

8. El procedimiento según una cualquiera de las reivindicaciones 1 a 7, en el que los datos de secuencia de fragmentos incluyen señales electroferográficas; y en el que el procedimiento comprende además mostrar, una al lado de otra, secciones de los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados, mostrándose los datos de secuencia de fragmentos de cada fragmento de ADN a lo largo de líneas separadas, como una secuencia de códigos de nucleótidos y como una señal electroferográfica, ajustándose los niveles de señal de las señales electroferográficas de manera individual para los diferentes tipos de nucleótidos basándose en ajustes obtenidos del usuario.

9. El procedimiento según una cualquiera de las reivindicaciones 1 a 8, en el que los datos de secuencia de fragmentos son generados por un secuenciador (5) y cargados a través de una red de telecomunicaciones (2) a un servidor (3); en el que la base de datos (4) está conectada al servidor (3); en el que los datos de secuencia de fragmentos y la secuencia de referencia seleccionada son alineados por el servidor (3); en el que un cóntigo es generado por el servidor (3) como una secuencia consenso; en el que los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados se muestran en un terminal de visualización (13) colocado donde se encuentra el usuario; en el que se obtienen del usuario instrucciones para ajustar, en el cóntigo, un código de nucleótido, a través de un terminal

ES 2 311 958 T3

de entrada de datos (1) colocado donde se encuentra el usuario; y en el que el cóntigo es almacenado por el servidor (3) en una base de datos (4) asignado a los datos de secuencia de fragmentos, la secuencia de referencia seleccionada, un identificador de usuario obtenido del usuario, y la información acerca de los códigos de nucleótidos modificados por el usuario.

5 10. Un sistema informático para validar datos de secuenciación de ADN a partir de los datos de secuencia de fragmentos de uno o más fragmentos de ADN, comprendiendo el sistema:

un medio para obtener de un usuario una especificación diana que identifica una secuencia génica diana;

10 una base de datos (4) que comprende secuencias de referencia (42);

un módulo de selección (30) configurado para que seleccione de modo automático, a partir de un conjunto de una o más posibles secuencias de referencia (42) relacionadas con la especificación diana y almacenadas en la base de datos (4), una secuencia de referencia seleccionada que tenga la mayor correlación con los datos de secuencia de fragmentos;

un módulo de alineamiento (31) configurado para alinear, de modo automático, los datos de secuencia de fragmentos con la secuencia de referencia seleccionada; y

un módulo de detección (33) configurado para identificar posiciones en la secuencia en que los códigos de nucleótidos de los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados no se corresponden.

11. El sistema según la reivindicación 10, en el que el medio para obtener la especificación diana incluye un servidor (3) configurado para obtener la especificación diana a partir del usuario a través de una red de telecomunicaciones (2); en el que la especificación diana identifica una secuencia génica diana; y en el que el módulo de selección (30) está colocado en el servidor (3) y configurado para seleccionar la secuencia de referencia seleccionada a partir de la base de datos (4) a partir de un conjunto de uno o más variantes de la secuencia génica diana.

12. El sistema según una cualquiera de las reivindicaciones 10 u 11, en el que el sistema incluye además un módulo de ensamblaje (32) configurado para generar un cóntigo como una secuencia consenso a partir de los datos de secuencia de fragmentos alineados con la secuencia de referencia seleccionada, insertando en la secuencia consenso, en las posiciones en la secuencia que tienen códigos de nucleótidos que no se corresponden en los datos de secuencia de fragmentos, un código especial que indica ambigüedad.

13. El sistema según una cualquiera de las reivindicaciones 10 a 12, que comprende además un módulo de usuario (14) configurado para mostrar secciones, una al lado de otra, de los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados, mostrándose los datos de secuencia de fragmentos de cada fragmento de ADN a lo largo de una línea separada, e indicándose de modo visual las posiciones en la secuencia con códigos de nucleótidos que no se corresponden, y obtener del usuario instrucciones para modificar un código de nucleótido en las posiciones en la secuencia que tienen códigos de nucleótidos que no se corresponden, y modificar los códigos de nucleótidos según las instrucciones obtenidas del usuario.

