

Propuesta de unificación de herramientas bioinformáticas para el análisis de secuencias proteómicas y genómicas de algas y hongos

Victor Terrón-Macias¹, Jezreel Mejía¹

{victor.terron, jmejia}@ciimat.mx

¹ Centro de Investigación en Matemáticas, Parque Quantum, Ciudad del Conocimiento, Av. Lassec, Andador Galileo Galilei, Manzana 3, Lote 7, CP 98160 Zacatecas, Zac., México.

DOI: 10.17013/risti.53.122-137

Resumen: El análisis de secuencias biológicas es esencial en bioinformática, pero la diversidad de herramientas disponibles presenta desafíos significativos. Este artículo propone una aplicación web que unifica diferentes funcionalidades útiles en el análisis de secuencias biológicas. Desarrollada en Python y Django, integra funcionalidades de BioPython, BioPandas, 3Dmol.js y BLAST, entre otras, que además posibilita su uso a usuarios sin conocimientos de programación. La herramienta abarca el análisis de secuencias de ADN y proteínas, transcripción, alineación, búsqueda de secuencias y más. Dentro de las características más destacables está la eficiencia, precisión al obtener datos cuantitativos del proceso que se está ejecutando y la facilidad de uso.

Palabras-clave: bio-informática, bio-tecnología, análisis de secuencias biológicas, sin código.

Proposal for the unification of bioinformatics tools for the analysis of proteomic and genomic sequences of algae and fungi

Abstract: Biological sequence analysis is essential in bioinformatics, but the diversity of available tools presents significant challenges. This article proposes a web application that unifies useful functionalities in biological sequence analysis. Developed in Python and Django, it integrates features from BioPython, BioPandas, 3Dmol.js, and BLAST, among others, making it usable for users without programming knowledge. The tool covers the analysis of DNA and protein sequences, transcription, alignment, sequence search, and more. Its most notable features are efficiency, precision in obtaining quantitative data from the running process, and ease of use.

Keywords: bioinformatics, biotechnology, biological sequence analysis, no code.

1. Introducción

El análisis de secuencias biológicas es una de las actividades fundamentales en bioinformática (Dimitrios Vasileiou, 2023) facilita la genómica estructural, comparativa y funcional, sirviendo como marco subyacente para predicciones sistemáticas de estructura/función, clasificación y anotación de proteínas. Este análisis ayuda a identificar funciones o características cruciales de los organismos (Katie Emelianova, 2023) (Azza E. Ahmed, 2021), como la complejidad de la regulación transcripcional y la información molecular, entre otras características o información valiosa.

Ante la gran cantidad de datos biológicos, es imperativo recurrir a software eficiente y probado que facilite estos análisis, más aún cuando se ejecutan a gran escala porque algunas de las herramientas actuales son complejas de utilizar (Azza E. Ahmed, 2021).

En este contexto, se ha identificado que existe una amplia variedad de software bioinformático como BioPython (Cock, 2009), cogent3 (Cogent3 Team), cactus (Joel Armstrong, 2020), PyCogent (Rob Knight, 2007), PyEvolve (Andrew Butterfield, 2004), entre otros; que realizan tareas específicas, y que a su vez promueven la dispersión de herramientas. Como resultado, profesionales relacionados al campo encuentran ineficiencias en sus tareas diarias al utilizar varios softwares en lugar de uno.

Jean-Baptiste Cazier en 2013 realizó una encuesta en donde obtuvo 47 respuestas de expertos en el campo (Cazier, 2013), en donde plantea la misma problemática sobre la existencia de muchas herramientas bioinformáticas, como resultado identificó los siguientes puntos:

1. 89% de las respuestas coinciden en que hay demasiadas herramientas bioinformáticas.
2. 34% de las opiniones coinciden en que es difícil encontrar software adecuado para tareas específicas.
3. 30% mencionan que es necesario compartir experiencias y conocimientos para hacer un mejor uso de las herramientas disponibles.
4. 13% de las opiniones indicaron que es necesario hacer un esfuerzo conjunto por unificar las herramientas en un solo entorno.

