



Bases de datos de genomas. Software de análisis de secuencias

Genome databases. Sequence Analysis software

 <https://doi.org/10.47230/Journal.TechInnovation.v2.n2.2023.95-105>

Recibido: 11-08-2023 **Aceptado:** 11-10-2023 **Publicado:** 01-12-2023

Yurisay Milanes Aguilar^{1*}

 <https://orcid.org/0009-0007-1977-0418>

Leidy Mariam Cruz Crespo²

 <https://orcid.org/0009-0001-308991-4801>

Itzyar Selene Rodríguez Roque³

 <https://orcid.org/0009-0001-6009-8527>

1. Facultad de Ciencias y Tecnologías Computacionales; Universidad de las Ciencias Informáticas; La Habana, Cuba.
2. Facultad de Ciencias y Tecnologías Computacionales; Universidad de las Ciencias Informáticas; La Habana, Cuba.
3. Facultad de Ciencias y Tecnologías Computacionales; Universidad de las Ciencias Informáticas; La Habana, Cuba.

Volumen: 2

Número: 2

Año: 2023

Paginación: 95-105

URL: <https://revistas.unesum.edu.ec/JTI/index.php/JTI/article/view/57>

***Correspondencia autor:** maguilar@estudiantes.uci.cu



RESUMEN

En el presente documento se discute la importancia de los sistemas de información y el soporte tecnológico, como las bases de datos, en el campo de la bioinformática. La bioinformática es una disciplina científica emergente que utiliza la tecnología de la información para organizar, analizar y distribuir información biológica para resolver problemas complejos en biología. El mismo explora las características de diferentes bases de datos genómicas y software de análisis de secuencias, y su utilidad en la identificación y análisis de la variabilidad genética en diferentes especies. Los autores buscan evaluar la eficacia de estas herramientas en la identificación y análisis de mutaciones y polimorfismos genéticos en patologías humanas y animales. El artículo proporciona una descripción general de varias bases de datos genómicas, incluyendo Oxford Academic, National Center for Biotechnology Information (NCBI), Human Genome Variation Society (HGVS) y Health Sciences Library System (HSLs). En el archivo se describen las categorías de información disponibles en cada base de datos, como bases de datos de secuencias de nucleótidos, bases de datos de secuencias de proteínas y bases de datos genómicas. También discute el papel del NCBI en la consolidación de varios genomas y en proporcionar acceso confiable y robusto a diferentes recursos, herramientas y códigos compartidos por la comunidad científica. Los autores destacan la importancia de estas tecnologías y herramientas en la facilitación y garantía de la factibilidad de la información proporcionada por ellas. En general, el artículo proporciona una descripción general de la importancia de las bases de datos genómicas y el software de análisis de secuencias en la biología moderna y la naturaleza multidisciplinaria de la bioinformática.

Palabras clave: Bioinformática, Bases de datos genómicas, NCBI, Softwares de análisis de secuencias, Variabilidad genética.

ABSTRACT

The present document discusses the importance of information systems and technological support, such as databases, in the field of bioinformatics. Bioinformatics is an emerging scientific discipline that uses information technology to organize, analyze, and distribute biological information to solve complex problems in biology. The article explores the characteristics of different genome databases and sequence analysis software, and their usefulness in identifying and analyzing genetic variability in different species. The authors aim to evaluate the effectiveness of these tools in identifying and analyzing genetic mutations and polymorphisms in human and animal diseases. The article provides an overview of various genome databases, including Oxford Academic, National Center for Biotechnology Information (NCBI), Human Genome Variation Society (HGVS), and Health Sciences Library System (HSLs). The authors describe the categories of information available in each database, such as nucleotide sequence databases, protein sequence databases, and genomics databases. The article also discusses the role of the NCBI in consolidating various genomes and providing reliable and robust access to different resources, tools, and codes shared by the scientific community. The authors highlight the importance of these technologies and tools in facilitating and ensuring the feasibility of the information provided by them. Overall, the article provides a comprehensive overview of the importance of genome databases and sequence analysis software in modern biology and the multidisciplinary nature of bioinformatics.

Keywords: Bioinformatics, Genome databases, Genetic variability, NCBI, Sequence analysis software.



