

SLD036 MÓDULO PARA EL ANÁLISIS DE SECUENCIAS DE LA PLATAFORMA DE SERVICIOS BIOINFORMÁTICOS

SLD036 SEQUENCE ANALYSIS MODULE OF THE WEB PLATFORM OF BIOINFORMATICS SERVICES

Andry Daniel Díaz León¹, Orlando Martínez Pérez¹, Alina Agramonte Delgado¹

¹Universidad de las Ciencias Informáticas, Cuba, addiaz@uci.cu, 97000.

²Universidad de las Ciencias Informáticas, Cuba, orlandomp@uci.cu.

³Universidad de las Ciencias Informáticas, Cuba, alinaad@uci.cu.

RESUMEN: *En el presente trabajo se propone un módulo integrado a la Plataforma Web de Servicios Informáticos que permite mediante un Portlet realizar el análisis de secuencias biológicas. El objetivo del módulo es permitir a especialistas en Bioinformática utilizar herramientas para realizar alineamientos de secuencia mediante varios programas como clustalw, t-coffee y muscle, conocer la información contenida en una determinadas secuencias, comparar unas con otras, mostrar gráficamente la información biológica contenida en las mismas, así como permitir su edición.*

Palabras claves: análisis de secuencia, alineamiento de secuencias, Portlet, servicios bioinformáticos

ABSTRACT: *This work proposes an integrated module of the Computer Services Web Platform through a portlet that allows to perform the analysis of biological sequences. The module aims to enable bioinformatics specialists to use tools for sequence alignment through several programs such as ClustalW, T-Coffee and muscle, know the information contained in sequences, to compare with each other, show graphically the biological information contained in them and allows edit them.*

Keywords: sequence analysis, sequence alignment, portlet, bioinformatics services

1. INTRODUCCIÓN

El desarrollo de la ciencia y la técnica aparejado con la innovación a logrado que el hombre en el devenir de la historia haya evolucionado hasta llegar a la sociedad actual, una sociedad donde los avances tecnológicos marcan nuevos paradigmas en todos sus ramas y avanzan a velocidades increíbles, estos avances han impactado en el modelo económico mundial y son transversales al desarrollo de las naciones.

Las Tecnologías de la Información y las Comunicaciones o TICs son aplicables en casi todos los sectores de la sociedad reportando mejoras y eficiencia. La ciencia se ha apoyado en los avances tecnológicos para aplicar los conocimientos y aprovecharlos para el progreso de la humanidad. Ciencias como la física, química y biología encuentran en las TICs la posibilidad de incrementar, mejorar y compartir su gama de conocimientos, y por ende contribuir al desarrollo de las mismas. El aprovechamiento de los recursos tecnológicos en el campo científico dio origen a nuevas ramas dentro de la ciencia como por ejemplo la Bioinformática.

La Bioinformática es un campo más grande que la biología computacional, constituye la aplicación de técnicas cuantitativas de análisis en el modelado de sistemas biológicos. Es una disciplina que evoluciona rápidamente en las que se aplican herramientas computacionales y técnicas para el manejo y análisis de datos biológicos. La meta de esta disciplina científica es permitir el descubrimiento de nuevas ideas, así como ofrecer una perspectiva global a partir de la cual se puedan discernir nuevos principios y paradigmas.

El acelerado desarrollo de las TICs ha propiciado el auge de Internet como plataforma internacional de intercambio de información, son muchas las aplicaciones que podemos encontrar en lo que también es llamado red de redes o WWW (del inglés World Wide Web), el desarrollo de herramientas basadas en la web brinda la posibilidad de utilizar las mismas sin tener en cuenta la posición geográfica en la que esté el usuario o el sistema operativo que tenga instalado, de modo que en todo momento están accesibles. Poco a poco los programas que en un principio teníamos que instalar en nuestro ordenador se han trasladado a la red de redes, y las herramientas destinadas el desarrollo de la Bioinformática no son la excepción.

Una de las actividades comunes en Bioinformática es el análisis de secuencia, se analizan genes con importantes funciones biológicas con la esperanza de aprender algo nuevo y relevante acerca de la

naturaleza química de los productos genéticos (David W. Mount, 2001). El análisis de secuencia incluye la realización de alineamientos, la muestra de estadísticas de ADN y proteínas, la búsqueda de patrones, la interconversión de formatos de secuencias entre otras actividades.