Otro aspecto relevante que se obtuvo de esta encuesta es que los expertos encuentran muchas dificultades en el uso de las herramientas bioinformáticas existentes, esto se presenta en los siguientes puntos:

1. 25% de las opiniones mencionan dificultad para encontrar software adecuado para tareas específicas.
2. 13% mencionan la falta de un esfuerzo conjunto para unificar las herramientas en un solo entorno.
3. 15% indicaron la complejidad y la curva de aprendizaje asociadas con algunas herramientas ante la necesidad de tener conocimientos de programación o de compilación de archivos fuente.
4. 11% resaltaron los problemas de documentación y falta de comprensión de los algoritmos subyacentes en algunas herramientas.
5. 15% mencionan la proliferación de herramientas que genera la confusión sobre cuál utilizar.

6. 17% indican la necesidad de actualización y mantenimiento continuo de las herramientas.
7. 9% mencionan la falta de recursos para evaluar y comparar las herramientas disponibles.
8. 15% indicaron las dificultades en la instalación y configuración de algunas herramientas.
9. 15% de los expertos coinciden en que existe variabilidad en los resultados obtenidos con diferentes herramientas para la misma tarea.

Ante este contexto este artículo presenta una propuesta de aplicación web con el objetivo de unificar diferentes herramientas de análisis en un mismo entorno enfocado en el análisis elemental y sin código que conjunte en un solo entorno diferentes herramientas atendiendo las problemáticas identificadas.

2. Plataformas relacionadas

Para identificar las funcionalidades que se pueden implementar dentro de la propuesta de unificación de herramientas se realizó una investigación explorando las principales plataformas científicas relacionadas al campo que incluyen diferentes herramientas, algunas de las principales plataformas identificadas son:

1. Expasy.
2. NCBI Tools.
3. Chimera.
4. BLAST.

En la siguiente subsección se presentan las funciones que cubre cada una de estas plataformas.

2.1. Expasy

Ofrece una amplia variedad de funciones relacionadas con el análisis y la manipulación de datos biológicos, dentro de las cuales están:

1. Análisis de Secuencias: Capacidad para analizar secuencias biológicas.
2. Predicción de Estructuras: Previsión de la estructura tridimensional de proteínas y ácidos nucleicos.
3. Búsqueda de Homologías: Identificación de secuencias similares en bases de datos.
4. Diseño de Primers: Herramientas para el diseño eficiente de cebadores para PCR.
5. Análisis de Expresión Genética: Evaluación de la expresión génica a nivel transcriptómico.
6. Anotación Funcional: Asignación de funciones biológicas a secuencias.
7. Bases de Datos: Acceso a una variedad de bases de datos biológicas.

2.2. NCBI Tools

La suite de herramientas de NCBI ofrece una amplia gama de recursos para la investigación biomédica:

1. Bases de Datos Biomédicas: Repositorios de datos biomédicos para investigación y análisis.
2. BLAST: Herramienta para comparar secuencias biológicas con bases de datos.
3. Entrez: Sistema de búsqueda y recuperación de información biomédica.
4. Información de Genes: Recursos para acceder y comprender información genética.
5. Exploración de Genomas y Especies: Herramientas para explorar información genómica.

2.3. Chimera

Chimera se centra en la visualización y modelado de estructuras moleculares:

1. Visualización de Estructuras Moleculares: Representación visual de estructuras tridimensionales.
2. Modelado de Estructuras: Creación y manipulación de modelos tridimensionales de macromoléculas.
3. Análisis de Interacciones: Evaluación de interacciones moleculares.
4. Renderización de Imágenes: Generación de imágenes de alta calidad para la presentación de datos.
5. Análisis de Archivos PDB: Exploración y análisis de datos estructurales en formato PDB.

2.4. BLAST

BLAST es una herramienta ampliamente utilizada para comparar secuencias biológicas:

1. Comparación de Secuencias: Identificación de similitudes entre secuencias biológicas.
2. Búsqueda Rápida en Bases de Datos: Acceso rápido a información relevante en bases de datos.
3. Alineación de Secuencias: Alineación de secuencias para identificar regiones conservadas.
4. Variantes y Mutaciones: Análisis de variaciones genéticas y mutacionales.
5. BLASTp, BLASTn, etc.: Variantes de BLAST para diferentes tipos de secuencias.

Considerando las funcionalidades identificadas en cada plataforma y con el objetivo de unificar la mayor cantidad de funcionalidades posibles, se realizó una comparativa entre ellas y la funcionalidad que cubren (ver Tabla 1).