Creative Commons Attribution 4.0
International (CC BY 4.0)

Introducción

Con la evolución de las tecnologías informáticas y el incremento de su utilización, los sistemas de información se han convertido en el pilar básico de muchas organizaciones. Aunque la aplicación normalmente utilizada de estos sistemas, se encuentren en el ámbito organizacional, los sistemas de información y soporte tecnológico, como las bases de datos se han extendido a otras esferas de aplicación, como los sistemas bioinformáticos. Estas técnicas han sido aplicadas a sistemas médicos, financieros, sin embargo las investigaciones no cesan y por ende se busca llevarlas a niveles superiores, más difíciles y desafiantes, como lo es la interpretación del genoma de datos (Romero, 2020), contribuyendo de esta forma al desarrollo de la bioinformática, que es una disciplina científica emergente que utiliza la tecnología de la información para organizar, analizar y distribuir información biológica para dar solución a problemas complejos en la biología, al ser un área de investigación multidisciplinaria abarcadora del estudio de más de una ciencia. Así, esta disciplina ha dado lugar al desarrollo de nuevas tecnologías, así tuvo lugar la implementación de softwares de análisis de secuencias biológicas (Gálvez, 2008; Darianis et al., 2023; González et al., 2023). En el desarrollo de este trabajo se abordarán contenidos relacionados con este tema por la transformación inminente que han representado estas tecnologías para la biología moderna y la gran importancia que tiene su conocimiento en la época actual (Cornelio et al., 2016; Cornelio & Fonseca, 2016; González & Mar Cornelio, 2013).

El objetivo de la realización de este trabajo, radica en la importancia que tiene estas tecnologías y herramientas desarrolladas para facilitar y garantizar la factibilidad de la información proporcionada por las mismas (Santos et al., 2023; Reyes et al., 2023). Explorar y comparar las características de distintas bases de datos de genomas y softwares de análisis de secuencias para

identificar su utilidad en el análisis de la variabilidad genética en distintas especies (Martínez-Vaz, 2014). A través de esta investigación se pretende lograr la comprensión de los futuros ingenieros en bioinformática acerca del papel de estas herramientas en su desarrollo como futuros profesionales y su utilización como fuentes de conocimientos en el desarrollo de nuevos descubrimientos.

Metodología

La investigación realizada se sustenta en las bases de la revisión bibliográfica de artículos científicos relacionados con el tema, para comprender el funcionamiento de diversas bases de datos y software de análisis de secuencias desarrollados como herramientas bioinformáticas. La selección de cada artículo se realizó de acuerdo a su relevancia y calidad de los mismos. El trabajo se encuentra dividido en epígrafes para facilitar la búsqueda del lector, proporcionando un enfoque sistemático y organizado a la hora de presentar la información (Vivas Tesón, 2015).

Los métodos y técnicas seleccionados para desarrollar la información se basaron en la necesidad de recopilar información precisa y relevante sobre el contenido que abarca el contenido de este trabajo, por la necesidad de conocimiento existente acerca de las herramientas bioinformáticas para el análisis de secuencias biológicas para los estudiantes de la carrera ingeniería en bioinformática en Cuba, por lo novedoso de esta carrera, así como para enriquecer el desempeño laboral de los trabajadores que desempeñan su labor en los centros de investigación del país u otros interesados en estos temas tan innovadores en la actualidad.

En la actualidad el manejo de bases de datos se ha vuelto muy común, por lo que fácilmente se encuentra información recopilada en las diferentes áreas de trabajo. Es así como, en el ámbito genómico existe una gran variedad de bases de datos genómicas.

Una base de datos genómica, o conocida también como base de datos del genoma, es el repositorio en el cual se puede realizar el mapeo genómico resultante de lo encontrado en las pruebas de laboratorio (Silva Logroño et al., 2010). Actualmente las bases de datos genómicas son herramientas de vital importancia ya que permiten realizar investigaciones y estudios en base al contenido o enfoque que se ha dado dentro del contenido de estas. Además contienen información detallada sobre la secuencia de ADN de los organismos, incluyendo información sobre genes, proteínas, regiones reguladoras y otras características genómicas (Sanz et al., 2006).

Con las bases de las investigaciones realizadas se han generado bases de datos (Melo et al., 2016). Con el paso del tiempo se han generado diversas bases de datos que contienen información genómica, así, estas bases de datos, e99stán disponibles en listados para su acceso público o privado (de pago o algún tipo de membresía).

Resultados

Algunas de estas bases de datos son: los listados proporcionados por Oxford Academic, National Center for Biotechnology Information (NCBI), Human Genome Variation Society (HGVS) y Health Sciences Library System (HSLs).