La Plataforma Web de Servicios Bioinformáticos de la Universidad de las Ciencias Informáticas (UCI) incluye herramientas desarrolladas por el departamento de Bioinformática como siRNA Design y BioSys, el objetivo del módulo de análisis de secuencia es brindar acceso a funcionalidades básicas de las más comúnmente usadas en el campo de la Bioinformática de manera que complementen la plataforma Otro aspecto relevante es que la plataforma de servicios pretende brindar acceso nacional a dichas herramientas, lo que sería de mucha utilidad en aquellos centros o universidades donde el acceso a Internet se encuentra restringido.

1.1 Herramientas para el análisis de secuencia

Actualmente se puede encontrar en Internet muchas plataformas especializadas en el área de Bioinformática que proveen a través de sus páginas web muchos servicios para el trabajo con secuencias, ejemplo de algunos de ellos son el Centro Nacional para Información Biotecnológica, NCBI (de sus siglas en inglés National Center for Biotechnology Information) disponible en <http://www.ncbi.nlm.nih.gov/>, Expasy (de sus siglas en inglés, Expert Protein Analysis System) disponible en <http://expasy.org/> y el Instituto Europeo de Bioinformática, EBI (del inglés European Bioinformatics Institute) en <http://www.ebi.ac.uk>.

1.1.1 Alineamientos de secuencias

El método más común para analizar secuencias es el alineamiento de secuencias, este provee un mapa explícito entre los residuos de dos o más secuencias (Andreas D. et al, 2001). Este método se usa para encontrar similitudes entre secuencias de ADN, ARN o estructuras proteicas.

Las secuencias de ADN y proteína marcan la función de las proteínas en los seres vivos. Cuando más similares sean dos secuencias más similares tenderán a ser las funciones de las proteínas codificadas por ellas. Las secuencias de un mismo gen en un conjunto de especies serán más distintas cuando más alejadas filogenéticamente estén las especies comparadas [1]. El alineamiento de secuencias se

puede realizar mediante varios programas informáticos desarrollados entre los que se pueden mencionar Clustalw [2], Muscle [3] y T-Coffee [4].

1.1.2 Edición de secuencias

Dada la gran variedad de formatos existentes para representar secuencias de ADN, ARN o proteínas es necesario a veces convertir de un formato a otro, lo que se conoce como traductores de formatos, un ejemplo de ello es el programa transeq incluido en el paquete de programas EMBOSS disponible en <http://emboss.bioinformatics.nl/>. En este sitio podemos de igual modo encontrar herramientas muy útiles para la edición de secuencias tales como: pasteseq, que permite copiar una secuencia dentro de otras, extractseq o cutseq, que permiten extraer o eliminar regiones, dotmatcher, que muestra en gráficas el umbral de coincidencia de dos secuencias, makeseq o newseq, los cuales generan secuencias de proteínas de forma aleatoria, sizeseq, que ordena secuencias por tamaño entre otras muchas funcionalidades. También se utilizan varios programas para extraer información y características de las secuencias entre los que se pueden mencionar diffseq, infoseq y infoalign.

1.1.3 Representación gráfica de secuencias

Debido a la gran complejidad que presentan las estructuras macromoleculares la visualización es esencial en la comprensión de sus estructuras biológicas. La mayoría de los investigadores encuentran dificultad a la hora de visualizar moléculas en el espacio tridimensional, por un lado, la mayoría de los programas construidos para este fin son de origen propietario y es difícil su adquisición y por otro, muchos de estos programas necesitan ser instalados en el ordenador para lo que se necesita conocimientos básicos de informática. Entre los programas más usados para la visualización molecular, debido a que son de código abierto y de licencia libre, se encuentran el PyMOL, VMD y el Jmol. El PyMOL es una aplicación de escritorio desarrollada en el lenguaje de programación Python que puede ser utilizado en múltiples plataformas, por su parte el VMD se desarrolló utilizando la librería gráfica OpenGL y también está disponible para varios sistemas operativos. Ambos programas poseen la peculiaridad de que deben ser descargados e instalados manualmente por el especialista para su posterior uso, lo que constituye sin duda alguna una desventaja en cuanto a facilidad de uso y disponibilidad.

Jmol es un un visor Java de código abierto para estructuras químicas en tres dimensiones, implementado en Java por lo que puede ejecutarse en cualquier plataforma, posee una miniaplicación: JmolApplet, que permite integrarlo en páginas web y visualizarlo a través de un navegador web como Mozilla Firefox, Internet Explorer y Google Chrome.