Funcionalidad	Expasy	NCBI Tools	Chimera	BLAST
Análisis de secuencias	✓	✓		
Predicción de estructuras	✓		✓	
Búsqueda de homología	✓			
Diseño de primers	✓			
Análisis de expresión genética	✓			
Anotación funcional	✓			
Bases de datos	✓	✓		

Funcionalidad	Expasy	NCBI Tools	Chimera	BLAST
Cálculo y generación de árboles filogenéticos			✓	
Análisis de archivos PDB y visor de PDB			✓	
Traducción y transcripción de secuencias				
BLAST (BLAST-P, BLAST-N, BLAST-X, TBLASTN, TBLASTX)		✓		✓
Entrez				
Información de genes		✓		
Archivos de artículos científicos		✓		
Exploración de genomas y especies		✓		
Transcripción inversa		✓		
Análisis de Propiedades de Proteomas y ADN		✓		
Alineación y traducción de secuencias				✓
Análisis de interacciones			✓	
Renderización de imágenes			✓	
Variantes y mutaciones				✓

Tabla 1 – Comparativa respecto a la funcionalidad cubierta de cada plataforma.

Con base a la anterior comparativa, se identifica que del total de funcionalidades identificadas la plataforma que cubre mayor cantidad de esas funciones es NCBI Tools con un 43% de cobertura. De acuerdo al resultado de la Tabla 1, se establecieron las funcionalidades que debería tener la herramienta web que permitiría unificar diversas funcionalidades estableciendo que la plataforma propuesta cubrirá con 52% del total de las funcionalidades, tal como se ve en la Tabla 2, los signos de guión significan que la propuesta cubre parcialmente con esa funcionalidad:

Funcionalidad	Expasy	NCBI Tools	Chimera	BLAST	Propuesta
Análisis de secuencias	✓	✓			✓
Predicción de estructuras	✓		✓		
Búsqueda de homologías	✓				
Diseño de primers	✓				—
Análisis de expresión genética	✓				✓
Anotación funcional	✓				
Bases de datos	✓	✓			✓
Cálculo y generación de árboles filogenéticos			✓		✓
Análisis de archivos PDB y visor de PDB			✓		✓
Traducción y transcripción de secuencias					✓

Funcionalidad	Expasy	NCBI Tools	Chimera	BLAST	Propuesta
BLAST		✓		✓	✓
Entrez		✓			—
Información de genes		✓			✓
Archivos de artículos científicos		✓			
Exploración de genomas y especies		✓			
Transcripción inversa		✓			✓
Análisis de Propiedades de Proteomas y ADN		✓			✓
Alineación y traducción de secuencias				✓	✓
Análisis de interacciones			✓		
Renderización de imágenes			✓		
Variantes y mutaciones				✓	

Tabla 2 – Cumplimiento de funcionalidades de la propuesta de plataforma.

La siguiente sección aborda la propuesta de unificación, así como la selección de tecnologías que ayudan a lograr unificar las funcionalidades en un solo entorno sin necesidad de que el usuario tenga conocimientos de programación o de compilación de archivos.

3. Propuesta de unificación

Se consideró un conjunto completo de software y herramientas bioinformáticas ya existentes que cuentan con una amplia gama de herramientas que facilitan el análisis de secuencias biológicas, como ExpASy (Gasteiger, 2003) y NCBI Services (Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information, 2017), con el objetivo de identificar las funciones que se pudieran unificar en un mismo entorno.

Esta propuesta de herramienta fue desarrollada utilizando Python como lenguaje de programación, misma que integra librerías que han sido probadas como: BioPython (Cock, 2009), BioPandas (Raschka, 2017) y 3Dmol (Rego, 2014). Mismas que fueron seleccionadas por su estabilidad, confiabilidad y funcionamiento extensivo en el análisis de secuencias biológicas.

3.1. Arquitectura de la aplicación

La arquitectura de la aplicación está estructurada para optimizar la integración de varias funcionalidades de diferentes librerías de Python bajo el framework DJANGO, por ello esta propuesta reúne herramientas que se enfocan en los siguientes puntos:

1. Análisis de secuencias de ADN (ver Figura 1, con la secuencia de entrada “ACGTAAACGT”).
 - a. Análisis de propiedades: recibe como entrada una secuencia de nucleótidos, como resultado entrega: secuencia complementaria, secuencia

complementaria inversa, distribución de los aminoácidos presentes con su gráfico.