Oxford Academic : La Oxford University Press es un departamento de la Universidad de Oxford, que tiene una sección llamada NAR (Nucleic Acid Research) destinada para la recepción y publicación de información científica orientada a los resultados de investigaciones de vanguardia sobre los aspectos físicos, químicos, bioquímicos y biológicos de los ácidos nucleicos y las proteínas implicadas en el metabolismo y las interacciones de los ácidos nucleicos Oxford University Press y los Editores de NAR lanzaron una iniciativa de acceso abierto para NAR en 2005. Actualmente NAR presenta un total de 1610 bases de datos, las cuales presentan una categorización, permitiendo

tener un gran número de datos para cada categoría.

Entre estas categorías se encuentran:

- Nucleotides Sequence Databases, que contiene información de secuencias de nucleótidos, ADN codificado y no codificado, estructura de genes, intrones y exones, sitios de empalme, factores de transcripción.
- RNA Sequence Databases , que se refiere a secuencias de RNA (Ácido Ribonucleico).
- Protein Sequence Databases, que posee información de secuencias generales, propiedades de las proteínas, localización de proteínas, clasificación de proteínas, familias de proteínas. Dentro de esta categoría resalta la llamada Protein Data Bank (PDB) el cual es el único repositorio global de estructuras tridimensionales de macromoléculas biológicas y sus complejos determinados experimentalmente. Este, se convirtió en el primer recurso digital de acceso abierto en las ciencias biológicas. De esta forma, actualmente alberga en su archivo alrededor de 130,000 entradas y es administrado por la organización Worldwide Protein Data Bank (wwPDB), que incluye a RCSB Protein Data Bank (RCSB PDB), Protein Data Bank Japan (PDBj), Protein Data Bank in Europe (PDBe) y BioMagResBank (BMRB. En la actualidad, el PDB archiva datos experimentales, metadatos asociados y modelos estructurales tridimensionales a nivel atómico derivados de tres métodos bien establecidos: cristalografía, espectroscopía de resonancia magnética nuclear (NMR) y microscopía electrónica (3DEM). - Structure Databases, que tiene información de moléculas pequeñas, hidratos de carbono, estructuras de ácidos nucleicos y estructuras proteicas.
- Genomics Databases (NONVERTEBRATE), esta categoría presenta información

de términos de anotación del genoma, ontologías y nomenclaturas, taxonomía, genomas virales, genomas procarióticos, genomas de eucariotas unicelulares, genomas fúngicos, genomas de invertebrados y genómica comparativa.

- Metabolic and Signaling Pathways, que indica información de enzimas, nomenclatura enzimática, rutas metabólicas, interacciones entre proteínas.
- Human and other Vertebrate Genomes, esta categoría brinda información de genoma de invertebrados, genómica comparativa, mapas y visores del genoma humano.
- Microarray Data And Other Gene Expression Databases, la cual contiene información general de genética humana, polimorfismo, datos específicos de genes, sistemas o enfermedades.
- Proteomics resources, que presenta información de recursos de proteómica.
- Other Molecular Biology Databases, que incluye base de datos de literatura, drogas, sondas moleculares.
- Organelle Databases, que asocia información de genes y proteínas mitocondriales.
- Plant Databases, que indica información de bases de datos de plantas, Arabidopsis thaliana, arroz entre otras.
- Immunological Databases, que contiene información referente a inmunología.
- Cell Biology, que presenta información de biología celular. National Center For Biotechnology Information.

El Centro Nacional de Información Biotecnológica (NCBI) inició con la finalidad de identificar los registros de GenBank, que era donde se recopilaba la secuencia del genoma de manera aceptada por un organismo particular. NCBI asume la responsabilidad de la base de datos de secuencias