1.2 Arquitectura Orientada a Servicios

Actualmente la gran diversidad de aplicaciones existentes y la gran cantidad de funcionalidades implementadas en ellas hacen necesario la integración de funciones de una aplicación en otra, sin embargo la arquitectura de software utilizada en la creación de estos sistemas atenta contra esta posibilidad. La industria del software ha buscado alternativas a este problema y una de ella es la Arquitectura Orientada a Servicios o SOA (siglas del inglés Service Oriented Architecture).

SOA ha demostrado ser eficiente y es un tópico bastante debatido en la actualidad, propone una serie de principios y conceptos que aumentan la reutilización de código y la interpolación de aplicaciones independientemente de las tecnologías usadas en su desarrollo. SOA puede establecer una abstracción de la lógica de negocio y la tecnología que puede introducir cambios en el modelo de proceso de negocio y la arquitectura, obteniendo como resultado un bajo acoplamiento entre estos modelos [5].

SOA utiliza los servicios web como base de la arquitectura orientada a servicios, una aplicación web que utilice servicios web para ejecutar la lógica de negocio basado en los principios de la orientación a servicios debe presentar un modelo similar al de la figura 1.

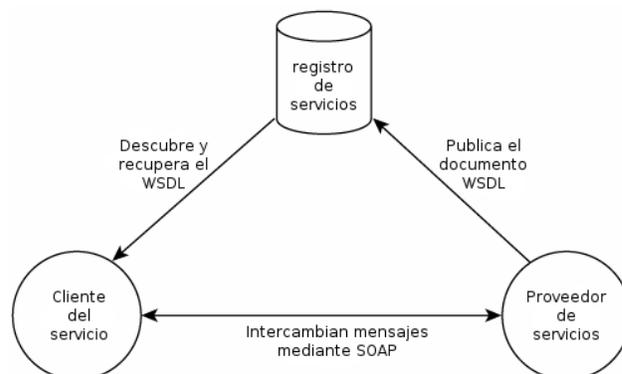


Figura. 1: Representación general de una arquitectura orientada a servicios

El cliente web (el que utilizará el servicio) consulta en un registro el documento WSDL (siglas del inglés Web Service Descriptor Language) que especifica los datos de conexión, así como las funciones y los parámetros necesarios para realizar las operaciones, luego el cliente se conecta al servicio web pasando los parámetros necesarios para que el servicio se ejecute correctamente, una vez terminado la ejecución el servicio devuelve una respuesta. El intercambio de información se realiza mediante el Protocolo Simple de Acceso a Objeto o SOAP (siglas del inglés Simple Object Access Protocol). El proveedor de servicios tiene la obligación de publicar el documento WSDL para que los clientes puedan consultarlo.

1.3 Características del sistema

El sistema está basado en la web, por lo que será accesible a través de un navegador web. El usuario solicita al servidor la página donde se encuentra el sistema, una vez allí el módulo presenta todas las funcionalidades que posee y el usuario decide la que utilizará. El procesamiento de la información se realiza en el servidor, que hace uso de los servicios web que implementan las funcionalidades para dar respuestas a las solicitudes del usuario, finalmente el servidor envía al usuario la respuesta. El intercambio de la información entre el usuario y el servidor se realiza mediante el protocolo de transferencia de hipertexto o HTTP (siglas del inglés HyperText Transfer Protocol) utilizando el lenguaje de marcado de hipertexto o HTML (siglas en inglés de HyperText Markup Language) que permite al navegador mostrar la página. La información que es necesaria preservar para futura utilización se almacena en una base de datos. El módulo implementa el patrón Modelo Vista Controlador o MVC por lo que separa la interfaz de usuario, la lógica de programación y la base de datos en capas diferentes, lo que posibilita una mayor organización en su desarrollo.

1.3.1 Estilos arquitectónicos y patrones

Las funcionalidades están representadas como servicios web mediante una Arquitectura Orientada a Servicios, estos servicios son llamados desde los controladores del sistema para realizar el procesamiento y recuperación de los datos biológicos. El sistema posee un estilo arquitectónico de llamada-retorno, ya que el cliente realiza una petición al servidor y este procesa la respuesta para después mostrársela, utiliza el patrón Modelo-Vista-Controlador o MVC (figura 2) para separar las capas de presentación(vista), de lógica de negocio (controlador)

y la capa de acceso a bases de datos (modelo), hace uso además del patrón DAO (del inglés Data Access Object) u objeto de acceso a datos que garantiza alta cohesión y bajo acoplamiento entre las clases que acceden a las bases de datos.

Aprovechando las funcionalidades brindadas por el marco de trabajo Spring Framework, el sistema trabaja con inyección de dependencia que centraliza la responsabilidad de asignar recursos y el patrón bean factory encargado de invocar a los constructores de las clases Javas correspondientes, además de un controlador frontal encargado de gestionar todas las peticiones procedentes desde el usuario y delegarlas a los controladores pertinentes [6].