14. El sistema según la reivindicación 13, en el que el sistema comprende además un medio para almacenar la información acerca de los códigos de nucleótidos modificados por el usuario; y en el que el módulo de usuario (14) está configurado para mostrar de forma selectiva, una al lado de otra, las secciones modificadas de los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados que contienen los códigos de nucleótidos modificados por el usuario; e indicar en las secciones modificadas, de modo visual, los códigos de nucleótidos modificados por el usuario.

15. El sistema según una cualquiera de las reivindicaciones 10 a 14, que comprende además máscaras de secuencias (43) almacenadas en la base de datos (4) asignadas a las secuencias de referencia (42), incluyendo cada una de las máscaras de secuencias (43) información del perfil relacionada con una o más posiciones de la respectiva secuencia de referencia; un medio para obtener del usuario información de interés; y un módulo de usuario (14) configurado para mostrar, una al lado de otra, secciones seleccionadas de los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados, determinándose las secciones seleccionadas basándose en la información de interés obtenida del usuario y la información del perfil incluida en la máscara de secuencia asignada a la secuencia de referencia seleccionada.

16. El sistema según la reivindicación 15, en el que cada una de las máscaras de secuencias (43) se almacena en la base de datos (4) asignada a un identificador de usuario; y en el que las secciones seleccionadas se determinan basándose en la máscara de secuencia asignada a un identificador de usuario obtenido del usuario.

17. El sistema según una cualquiera de las reivindicaciones 10 a 16, en el que los datos de secuencia de fragmentos incluyen señales electroferográficas; y en el que el sistema comprende además un módulo de usuario (14) configurado para mostrar, una al lado de otra, secciones de los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados, mostrándose los datos de secuencia de fragmentos de cada fragmento de ADN a lo largo de líneas separadas, como una secuencia de códigos de nucleótidos y como una señal electroferográfica, ajustándose

ES 2 311 958 T3

los niveles de señal de las señales electroferográficas de manera individual para los diferentes tipos de nucleótidos basándose en ajustes obtenidos del usuario.

18. El sistema según una cualquiera de las reivindicaciones 10 a 17, en el que el sistema incluye un servidor (3) configurado para recibir los datos de secuencia de fragmentos, generados por un secuenciador (5), a través de una red (2) de telecomunicaciones; en el que la base de datos (4) está conectada al servidor (3); en el que el módulo de alineamiento (31) está colocado en el servidor (3); en el que el sistema incluye además un módulo de ensamblaje (32) colocado en el servidor (3) y configurado para generar un cóntigo como una secuencia consenso; y un módulo de usuario (14) configurado para mostrar los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados en un dispositivo visualizador (13) colocado donde se encuentra el usuario; en el que el módulo de usuario (14) está configurado para obtener del usuario, a través de un terminal de entrada de datos (1) colocado donde se encuentra el usuario, instrucciones para ajustar, en el cóntigo, un código de nucleótido; y en el que el servidor (3) está configurado para almacenar el cóntigo en la base de datos (4) asignado a los datos de secuencia de fragmentos, la secuencia de referencia seleccionada, un identificador de usuario obtenido del usuario, y la información acerca de los códigos de nucleótidos modificados por el usuario.

19. Un producto de programa informático que comprende un medio de código de programa informático para controlar uno o más procesadores de un sistema informático para validar datos de secuenciación de ADN a partir de los datos de secuencia de fragmentos de uno o más fragmentos de ADN, de forma que el sistema:

obtiene de un usuario una especificación diana que identifica una secuencia génica diana;

selecciona, de modo automático, a partir de un conjunto de una o más posibles secuencias de referencia (42) relacionadas con la especificación diana y almacenadas en la base de datos (4), una secuencia de referencia seleccionada que tenga la mayor correlación con los datos de secuencia de fragmentos;

alinea, de modo automático, los datos de secuencia de fragmentos con la secuencia de referencia seleccionada; e

identifica, de modo automático, las posiciones en la secuencia en que los códigos de nucleótidos de los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados no se corresponden.