de ADN de GenBank en octubre de 1992. A partir de la consolidación de varios genomas de los investigadores, NCBI permitió que se pueda de manera confiable y robusta la descarga, visualización o análisis de diferentes recursos, herramientas y códigos compartidos por la comunidad científica. Actualmente se puede encontrar material como genes, genomas, secuencias de referencia, anotaciones de genomas eucarióticos o procarióticos, variaciones, entre otros. Algunos recursos reflejan los principios organizadores naturales de datos genómicos y apoyan el acceso desde una perspectiva centrada en genes u organismos, mientras que otros se desarrollaron en respuesta a un brote de enfermedad particular o en colaboración, comentarios o solicitudes de la comunidad. A lo largo del camino también se desarrolló un conjunto de plataformas de visualización que incluyen Visores como Map Viewer, visor de secuencia gráfica y una interfaz gráfica de usuario Genome Workbench que puede ser descargado y es multiplataforma. Actualmente NCBI contiene 41 bases de datos las cuales han sido categorizadas, siendo así que abarca diferentes ámbitos en la que los investigadores de manera pública pueden acceder a los contenidos. El personal de NCBI tiene capacitación avanzada en biología molecular, que son quienes construyen la base de datos a partir de secuencias enviadas por laboratorios individuales y mediante el intercambio de datos con las bases de datos de secuencias de nucleótidos internacionales, European Molecular Biology Laboratory (EMBL) Base de datos de ADN de Japón (DDBJ).

Literature Orientada al contenido de libros, reportes o resúmenes que indican resultados vinculados con el genoma. GENES Posee información sobre secuencias etiquetadas, transcripciones, expresiones génicas, genes homólogos, estudios filogenéticos y de población. GENETICS Se relaciona a variaciones humanas, genotipos, fenotipos, variación estructural del genoma, registros

de pruebas genéticas, herencia mendeliana en el hombre.

Proteins Posee información de dominios de proteínas conservadas, secuencias de proteínas, estructuras biomoleculares.

Genomes Contiene información de secuencias de AND y ARN, secuencias genómicas, taxonomía, nomenclaturas.

Chemicals Presenta información de rutas metabólicas, estudios de detección de bioactividad, información química con estructuras, información química.

La Sociedad de Variación del Genoma Humano, Human Genome Variation Society (HGVS) es un afiliado de la Federación Internacional de Sociedades de Genética Humana (IFHGS) y también de la Organización del Genoma Humano (HUGO), que tiene como objetivo fomentar el descubrimiento y la caracterización de las variaciones genómicas, así como también la distribución de la población y las asociaciones fenotípicas. Además, promueve la recopilación, documentación y distribución gratuita de la información de variación genómica y las variaciones clínicas asociadas. Otro de sus objetivos es fomentar el desarrollo de la metodología y herramientas necesarias para ayudar en la investigación. La misión de HGVS es mejorar la salud humana a través de la identificación y caracterización de los cambios en el genoma que conducen a la susceptibilidad de la enfermedad. De esta manera HGVS permite cotejar la información genómica necesaria para el diagnóstico molecular, la investigación sobre los mecanismos básicos y el diseño de los tratamientos de las dolencias humanas.

Health Sciences Library System

-El Sistema de Bibliotecas de Ciencias de la Salud (HSLs) de la Universidad de Pittsburgh ofrece una amplia gama de servicios de información, oportunidades educativas y recursos en formato impreso y electrónico para profesores, estudiantes e investigado-

res en las escuelas de ciencias de la salud (Medicina, Medicina Dental, Farmacia, Enfermería, Ciencias de la Salud y Rehabilitación, y Salud Pública). Es así como en su sitio web presenta un listado de bases de datos relacionados con el ámbito genómico que son de acceso público. Actualmente los repositorios de HSLs posee información de investigaciones sobre temas de salud y biomédicos, revisiones sistemáticas y otras búsquedas avanzadas de literatura, gestión de datos de investigación, recursos de biología molecular y herramientas de software, entre otros. El HSLs maneja, en la actualidad 2458 bases de datos.

Existen otras bases de datos utilizadas en la actualidad, que también contribuyen al desarrollo de las ramas de investigación relacionadas con estos temas.

- Completed genomes at EBI: sitio del European Bioinformatics Institute.
- The Wellcome Trust Sanger Institute, que, como TIGR, ha participado activamente en la secuenciación de muchos genomas.
- Ensembl, genomas eucariotas
- SGD, Saccharomyces Genome Database, levadura.
- FlyDB, dedicada a Drosophila.