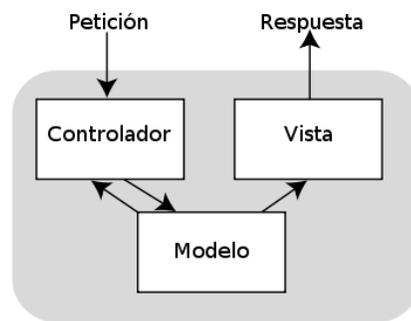


Figura. 2: Patrón Modelo Vista Controlador

1.3.2 Arquitectura general

La arquitectura de software del sistema queda representada en la figura 6. De esta forma queda planteada una plataforma de servicios que facilita la inclusión de nuevas funcionalidades para procesar los datos biológicos, los cuáles pueden ser accedidos mediante el Portal de Servicios Bioinformáticos de la Universidad de las Ciencias Informáticas o utilizados por otras aplicaciones.

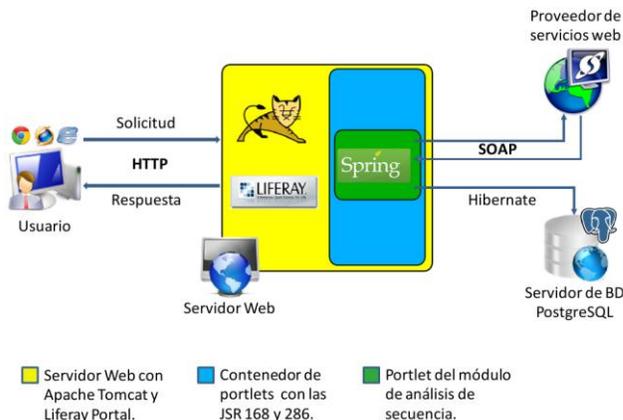


Figura. 3: Arquitectura de software del módulo de

análisis de secuencia

2. CONTENIDO

Para el desarrollo del sistema propuesto se tuvo en cuenta las políticas de migración del país hacia el software libre, descartando aquellas plataformas o herramientas de carácter lucrativos y abogando por el código abierto y las licencias de libre acceso.

2.1 Herramientas

Para el desarrollo del módulo se utilizó el lenguaje de programación Java, concretamente la plataforma destinada al desarrollo de aplicaciones web Java 2 Enterprise Edition. Todo el sistema se centra en el sistema gestor de contenidos o CMS (siglas del inglés Content Management System) Liferay Portal 6.0.5, que es un CMS escrito en Java de software libre y código abierto [7]. Liferay Portal puede usarse con varios servidores web, en este caso se utilizó el Apache Tomcat 6.0.26 encargado de procesar las páginas JSP (siglas del inglés Java Server Pages) que constituyen las vistas de la aplicación.

Liferay Portal implementa las especificaciones java JSR (siglas en inglés de Java Specifications Request) 168 y 286 por lo que es también un contenedor de portlets (componentes web reutilizables, que producen lenguaje de marcado), aprovechando esta capacidad el módulo está desarrollado bajo el paradigma de portlets, esto significa que puede ser transportado hacia otro contenedor compatible con las JSR mencionadas anteriormente y mantenerse completamente funcional.

Para el almacenamiento de la información se escogió como gestor de base de datos el PostgreSQL en su versión 8.4, un gestor de código abierto ampliamente diseminado en la comunidad internacional. Como marco de trabajo o framework se optó por utilizar Spring Framework ya que a través del módulo Spring Portlet MVC permite desarrollar portlets, separa el sistema en diferentes capas como son vista, controlador y modelo de base de datos aplicando el patrón Modelo-Vista-Controlador o MVC y provee una serie de recursos que facilitan el desarrollo de aplicaciones web, es además un marco de trabajo libre y de código abierto. Para lograr una mayor abstracción de la base de datos se determinó utilizar la herramienta de mapeo relacional o ORM (siglas de Object Relational Mapping) Hibernate [8], que viene además integrado con un lenguaje de consultas muy sencillo HQL (siglas de Hibernate Query Language).