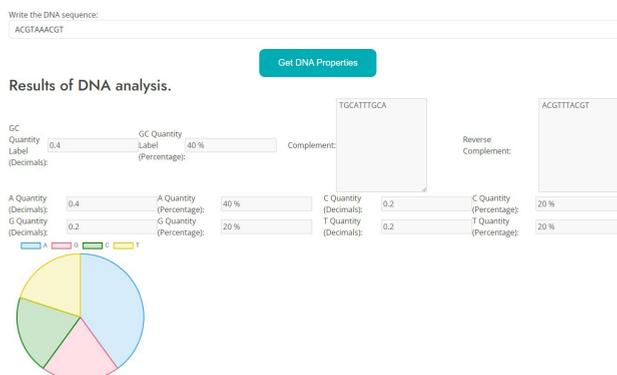


Figura 1 – Análisis de secuencias de ADN.

- i. Como resultado nos entrega una cantidad de GC de 0.4, equivalente al 40% porcentualmente.
 - ii. Adicionalmente nos entrega su secuencia complementaria “TGCATTTGCA” y su secuencia reversa complementaria “ACGTTTACGT”.
 - iii. Finalmente nos entrega la distribución de los elementos de la secuencia en decimales y porcentualmente A 40%,GCT con una distribución del 20% cada uno.
2. Análisis de secuencias proteómicas (ver Figura 2, con la secuencia de entrada *MEQPATKRRKQHNNENLETEPTGADYERMSAWLQREGADFANVTIHKSHESSEGYGIYAARA*).

Results of PROTEOME analysis.

#	Parameter	Value
1	Length of the Proteome:	60
2	Molecular weight:	6878.4238000000005
3	Aromaticity:	0.08333333333333333
4	Inestability index:	64.68499999999997
5	Iso-electric point:	5.849550819396971
6	Secondary Structure Fraction:	(0.16666666666666666, 0.2, 0.3333333333333333)
7	Molar Extinction Coefficient (E):	(9970, 9970)
8	E Reduced cysteines:	9970
9	Disulfide bridges:	9970
10	Gravy:	-1.2133333333333336
11	Charge at pH:	11.926312289610369

Figura 2 – Análisis de propiedades.

- a. **Análisis de propiedades:** recibe como entrada una secuencia de proteínas y un pH, como resultado entrega: tamaño de la secuencia, peso molecular, aromaticidad, índice de inestabilidad, punto isoeléctrico, estructura secundaria, coeficiente de extinción molar, numero de puentes de disulfuro, índice de hidropaticidad, carga eléctrica e hidrofobicidad.
 - i. Como resultados entrega la longitud de la secuencia: 60, peso molecular: 6878.4238 g/mol, aromaticidad: 0.083, índice de inestabilidad: 64.68, entre otros.
3. **Transcripción y transcripción inversa** (ver Figura 3, con la secuencia de entrada de GTGAAAAAGATGCAATCTATCGTACTCGCACTTCCCTGGTTC TGGTCGCTCCCATGGCAGCACAGGCTGCGGAAATTACGTTAGTCCCGTC AGTAAAATTACAGATAGGCGATCGTGATAATCGTGGCTATTACT GGGATGGAGGTCACTGGCGCGACCACGGCTGGTGGAAACAACAT GAGGCAATCGCTGGCACCTACACGGACCGCCGCCACCGCCGCGCCAC CATAAGAAAGCTCCTCATGATCATCACGGCGGTCATGGTCCAGGCAAACAT CACCGCTAA).
 - a. **Transcripción:** recibe como entrada una secuencia de nucleótidos, como salidas entrega la transcripción de la secuencia a mRNA y hebras plantilla.
 - b. **Transcripción inversa:** recibe como entrada una secuencia de mRNA y como salida entrega la secuencia de nucleótidos.

Important: Only Coding DNA.