El Instituto Pasteur, mantiene varias bases de datos específicas de ciertas bacterias, donde todas tienen una interfaz gráfica parecida, bien diseñada y fácil de usar. Además de permitir la extracción de información sobre genes y de mostrarla en forma gráfica, tiene la posibilidad de hacer búsquedas de motivos, permitiendo, por ejemplo, la identificación de sitios de unión de proteínas reguladoras (mediante la opción "Search Pattern"). Permitiendo el análisis de bacterias como:

- Subtilist, *Bacillus subtilis*.
- Colibri, *Escherichia coli*
- Tuberculist, *Mycobacterium tuberculosis*
- MypuList, *Mycoplasma pulmonis*
- Leproma, *Mycobacterium leprae*

También existen bases de datos especializadas en genómica comparativa:

Así como las herramientas que se han analizado hasta ahora iban principalmente encaminadas a predecir la función de nuestras secuencias, las comparaciones nos ofrecen informaciones adicionales: pertenencia a familias conocidas, perfil filogenético (presencia en otros organismos), pertenencia a operones o clusters de genes (genes que se presentan agrupados en diferentes organismos, lo que a menudo tiene implicaciones funcionales), etc.

La información comparativa puede incluso ayudarnos a predecir la función para ORFs de función desconocida en el organismo que analizamos: comparando diversos organismos podemos conocer aquellos genes/proteínas que realicen funciones esenciales y que no hayan sido descubiertas en este. Las bases de datos de metabolismo son de considerable ayuda en este punto. La información posicional también es muy importante: la función de algunos ORFs puede conocerse de acuerdo a su vecindad con ORFs de función conocida, si esta disposición está conservada en diferentes organismos.

Algunas bases de datos especialmente orientadas al estudio comparativo de genomas son:

COGs. Esta base de datos del NCBI consiste en unas 2800 familias de proteínas conservadas o COGs, que contienen grupos de proteínas ortólogas. Por definición, dos proteínas, o genes, son ortólogos si han evolucionado a partir de un mismo ancestro. Típicamente, genes o proteínas ortólogas tienen la misma función; no tiene por qué ocurrir lo

mismo entre genes o proteínas parálogas, que en vez de haber sido originadas por especiación (divergencia entre especies), han sido originadas por duplicación de un gen y posterior evolución hasta, a veces, desarrollar una nueva funcionalidad. Dado que cada COG incluye proteínas de organismos filogenéticamente muy diversos, este sitio es especialmente útil para asignar función a proteínas que no muestran un grado significativo de similitud con otras proteínas. Para ello, el usuario tiene a su disposición de enviar secuencias a través del interfaz COGNITOR, que serán analizadas y asignadas a alguno de los COGs existentes. Otras herramientas disponibles en COGs permiten analizar la distribución de genes en genomas (Phylogenetic patterns) o analizar semejanzas entre genomas mediante análisis de componentes principales.

KEGG. Esta es la Kyoto Encyclopedia of Genes and Genomes, y está centrada fundamentalmente en la comparación de genomas en cuanto a su capacidad para codificar diferentes rutas metabólicas. El usuario puede obtener mapas metabólicos e identificar qué enzimas están presuntamente codificadas por un genoma en concreto. Otro tipo de análisis posible con KEGG es la identificación de clusters de genes (o grupos de genes consecutivos) conservados entre dos especies.

MBGD. La Microbial Genome Database es otro sitio muy útil para identificar genes o proteínas homólogas, en varios genomas.

Pedant es un sistema para análisis de genomas. Aunque no es accesible desde el exterior (no podemos analizar nuestras secuencias), si podemos examinar los análisis ya realizados para gran cantidad de organismos, incluso algunos aun no completos.

String. Es una base de datos en la que se integran predicciones sobre interacciones funcionales, basadas en tres técnicas que hacen uso de comparación de genomas: conservación del contexto génico, fusiones génicas y perfiles filogenéticos.

Para analizar y comprender la información biológica, se utilizan diversas técnicas y herramientas como es conocido, dentro de estas herramientas utilizadas, los softwares bioinformáticos especializados en el análisis de secuencias biológicas son muy empleados. Estos softwares son esenciales para el avance de la investigación en biología, medicina y biotecnología, ya que permiten la identificación de patrones y relaciones en la información genética y la comprensión de la estructura y función de los genes.

Antheptot 2000 V6.0 release 1.1.54, es una herramienta para el análisis de proteínas.

BioEdit 7.0.9 es un programa que contiene múltiples herramientas para el análisis de secuencias biológicas.

Blast 2.2.21 ia32 win32, es una versión "standalone" del programa BLAST para la búsqueda en bases de datos.

Chroma, es un programa para la edición de alineamientos múltiples de secuencias.

Cinema 5.021 beta, es un programa para la edición de alineamientos múltiples.

Clustalw 2.0.11, es un programa para calcular alineamientos múltiples. Versión con interfaz en modo textual (Para DOS y XP).