En la realización de los servicios web se utilizó Apache Axis 2 tanto en la parte del cliente

(consumidor del servicio) como en el servidor (proveedor de servicios). En el desarrollo del lado del cliente se utilizó el marco de trabajo Javascript Jquery 1.6 y para utilizar la técnica Ajax (del inglés Asíncrono Javascript and XML) se usó Direct Web Remoting o (DWR por sus siglas en inglés). Como entorno de desarrollo integrado o IDE (siglas en inglés de Integrated Development Environment) se utilizó el Eclipse Helios 3.6, que permite la adición de complementos que facilitan el trabajo con Liferay Portal, Hibernate y Spring.

2.2 Resultados y discusión

El sistema está listo para ser utilizado por los especialistas del país a través de la Plataforma de Servicios Bioinformáticos de la Universidad de las Ciencias Informáticas y puesto en la red nacional reduce la dependencia de muchos centros investigativos con respecto a Internet, lo cual sin duda alguna es un paso de avance en la informatización de la sociedad.

2.2.1 Funcionalidades disponibles

El módulo actualmente cuenta con varias funcionalidades, organizadas en tres secciones: Alineamientos, Información gráfica e Información y edición. En la figura 4 se muestra la página inicial del sistema a través de un componente Portlet.



Figura. 4: Módulo de análisis de secuencia a través de un Portlet.

Como se puede apreciar, los alineamientos de secuencia se realizan con varios programas como clustalw, t-coffee o muscle. Tres herramientas muy utilizadas en la comunidad científica. También se pueden consultar los alineamientos realizados por el especialista en cuestión, ya que por el considerable tiempo que demoran estos procedimientos, son almacenados en la base de datos para su posterior

consulta.

La gestión de los alineamientos se realiza a través de la interfaz mostrada en la figura 5, en la que se muestran datos como fecha de inicio y fin del proceso, si ya ha sido consultado o no y además las opciones de descargar y eliminar el resultado.

El sistema permite la traducción de ácidos nucleicos a proteína mediante el programa transeq, la figura 6 muestra el formulario de entrada de secuencia y la figura 7 el resultado brindado por el sistema.

| No. | Tipo | Fecha de inicio | Fecha de fin | Estado | Opciones |
|-----|----------|---------------------------------|---------------------------------|---------------|----------|
| 3 | Muscle | Sun, 20 May 2012 15:57:03 +0000 | Sun, 20 May 2012 15:57:10 +0000 | Sin consultar | [Iconos] |
| 7 | Clustalw | Mon, 21 May 2012 01:30:05 | Mon, 21 May 2012 01:30:30 | Sin consultar | [Iconos] |
| 18 | Clustalw | Mon, 21 May 2012 01:45:14 | Mon, 21 May 2012 01:45:29 | Sin consultar | [Iconos] |
| 19 | Muscle | Tue, 22 May 2012 01:35:33 +0000 | Tue, 22 May 2012 01:35:33 +0000 | Sin consultar | [Iconos] |
| 20 | Clustalw | Tue, 22 May 2012 01:35:50 | Tue, 22 May 2012 01:35:51 | Sin consultar | [Iconos] |
| 22 | Muscle | Tue, 22 May 2012 01:47:51 +0000 | Tue, 22 May 2012 01:48:17 +0000 | Sin consultar | [Iconos] |
| 25 | Clustalw | Tue, 22 May 2012 17:02:18 | Tue, 22 May 2012 17:02:19 | Sin consultar | [Iconos] |
| 26 | Muscle | Tue, 22 May 2012 17:02:21 +0000 | Tue, 22 May 2012 17:02:22 +0000 | Sin consultar | [Iconos] |
| 27 | Muscle | Tue, 22 May 2012 17:02:37 +0000 | Tue, 22 May 2012 17:04:39 +0000 | Sin consultar | [Iconos] |
| 28 | Muscle | Tue, 22 May 2012 17:01:59 +0000 | Tue, 22 May 2012 17:11:05 +0000 | Sin consultar | [Iconos] |

Figura. 5: Gestión de los alineamientos de secuencias

Para esta ejecución se utilizó el genoma completo de la Neisseria Meningitidis, más conocida como la meningococo, enfermedad que continúa siendo un problema para la salud pública mundial y que afecta a todos los países del mundo ya sean desarrollados o subdesarrollados.