Write the DNA sequence:

GTGAAAAAGATGCAATCTATCGTACTCGCACTTCCCTGGTTCGTTGCTCGCTCCCATGGCAGCACAGGCTGCGGAAATTACGTTAGTCCCGTCAGTAAAATTACAGATAGGCGATCGTATAATCG

Transcript DNA sequence

DNA Template Strand

5' to 3':

TTAGCGGTGATGTTTGCTGGACCATGCCCGGTGATGATCATGAGGAGCTTCTTATGGTGGCGCGCGGTGGCGCGGTCGGTGTAGGTGCCAGCGATTGCCTCGCCATTCATAATGTTGTTCCACCGCCG

3' to 5':

GTGAAAAAGATGCAATCTATCGTACTCGCACTTCCCTGGTTCGTTGCTCGCTCCCATGGCAGCACAGGCTGCGGAAATTACGTTAGTCCCGTCAGTAAAATTACAGATAGGCGATCGTATAATCGTGGCTATTACTGC

Figura 3 – Transcripción y transcripción inversa.

- i. En la figura se aprecia el resultado, nos entrega la cadena completa de 5' a 3' y de 3' a 5'.
4. **Traducción de secuencias** (ver Figura 4, con la secuencia de entrada AUUU UCUUUGCUCUUGAGCUCUGGCACUUCUCUGCUGCUGUCUG).
 - a. Como entradas recibe secuencias de mRNA, selecciona la tabla de codón a utilizar y como salida entrega el tipo de secuencia y la traducción de mRNA a ADN de la secuencia.

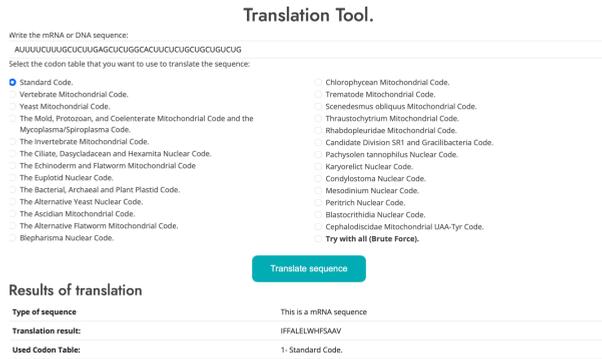


Figura 4 – Traducción de la secuencia.

- i. Como resultados nos entrega, el tipo de secuencia de entrada, el resultado de la traducción de la secuencia que es: “IFFALELWHFSAAV” y la tabla de codones utilizada para la traducción: tabla de codón estándar.
5. Alineación de secuencias por pares (ver Figura 5, con secuencia de entrada 1: “ATTTTCCTTGCTCTTGAGCTCTGGCACTTCTCTGCTGCTGTCTG” y secuencia de entrada 2: “ATTTTCCTTGCTCTTGAGCTCTGGCACTTCTCTG CTGCTGTC”).
- a. El usuario ingresa dos secuencias que quiera alinear, seleccionando también el modo de alineación, si es global o local y como salida la herramienta entrega estadísticas del proceso de alineación, así como las secuencias alineadas.

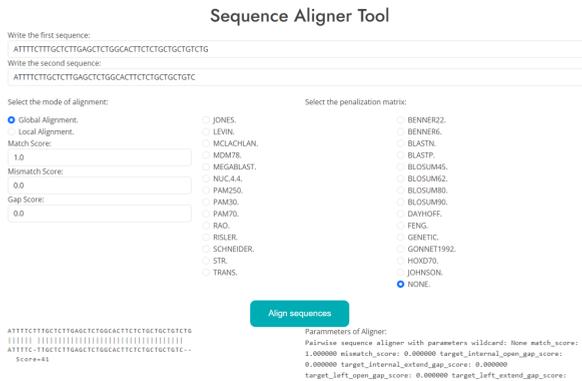


Figura 5 – Alineación de secuencias por pares.

- i. Como resultados nos entrega la secuencia alineada y las estadísticas del alineador (puntaje de coincidencia, puntaje de no coincidencia, tipo de alineamiento, entre otros).

6. Búsqueda de secuencias.

- a. El usuario proporciona una secuencia que puede ser proteína, nucleótido o gen, la aplicación hace una búsqueda utilizando BLAST y como salida entrega una lista de organismos que comparten similitudes con la secuencia. De esta funcionalidad no se coloca imagen por la extensión de esta.

7. Visualizador de archivos PDB (ver Figura 6, como valores de entrada recibe un archivo PDB que puede ser descargado en cualquier portal que ofrezca este tipo de visualizaciones de las proteínas de los organismos).