Clustalx 2.0.11, es un programa para calcular alineamientos múltiples. Versión con interfaz gráfica para Windows.

Cn3D 4.1, es un programa para la modelación de estructuras de biomoléculas.

DAMBE 5.0.86 para Windows XP, es un programa para el análisis de datos en Biología Molecular y Evolución.

DeepView 4.0.1, es un visor de estructuras de SwissProt. Programa avanzado para mostrar archivos PDB y llevar a cabo alineamiento estructural. DnaSP 4.20.2 es un programa para estudiar polimorfismos de datos de secuencias de DNA para estudios de genética de poblaciones y evolución molecular.

DSSP, es un programa diseñado por Wolfgang Kabsch y Chris Sander para estandarizar la asignación de estructura secundaria a partir de coordenadas 3D.

GeneDoc 2.7, es un programa para la edición de alineamientos múltiples.

HMMER 2.3.2, es un conjunto de programas para la creación de modelos ocultos de Markov. Incluye varias herramientas para el análisis de secuencias, entre ellas herramientas para el alineamiento múltiple. Versión compilada para Windows.

Jalview editor versión 2.2.1, es un programa para editar alineamientos múltiples de secuencias.

JModelTest, es un nuevo programa para la selección de un modelo filogenético empleando PhymI. Implementa los métodos hLRTs, dLRTs, AIC, BIC y DT methods.

Macawz, es un programa para el alineamiento de secuencias mediante bloques de similitud localizada.

Mage Display 6.44 es un programa para visualizar imágenes de moléculas ("kinameges").

MEGA 4.0, es un programa para análisis filogenético.

Melting4.3b, es un programa para analizar la estabilidad de moléculas de ácidos nucleicos.

Microarrays Einsen, es un conjunto de herramientas del Lawrence Berkeley National Lab para el análisis de datos de microarreglos.

MODELLER 9v7, es un programa para modelación por homología de estructuras MrBayes 3.1.2 <http://mrbayes.csit.fsu.edu/download.php>

Programa para la estimación bayesiana de filogenia NetBlast 2.2.21. ia32 win32 Este programa es una versión cliente del Blast del NCBI con algunas propiedades impor-

tantes para someter la búsqueda de gran número de secuencias, que no son proporcionadas por la versión WEB de Blast.

OligoArray 2.1 y OligoArrayAux 3.7, OligoArray permite el diseño de sondas de ácidos nucleicos para microarreglos empleados en estudios de expresión diferencial. OligoArrayAux es una extensión del programa mFold de M. Zuker empleada para la predicción de estructura secundaria de ácidos nucleicos, empleada por el programa OligoArray durante el diseño de sondas. PAML 4.3 ("Phylogenetic Analysis by Maximum Likelihood" o análisis filogenéticos de secuencias de ADN y proteínas utilizando máxima verosimilitud), es un programa para calcular árboles filogenéticos por el métodos de la máxima probabilidad (verosimilitud).

Son muchas las herramientas desarrolladas para la facilitación del trabajo en investigaciones bioinformáticas.

Una de estas tecnologías, se ha materializado a través de las llamadas redes neuronales, las cuales son un método de la inteligencia artificial ayuda a las computadoras a procesar datos de una manera inspirada en la forma en que lo hace el cerebro humano. Un sistema de red neuronal basado en el conocimiento neuronal fue aplicado al análisis de la secuencia de ADN. Existen dos buscadores de genes, que dieron lugar a las Redes Neuronales Artificiales. GRAIL, es el primer programa buscador de genes, diseñado para identificar genes, exones y varias características en las secuencias de ADN, éste programa, utiliza una red neuronal, que combina una serie de algoritmos de codificación de predicción para reconocer el potencial de codificación en ventanas de longitud fija sin buscar características adicionales.

Otro sistema de buscador de genes es Gene Parser, que fue diseñado para identificar y determinar la fina estructura de los genes de la proteína en las secuencias de ADN genómico. Un sistema neuronal arti-

ficial para clasificación de genes llamado GenCANS fue desarrollado para analizar y gestionar un gran volumen de datos de secuenciación molecular del Proyecto del Genoma Humano. El algoritmo genético ha sido aplicado con éxito para resolver muchos problemas prácticos en muchas disciplinas, en particular, en la bioinformática.