Traducir secuencia de ácidos nucleicos

Colapsar todo Desplegar todo Regresar a la página inicial

Opciones primarias

Pegue las secuencias en esta región

```
>gi|77358697|ref|NC_003112.2| Neisseria meningitidis MC58, complete genome
TTCCGGCTTAAACCTTATCCACATCCAAACCGGATAACCGTAACCCATTACCGTTATGGAAATGTCGCC
GACAAACCCAGCCGGAATGATCATAAAATATTTGACATCAGGCGTATAAAGATACAAGAACCTTA
TCCCCAGCGAAGCGCTGCGCTATGCAATGGGGCAGCAGCCCTCTGCCAATGCTTTCCGGCATATT
CAGGTAACAAAGACATCCCAACCAATATTATACCGTGGAAACTTTCCATCATCGCGCTTG
ACCGCAGCCGAAACCCAGGATTCGGGAATCCACAGCCGCAAAATGCCAGCGCCAGTTCGTCAT
CCTTCAACACCTCGCGTAAATAGGCATGAATCTTATCCACAGAAGACCGGTTCAAATCCGTCACCT
CCTCAAAACAGCCGTAACCAACCTGCGGATATGCCCGGCTTCCAGCCGTTAAATGAACAGATTG
TCCCAAAAGAGGGAATTCATCGTCAATTCGCCAGCCTTCCGTCGCGGCGTAAACCGCATTC
CAAGCATGGTCCAAACCGCCTCGATTTGCTCAAACTTCAGCCGCGGCGTTTTTGGCCATTGCTG
CAGGAATTTCCGCTTCCAAACCGGCGATGTCGCTGAGCCGCTGCAACCGCCGCGGCGCATCTCCA
AATCCGACTGCATCCGATGATTTTCCGTCAGATTTTTCGTTTTCGTAATAAGCCGGGATCCCGGA
TTGGATGCTGAGCAATTTGCTTACGATCCCGTCCCATACCGTTGTAGAAAAAACAAACATCAGAAA
ATAAAATATTTTTTCAITTTTAACTTCATTTAAATGCTGCTGCAAGCCGATTTCCGACATCAGACGGCA
TCGCCACCGCTGGGATAACCTTAAAGCGGGATGCGTTTCAACACTTCTTCTTTCGGGATTAAGTCCAAC
ACAGCATCGACGCTGGGGTTTTCCGCTACCCGAGAGCGGCAAGGCGCAGGGGCGATGCCGAGTTTCC
CCATTTTAAAGCTTCTGTCGCAAGAGGTTTGAAGAGGTCGAGATGCTTCCGCTTCCGCTTCCAGTTC
AGCCCTTCCAGGCGTCCGCAAGCGCAGCATACCGGCGCGGCTTCCATGCTCCGACTGTTCTGCAAG
TCTGCTTCCGCGAGCGTTTTGTTGAGCTAGAAGTAGAAGCACTCGTCGCAAGCGGTTTCAAGTCTTGG
GGCGGCTTTCAGCACTTCCCAACATCTTCCAAAGCAGGTTTTTCGTTTCAATGATATCGGCAAC
```

Figura. 6: Interfaz para la traducción de ácidos nucleicos a proteína

```
>NC_003112.2_1 Neisseria meningitidis MC58, complete genome
FGLNLIHIQTHNRNPFVME MSPDNHPAE*FIKYLHRRIRIKIQELYPQRTRCAYAVGDQP
PANAFSAIFR*NKDIPOQIFIPWKTFFHIMPLDRSRQQDSDGIHRSRCKRQFVILQTPAV
IGMNLIHRRPRFKSVPLLKQRLNQPADMPGFQPCNENSIVHKEGIIHRSIPRRLRSPCAVN
RIPSMVQTHSDLPQIFSLPGLRHCCRNFRFQGTGVDCLSRQLQTPARIFQIRLHPDDFSVQ
IVLLQ*GAVTGLDAEQLSSASPHTLVEKTTIRK*NIFFINFHNLAV*SRIPSTSDGIA
HACG*LRKGCVSTLLCRLMPTQHRRWGFSFYRRRQGAGACRVCPF*CLLVAEGFEVVD
GFGIPVFPQFEASAKRSIRAAAASSQCFCSTASAGVCLT*K*KHSSASVFKSWGRSLTSS
NTSSKAGFSVS*ISRNARRGLTSSASLPLGVILMCSPLIQ*SFFKSIIRLGDGETSFKSNH
SMNCSIVKNSSSPCAQPKRAR*LSIASGRMPIAPKSVMATVSPRLRLEIFLPCSLRIGRW
PYSGRFASMAFKMLICFGVLF TWSSPRITWVTPMS*STTTTQKL*VGVPSARAIIRSSSA
SLGMEISPLTKSVHLVTPSKGVLKRTTGCSDGISGRVLPSTGQRRS*VAEPPSFAFSR
MASSSSLLQ*Q*AWPFSKSSAMTSL*RSRVRV*TTLSALS*LRPHTFMPSRMLTDSA
VERAKSVSSIRNRSNPL*WRANAHEKAVRTPMCR*PVGLGAKRVLTVHMAPKSKAFI
LVLVPLCGIKNAV*TLADKVDGIFLTVFNASARGTVYHAPTEFLRDCVQKQV*NS*L
KKSRRPHFSKR*RDNAYPSCSIKIFLFPASNRRVGVAD I*TRTPKSSAISAARAKW*GSL
TTTTLAIPLARKTGRMLFRFS*VLRVFSNRMLRAGTCLLSAYRRPTSASVI*PFLVRPP
VNTILPTLRL*SAWAWLMRSRKTGEGRLSHAAAPNTSAAARTYGGKTCPPVR*TAQTD
EANTSKSGKTLKQKPPNR*RKPLRLQNSRLFTIPLRYFPAVCRQPPYRPLKQSDAPTGR
```