20. El producto de programa informático según la reivindicación 19, que comprende además un medio de código de programa informático para controlar los procesadores del sistema informático, de forma que un servidor (3) del sistema obtiene del usuario la especificación diana que identifica una secuencia génica diana, a través de una red de telecomunicaciones (2); y el servidor (3) selecciona la secuencia de referencia seleccionada a partir de la base de datos (4) a partir de un conjunto de uno o más variantes de la secuencia génica diana.

21. El producto de programa informático según una de las reivindicaciones 19 ó 20, que comprende además un medio de código de programa informático para controlar los procesadores del sistema informático, de forma que el sistema genera un cóntigo como una secuencia consenso a partir de los datos de secuencia de fragmentos alineados con la secuencia de referencia seleccionada, insertando en la secuencia consenso, en las posiciones en la secuencia que tienen códigos de nucleótidos que no se corresponden en los datos de secuencia de fragmentos, un código especial que indica ambigüedad.

22. El producto de programa informático según una de las reivindicaciones 19 a 21, que incluye además un medio de código de programa informático para controlar los procesadores del sistema informático, de forma que el sistema muestra secciones, una al lado de otra, de los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados, mostrándose los datos de secuencia de fragmentos de cada fragmento de ADN a lo largo de una línea separada; indica, de modo visual, las posiciones en la secuencia con códigos de nucleótidos que no se corresponden; obtiene del usuario instrucciones para modificar un código de nucleótido en las posiciones en la secuencia que tienen códigos de nucleótidos que no se corresponden; y modifica los códigos de nucleótidos según las instrucciones obtenidas del usuario.

23. El producto de programa informático según la reivindicación 22, que comprende además un medio de código de programa informático para controlar los procesadores del sistema informático, de forma que el sistema almacena la información acerca de los códigos de nucleótidos modificados por el usuario; muestra de forma selectiva, una al lado de otra, las secciones modificadas de los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados que contienen los códigos de nucleótidos modificados por el usuario; e indica en las secciones modificadas, de modo visual, los códigos de nucleótidos modificados por el usuario.

24. El producto de programa informático según una de las reivindicaciones 19 a 23, que comprende además un medio de código de programa informático para controlar los procesadores del sistema informático, de forma que el sistema almacena en la base de datos (4) las máscaras de secuencias (43) asignadas a las secuencias de referencia (42), incluyendo cada una de las máscaras de secuencias (43) información del perfil relacionada con una o más posiciones de la respectiva secuencia de referencia; obtiene del usuario información de interés; y muestra, una al lado de otra, secciones seleccionadas de los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados, determinándose las secciones seleccionadas basándose en la información de interés obtenida del usuario y la información del perfil incluida en la máscara de secuencia asignada a la secuencia de referencia seleccionada.

ES 2 311 958 T3

5 25. El producto de programa informático según la reivindicación 24, que comprende además un medio de código de programa informático para controlar los procesadores del sistema informático, de forma que el sistema almacena cada una de las máscaras de secuencias (43) en la base de datos (4) asignada a un identificador de usuario; y determina las secciones seleccionadas basándose en la máscara de secuencia asignada a un identificador de usuario obtenido del usuario.

10 26. El producto de programa informático según una de las reivindicaciones 19 a 25, que comprende además un medio de código de programa informático para controlar los procesadores del sistema informático, de forma que el sistema muestra, una al lado de otra, secciones de los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados, mostrándose los datos de secuencia de fragmentos de cada fragmento de ADN a lo largo de líneas separadas, como una secuencia de códigos de nucleótidos y como una señal electroferográfica, ajustándose los niveles de señal de las señales electroferográficas de manera individual para los diferentes tipos de nucleótidos basándose en ajustes obtenidos del usuario.