Las bases de datos del genoma y el software bioinformático son herramientas fundamentales para la investigación en biología y biotecnología. Estas herramientas permiten la identificación de patrones y relaciones en la información genética y la comprensión de la estructura y función de los genes. Los ingenieros en bioinformática y los investigadores en biotecnología utilizan estas herramientas para analizar grandes cantidades de secuencias biológicas y datos del genoma, lo que les permite realizar importantes descubrimientos sobre biología molecular y genética.

Las bases de datos del genoma contienen información completa sobre el genoma de diferentes organismos, lo que permite a los investigadores acceder a información sobre la estructura y función de los genes. El software de bioinformática, por otro lado, permite el análisis e interpretación de datos de genomas y secuencias biológicas. Estos softwares utilizan algoritmos y técnicas estadísticas para identificar patrones y relaciones en los datos, lo que permite a los investigadores realizar importantes descubrimientos en biología molecular y genética.

Además, las redes neuronales son herramientas poderosas para el análisis de datos de genomas y secuencias biológicas. Estas redes son capaces de aprender patrones complejos en los datos, lo que les permite identificar relaciones y hacer predicciones precisas sobre la estructura y función de los genes. Los ingenieros en bioinformática y los investigadores en biotecnología utilizan estas redes para analizar grandes cantidades de datos de secuencias biológicas y genómicas, lo que les permite realizar im-

portantes descubrimientos en biología molecular y genética.

En definitiva, las bases de datos genómicas y el software bioinformático, incluidas las redes neuronales, son herramientas fundamentales para la investigación en biología y biotecnología. Estas herramientas permiten a los ingenieros en bioinformática y a los investigadores en biotecnología analizar grandes cantidades de datos de secuencias biológicas y genómicas, lo que les permite realizar importantes descubrimientos sobre biología molecular y genética.

El trabajo presentado, constituye una valiosa contribución al campo de la bioinformática, pues proporciona información detallada sobre diferentes bases de datos genómicas y softwares de análisis de secuencias, y cómo estos pueden ser útiles en el análisis de la variabilidad genética en distintas especies haciendo posible facilitar las diferentes actividades investigativas, de esta forma el presente trabajo ha permitido:

La identificación de nuevas herramientas y bases de datos genómicas: El trabajo presenta una amplia variedad de bases de datos genómicas y softwares de análisis de secuencias que pueden ser útiles en el análisis de la variabilidad genética en distintas especies. Esto puede ayudar a los investigadores a identificar nuevas herramientas y bases de datos que pueden ser útiles en sus investigaciones.

La comprensión de las mejoras en la eficacia del análisis de la variabilidad genética, pues el trabajo evalúa la eficacia de diferentes herramientas y bases de datos en la identificación y análisis de mutaciones y polimorfismos genéticos en patologías humanas y animales. Esto puede ayudar a los investigadores a mejorar la eficacia del análisis de la variabilidad genética y, por lo tanto, a mejorar la comprensión de las enfermedades genéticas.

Contribución al desarrollo de la bioinformática: el conocimiento de estas herramientas

puede contribuir al desarrollo de la bioinformática, esto se debe a que proporciona información detallada sobre diferentes bases de datos genómicas y softwares de análisis de secuencias. Esto puede ayudar a los investigadores a mejorar sus investigaciones y a desarrollar nuevas herramientas y bases de datos genómicas.

Conclusiones

En este trabajo se exploraron diferentes bases de datos genómicas y herramientas de análisis de secuencias, evaluando su utilidad en la identificación y análisis de mutaciones y polimorfismos genéticos en patologías humanas y animales. La existencia de diversas bases de datos genómicas y herramientas de análisis de secuencias disponibles, cada una con sus propias características y ventajas. A través de lo antes expuesto acerca de este tema tan novedoso, queda demostrada la necesidad de estudio de las mismas, pues su contribución en las investigaciones científicas es evidente. Por tanto es necesaria la continuación de las investigaciones que propicien el surgimiento de nuevas técnicas relacionadas que contribuyan al desarrollo biotecnológico. Es necesario destacar que uno de los principales objetivos de la realización de esta revisión bibliográfica fue lograr el conocimiento de estos métodos de estudio con respecto a los ingenieros en bioinformática que se forman hoy en Cuba, pues la necesidad de crear sitios como los antes mencionados en el país es inminente, pues el bloqueo económico y financiero impide que se desarrollen en su totalidad las investigaciones. Por lo que es necesaria la creación de herramientas bioinformáticas propiamente cubanas para impedir la ralentización de las presentes y futuras investigaciones que desempeñen nuestros profesionales.