Figura. 7: Resultado de la traducción a proteína de la Neisseria Meningitidis

Además se pueden editar secuencias, insertar secuencias dentro de otras, eliminar o copiar regiones dentro de una secuencia, extraer secuencias más pequeñas desde otra, ordenar secuencias por tamaño y generar secuencias de forma aleatoria

Traducir secuencia de ácidos nucleicos

Regresar a la página inicial

>Generate_1
GATTCGGTACAACACCCGGAGATTGGTATGACTCCAGCCAGGGAGACCCGG
TAAAGTGGTGTAGCTCCCTAGAGCACTCAGCGGCAAGCCAGGATTCATCT
CGAATGATGATGAGCGCTGAGATCGGTAATGGTATAAACTAAGTGGTGGT
ACCTTCTAATA
>Generate_2
TGTACCGCGTGGATAGAGGATACACGACGCTGTTATCCGCATACGGACCGGT
TGGGCTATCCCTACGCGCTGAGAGAGGCAATCAAGTCCCTTCTGACCACT
CCTTGGCGTTGATGTTGCTCTATGGAATCAGATGAGAGGAGGCTCAGCATTTG
ATCATCGTGGCTTAAACAA

Aceptar cambio

>Generate_1
ReverseTATATTAGAAGTAAACCCACCCCTGATTATAACCATTACCGATCTCA
CGCTCCATATCAATTCAGAGATGAATCCGTTGGCTGCGGAGTGCCTAGG
GAGCTAGCAACCACTTACCGGCTCCCTGCGTGGAAAGTCAACCAAACTCCCG
GGTGTGACCGAATC
>Generate_2
ReverseTGTGTTAAGAGGCCAGATGATACAATATCGTAGGCTCCTTGCATCG
ATTCATAGAGCAACATCAACCCCAAGGAGTGTGAGAAAGGAGCACTGAT
GACCTTCTCAGCGGTAGAGGATAGCCCAACCGCGGTCCAGTATGGGATAAC
AGCTGCTGATCTCTATCCAGCCGTACA

Extrair subsecuencia
Unir secuencias
Ordenar secuencias
Invertir y complementar
Invertir
Complementar
OK
Crear Secuencia Nucleotica
Crear Secuencia Proteica

mediante el editor de secuencia (Figura 8).

Figura. 8: Editor de secuencias del módulo de análisis de secuencias

En la figura 8 se generaron primeramente dos secuencias nucleotídicas aleatoriamente las que fueron invertidas y complementadas, las secuencias de entrada están en el recuadro superior y el resultado de la operación se visualiza en el recuadro resaltado en color amarillo.

Otras de las funcionalidades del módulo es la visualización del umbral de coincidencia entre dos

secuencias, sacar la matriz de distancia de un alineamiento, mostrar la información de un alineamiento, concatenar varias secuencias en una sola, mostrar estadísticas de secuencias proteicas y mostrar en gráficas e imágenes estadísticas de secuencias proteicas y nucleotídicas.

Por otro lado el sistema puede visualizar en el espacio tridimensional estructuras cristalizadas, las cuáles están representadas en formato PDB (Figura 9). Para ello se utilizó el JmolApplet, garantizando de esta forma el acceso a través de un navegador web.