- a. La herramienta recibe como entrada un archivo PDB, y como salida entrega la estructura 3D de la proteína con sus diferentes modos de visualización.

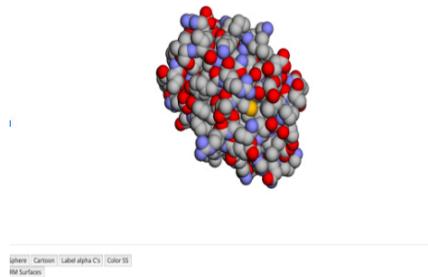


Figura 6 – Visualizador de archivos PDB.

- i. Como resultados entrega la visualización, así mismo cuenta con varios modos de visualización (palos, caricatura, esferas, vectores, etc.) dentro de los cuales el usuario puede interactuar para cambiar el modo de visualización.

8. Árboles filogenéticos (ver Figura 7).

- a. La herramienta recibe como entrada un archivo en formato CLUSTAL, como salidas entrega el cálculo de distancias entre las secuencias biológicas y genera un árbol filogenético.

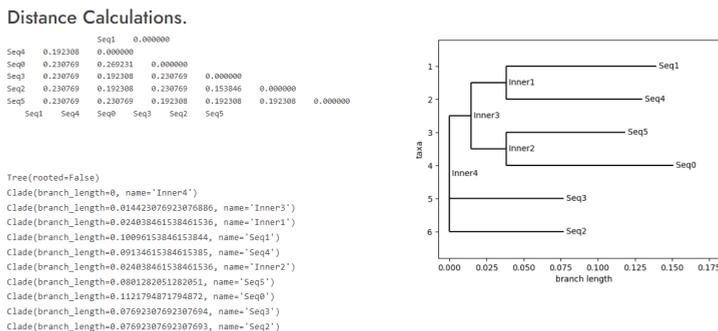


Figura 7 – Árboles filogenéticos.

- i. Como resultado, entrega el cálculo de distancias entre cada rama de las secuencias del archivo CLUSTAL, un diagrama que permite visualizar las distancias de las secuencias y el texto que forma cada rama del diagrama.
9. Análisis de MOTIFS y expresiones regulares (ver Figura 8, con las secuencias de entrada: TACAA, TACGC, TACAC, TACCC, AACCC, AATGC, AATGC).
- a. Análisis de MOTIFS: El usuario recibe una lista de MOTIFS, como salidas entrega una matriz de conteo, la secuencia de consenso, secuencia de consenso degenerado, reversa complementaria y la secuencia de consenso complementaria degenerada.
 - b. Expresiones regulares: la herramienta recibe una expresión regular de lo que va a buscar globalmente en la base de datos que se le cargue.

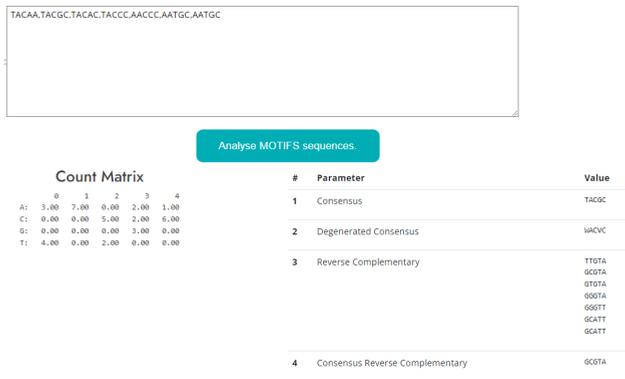


Figura 8 – Análisis de MOTIFS.

- i. Como resultados nos entrega la matriz de conteo, la secuencia consenso, la secuencia de consenso degenerado, la secuencia reversa complementaria y la secuencia consenso reversa complementaria.
10. Conexión con base de datos de hongos y algas con datos provenientes del JGI (JGI, 2022) o de cualquier otra fuente.
- a. El sistema se conecta a una base de datos MONGODB que debe ser instalada previamente, misma que contiene todas las secuencias que el usuario decida cargar ya sea de hongos, algas o del tipo de organismo que el usuario trabaje.

En las subsecciones siguientes se aborda la interfaz de usuario, el alcance y adaptabilidad, y otros aspectos relevantes para el desarrollo de esta propuesta de herramienta.