Bibliografía

- Cornelio, O. M., Ching, I. S., Fonseca, B. B., & Díaz, P. M. P. (2016). Herramienta para la simulación de sistemas dinámicos integrado al sistema de laboratorios virtuales ya distancia. *Anais do Encontro Virtual de Documentação em Software Livre e Congresso Internacional de Linguagem e Tecnologia Online*,
- Cornelio, O. M., & Fonseca, B. B. (2016). Procedimiento multicriterio multiexperto para determinar el índice de control de una organización. *Scientia et Technica*, 21(3), 234-238.
- Darianis, P., Milané, M., & Cornelio, M. (2023). Diseño de sistema basado en reglas para apoyar la toma de decisiones de la población en la obtención de los medicamentos. *Revista Cubana de Informática Médica*, 15(2), 623.
- Gálvez, C. (2008). Minería de textos: la nueva generación de análisis de literatura científica en biología molecular y genómica. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*(25), 1-14. <https://www.redalyc.org/pdf/147/14702502.pdf>
- González, E. R., Cornelio, O. M., García, A. L. G., & Fonseca, B. B. (2023). Herramientas computacionales para el apoyo al diagnóstico de pacientes con Parkinson: una revisión sistemática. *Revista Cubana de Ciencias Informáticas*, 17(3).
- González, J. L. G., & Mar Cornelio, O. (2013). Propuesta de algoritmo de clasificación genética. *Revista Cubana de Ingeniería*, 4(2), 37-42.
- Martínez-Vaz, B. M. (2014). Conceptos claves, laboratorios de investigación y bases de datos: estrategias para la enseñanza de Bioquímica en siglo XXI. *Química Viva*, 13(1), 5-17. <https://www.redalyc.org/pdf/863/86330863004.pdf>
- Melo, S. R., Draghi, D., & Saldivia, F. L. (2016). Enseñando geometría utilizando el software dinámico GeoGebra. *Informe Científico Técnico UNPA*, 8(1), 221-244. <https://dialnet.unirioja.es/descarga/articulo/5454195.pdf>
- Reyes, L. B., Suárez, J. E., & Cornelio, O. M. (2023). Técnicas de Inteligencia artificial para el diagnóstico de pulsioximetría de apnea de sueño. *Serie Científica De La Universidad De Las Ciencias Informáticas*, 16(4), 1-10.
- Romero, P. E. (2020). Escasa información genómica en bases de datos públicas para investigar el SARS-CoV-2 en Latinoamérica. *Revista Peruana de Medicina Experimental y Salud Publica*, 37, 374-374.
- Santos, L. A., Fonseca, B. B., & Hernández, K. D. (2023). Sistema para la gestión de información como de apoyo al diagnóstico médico basado en mapa cognitivo difuso. *Revista Científica Arbitrada Multidisciplinaria PENTACIENCIAS*, 5(2), 145-158.
- Sanz, R., Bodí, V., Sanchís, J., Moratal, D., Núñez, J., Palau, P., García, D., Rieta, J. J., Sanchís, J. M., & Chorro, F. J. (2006). Desarrollo de software para la reconstrucción tridimensional y cuantificación automática de secuencias de ultrasonido intravascular. *Experiencia inicial. Revista Española de Cardiología*, 59(9), 879-888. https://www.academia.edu/download/45186680/Desarrollo_de_software_para_la_reconstru20160428-111560-1mtqfk.pdf
- Silva Logroño, J. F., Berdun, L. S., Armentano, M. G., & Amandi, A. (2010). Análisis de secuencias discretas para la detección de Patrones de Diseño de Software. *Simposio Argentino de Inteligencia Artificial (ASAI 2010)-JAIIO 39 (UADE, 30 de agosto al 3 de septiembre de 2010)*,
- Vivas Tesón, I. (2015). La internacionalización de las bases de datos genéticos: Beneficios y riesgos. *Cadernos Ibero-Americanos de Direito Sanitário*, 4(3), 128-142. <https://www.cadernos.prodisa.fiocruz.br/index.php/cadernos/article/download/202/368>

Cómo citar: Milanes Aguilar, Y., Cruz Crespo, L. M., & Rodríguez Roque, I. S. (2023). Bases de datos de genomas. Software de análisis de secuencias. *Journal TechInnovation*, 2(2), 95-105. Recuperado a partir de <https://revistas.unesum.edu.ec/JTI/index.php/JTI/article/view/57>