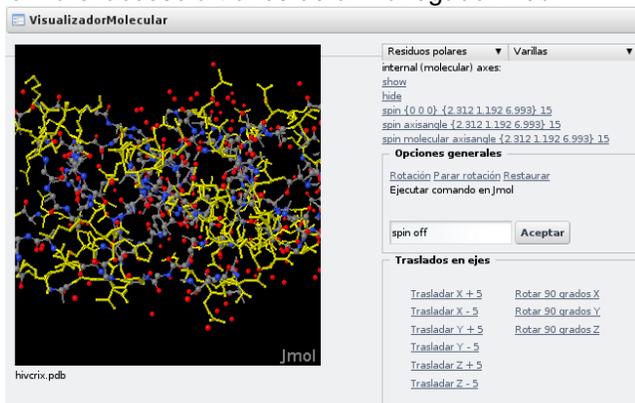


Figura. 9: Visualizador molecular del módulo de análisis de secuencia

En la figura nueve se visualizó a través del sistema el Virus de Inmunodeficiencia Humana (VIH) tipo 1, donde los residuos polares se mostraron como varillas y de color amarillo. La interfaz provee una gama de acciones que permiten configurar la visualización de la estructura a las necesidades del especialista, además el propio Jmol define una serie de comandos para este fin, que en caso de no estar conforme con las opciones básicas, pueden ejecutarse mediante la opción "Ejecutar comando en Jmol".

2.3 Perspectivas futuras

El módulo continúa en perfeccionamiento y de manera general se pretende adicionar nuevas funcionalidades para lograr una herramienta mucho más completa, también se estudia la posibilidad de realizar flujos de trabajos que involucren las funcionalidades básicas, por ejemplo entrar una secuencia al sistema, crear una secuencia aleatoria y realizar el alineamiento entre ellas, extraer de este alineamiento la matriz de distancia y mostrar finalmente los resultados en gráficas.

3. CONCLUSIONES

En la elaboración de este trabajo se analizaron las

herramientas existentes que permiten realizar análisis de secuencias biomoleculares y se utilizaron algunos programas para desarrollar un sistema a través de la web.

Se obtuvo un sistema acorde a las principales tendencias de las arquitecturas de software modernas que permite en un futuro utilizar las mismas funcionalidades en otras aplicaciones independientemente de la tecnologías que estas utilicen. Por último se obtuvo un módulo de análisis de secuencia listo para ser integrado a la Plataforma Web de Servicios Bioinformáticos que permite realizar algunas de las funcionalidades básicas en el campo de la Bioinformática, que constituye un paquete de herramienta indispensable para el desarrollo de los especialistas del país.

4. REFERENCIAS BIBLIOGRÁFICAS

1. Mount, D. W.. *Bioinformatics sequence and genome analysis*. Lugar de publicación desconocido. Cold Spring Harbor Laboratory Press. 2001. 565.
2. Thompson J. D.; Higgins D. G.; Gibson T. J.. 1994. *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice*. *Nucleic Acids Research*, 22, (22): 4673-4680
3. Edgar, R. C.. 2004. *MUSCLE: multiple sequence alignment with high accuracy and high throughput*. *Nucleic Acids Research*, 32, (5): 1792-1797
4. Notredame, C.; Higgins, D.; Heringa, J.. 2000. *T-Coffee: A novel method for fast and accurate multiple sequence alignment*. *J Mol Biol.*, 302, (1): 205-217
5. Erl, T.. *Service-Oriented Architecture: Concepts, Technology, and Design*. Lugar de publicación desconocido. Prentice Hall PTR. Agosto, 2005. 792.
6. Wall, C.; Ryan Breidenbach, S.. *Spring in Action 2*. E.E. U.U.. Manning Publications Co.. 2008. 765.
7. RICH SEZOV, JR.. *Liferay in action*. Shelter Island, New York, E.E. U.U.. Manning Publications Co.. 2010. 378.
8. Peak, P.; Heudecker, N.. *Hibernate Quickly*. Greenwich, Londres, U.K.. Manning Publications Co.. 2006. 453.

6. SÍNTESIS CURRICULAR DEL AUTOR

Andry Díaz: Graduado de la Universidad de las Ciencias Informáticas con el título de Ingeniero en Ciencias Informáticas. Se desempeña como Líder de Proyecto de la Plataforma de Servicios

Díaz, A.; Martínez, O.; Agramonte, A. | "MÓDULO PARA EL ANÁLISIS DE SECUENCIA DE LA PLATAFORMA DE SERVICIOS BIOINFORMÁTICOS"

Bioinformáticos de la UCI, que tiene como objetivo proveer a los especialistas bioinformáticos y centros del país en este campo, una serie de herramientas que actualmente solo están disponibles en Internet, además de varias herramientas desarrolladas en la propia universidad. Estuvo presente como coautor en el evento internacional de Informática 2011 en el trabajo "Portal Web de Servicios Bioinformáticos". Labora en el desarrollo de aplicaciones informáticas para el procesamiento de datos biológicos con inteligencia artificial.