15 27. El producto de programa informático según una de las reivindicaciones 19 a 26, que comprende además un medio de código de programa informático para controlar los procesadores del sistema informático, de forma que un servidor (3) del sistema recibe los datos de secuencia de fragmentos, generados por un secuenciador (5), a través de una red de telecomunicaciones (2); el servidor (3) alinea los datos de secuencia de fragmentos y la secuencia de referencia seleccionada; el servidor (3) genera un cóntigo como una secuencia consenso; un dispositivo visualizador (13) del sistema colocado donde se encuentra el usuario muestra los datos de secuencia de fragmentos y la secuencia de referencia seleccionada alineados; un terminal de entrada de datos (1) colocado donde se encuentra el usuario obtiene instrucciones para ajustar, en el cóntigo, un código de nucleótido; y el servidor (3) almacena el cóntigo en la base de datos (4) asignado a los datos de secuencia de fragmentos, la secuencia de referencia seleccionada, un identificador de usuario obtenido del usuario, y la información acerca de los códigos de nucleótidos modificados por el usuario.

25

30

35

40

45

50

55

60

65

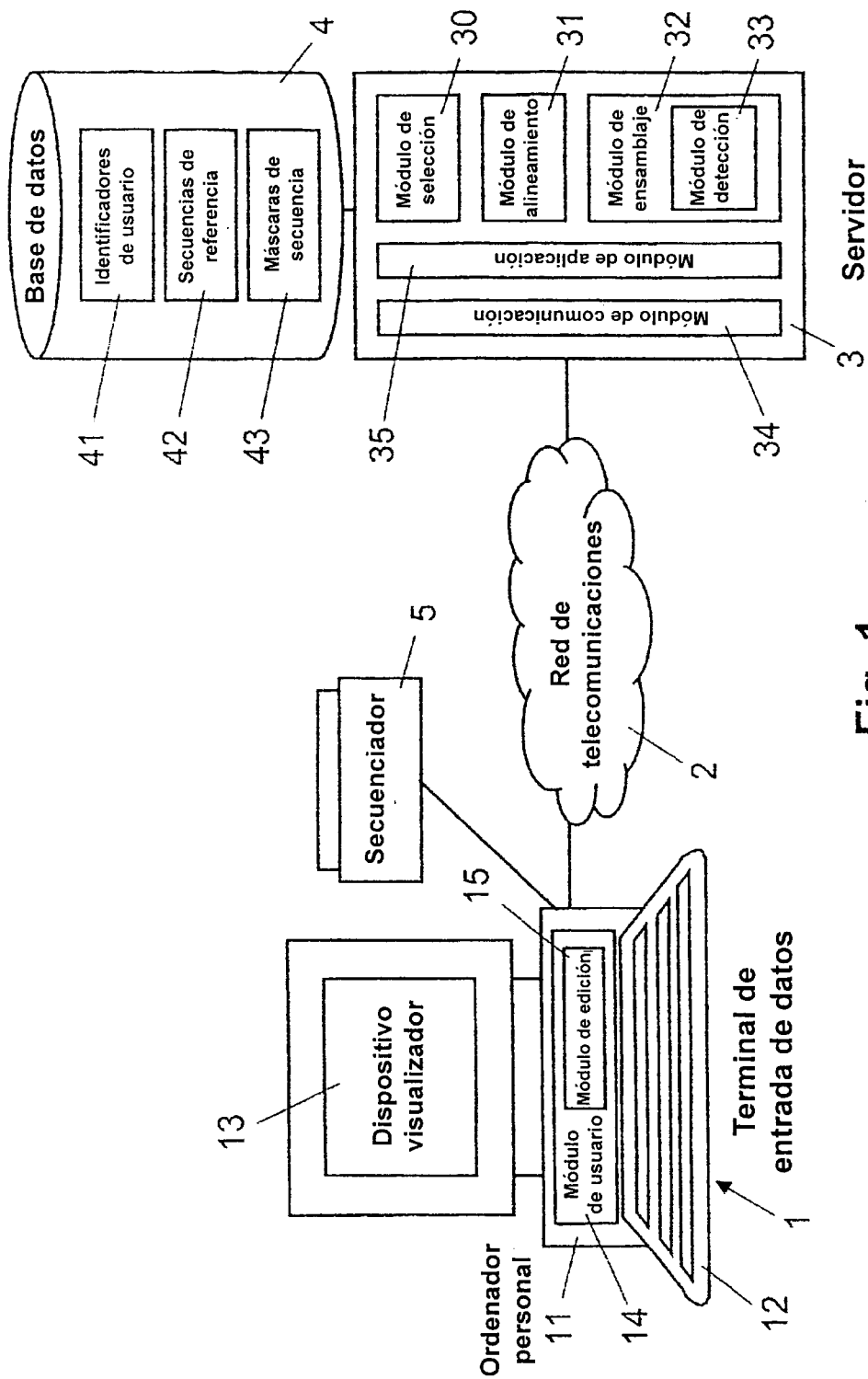


Fig. 1

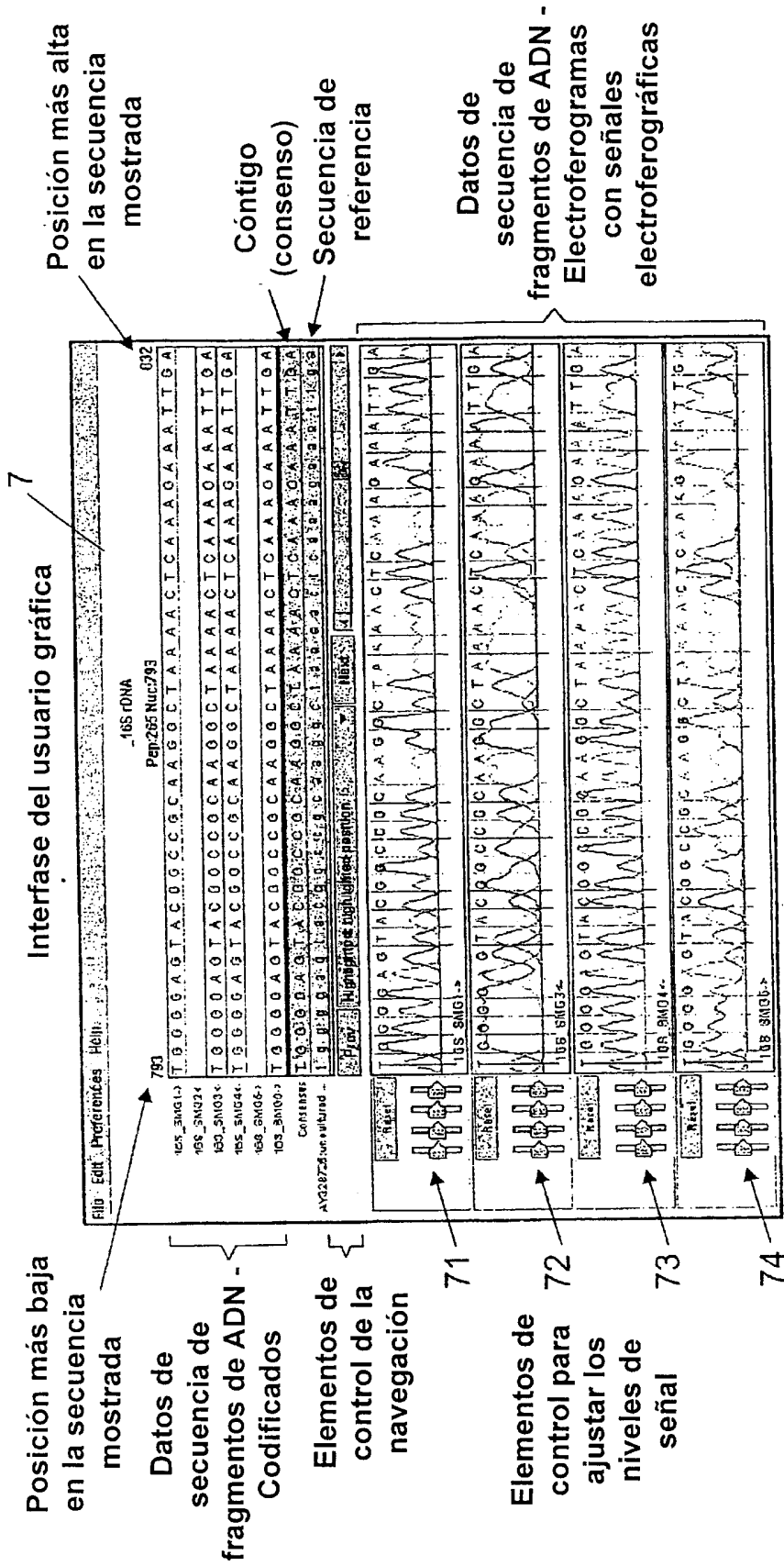


Fig. 2

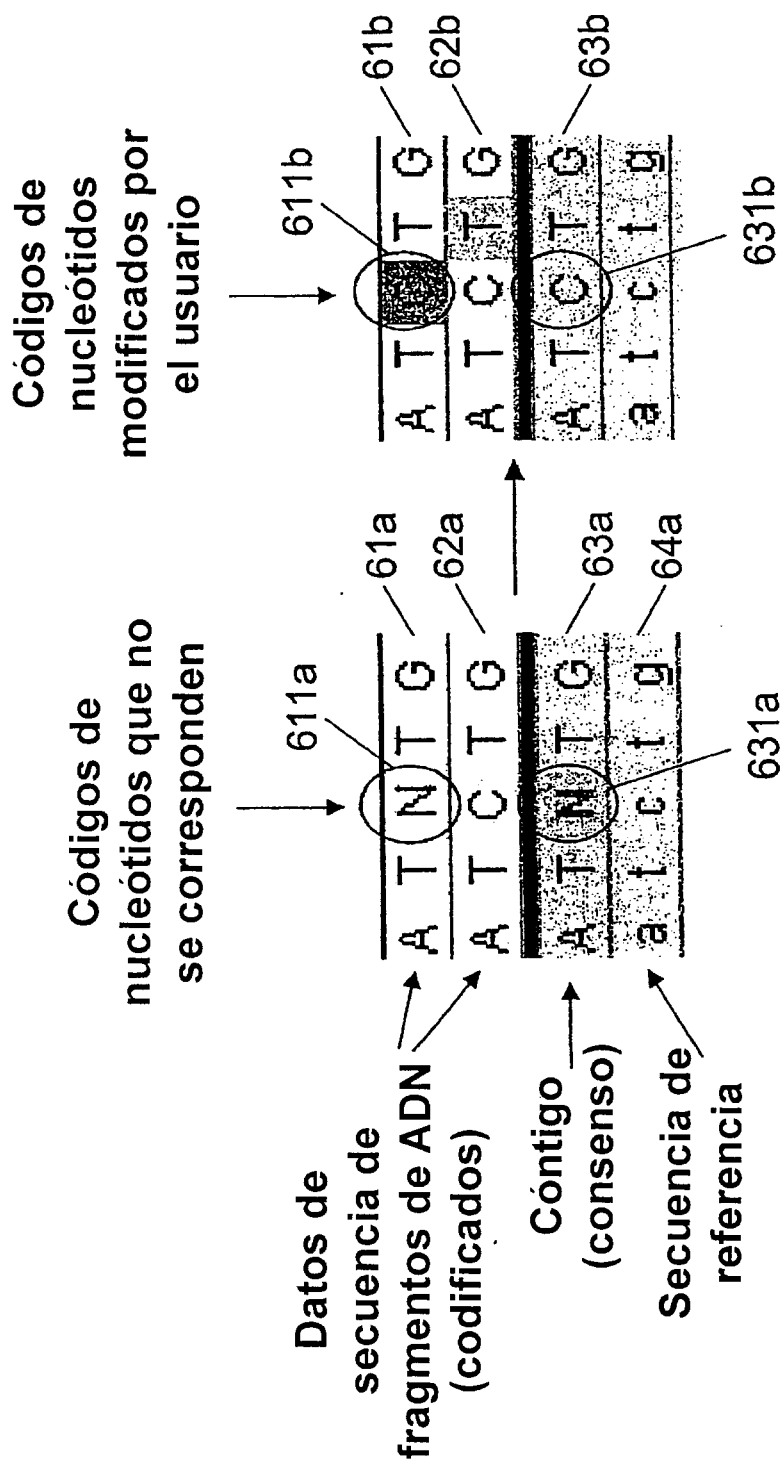


Fig. 3

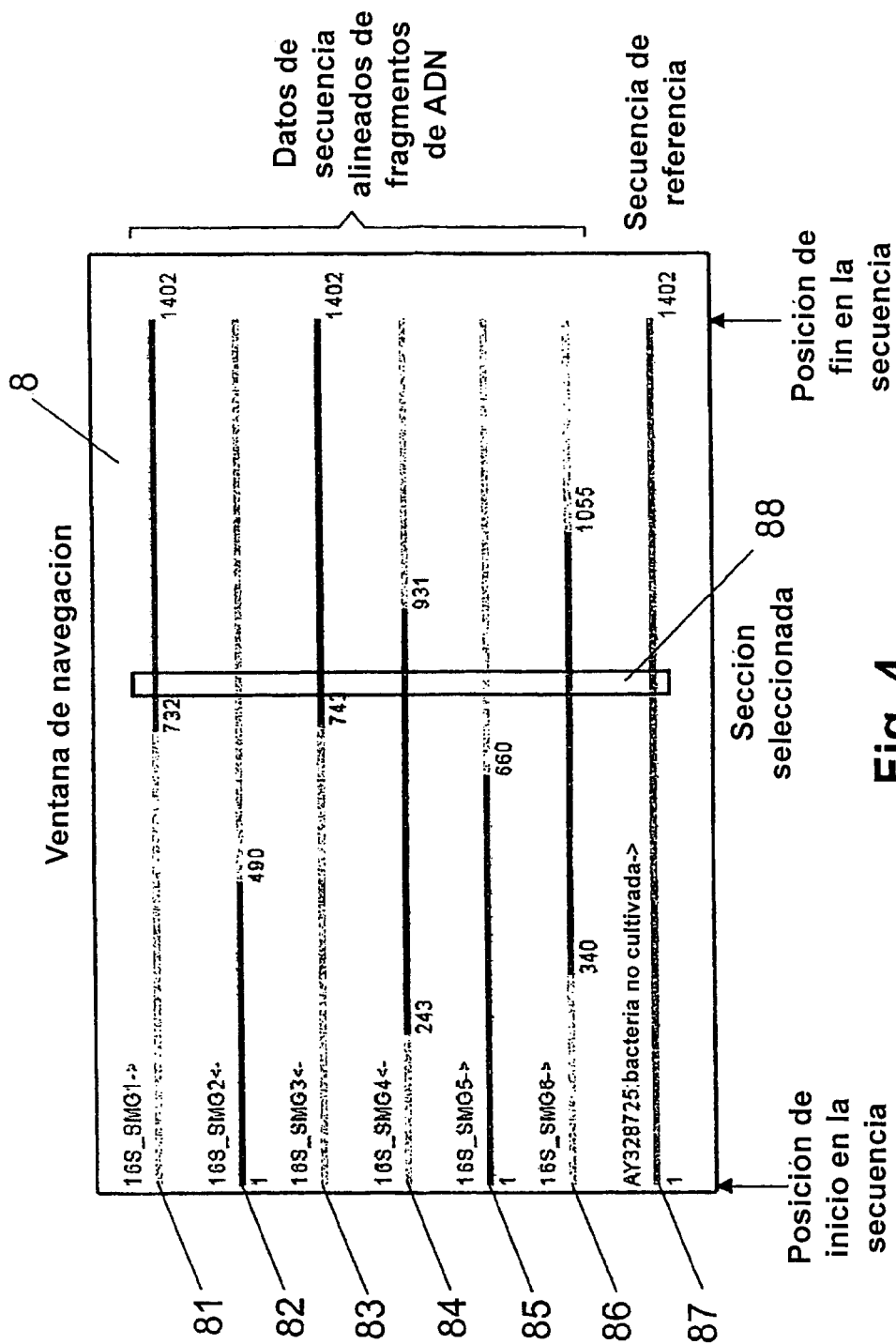


Fig. 4