3.2. Interfaz de usuario

Como se mencionó inicialmente, la propuesta de herramienta posibilita a que usuarios sin conocimiento en programación puedan usar este tipo de herramientas, por ello el

desarrollo de una interfaz de usuario intuitiva fue necesaria para lograr este cometido (ver Figura 9).

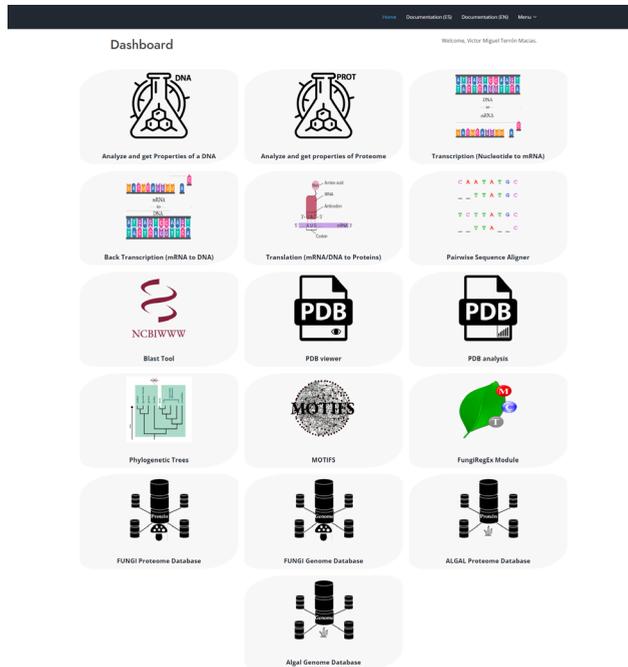


Figura 9 – Interfaz de usuario de la propuesta de herramienta.

La Figura 9 muestra la interfaz de usuario de la propuesta de herramienta con sus herramientas que están integradas dentro de la propuesta de aplicación.

3.3. Alcance y adaptabilidad

Esta propuesta de herramienta está diseñada para analizar secuencias biológicas de organismos de hongos y algas. Sin embargo, el software se adapta a una gama más amplia de secuencias biológicas sin importar si es o no de este tipo de organismos, lo que destaca su versatilidad. La adaptabilidad de la herramienta garantiza su utilidad en diversos contextos de investigación más allá del alcance inicial.

El tipo de secuencias que pueden analizar con la propuesta son secuencias genómicas y proteómicas de cualquier organismo, las secuencias genómicas son secuencias de nucleótidos o bases en una molécula de ADN (Brown, 2002). Por otra parte, las secuencias proteómicas son el total de moléculas proteicas presentes en una célula, tejido u órgano (Maino, 2000).

El alfabeto de las secuencias genómicas está conformado por 4 elementos: A, C, G y T; mientras que el alfabeto de las secuencias proteómicas se conforma por 26 elementos.

3.4. Optimización del flujo de trabajo

La integración de diferentes componentes de software en un mismo entorno es un punto clave del desarrollo de esta herramienta, ya que esta integración agiliza el flujo de trabajo eliminando la necesidad del usuario de navegar entre diferentes aplicaciones o software diferentes para el uso de sus herramientas o de tener conocimientos de programación.

3.5. Entrenamiento del usuario

Se ha preparado la documentación de esta propuesta de herramienta para facilitar su uso, con el objetivo de dotar al usuario del conocimiento necesario para el uso apropiado de la propuesta.

4. Resultados

La propuesta de herramienta presentó desafíos en la integración de tecnologías y componentes dentro de un mismo entorno principalmente por la diferencia entre los lenguajes de programación. Sin embargo, gracias al uso de diferentes tecnologías se logró una adecuada integración. A continuación, se presentan los resultados obtenidos, destacando cómo el uso de estas librerías ha contribuido a desarrollar la funcionalidad integral del software.

4.1. Integración de funcionalidades

La combinación de diferentes librerías en el desarrollo de la herramienta ha permitido la integración fluida de una amplia gama de funcionalidades relacionadas con el análisis de secuencias biológicas. Desde el análisis de propiedades de secuencias de ADN y proteínas hasta la generación de árboles filogenéticos, cada herramienta se ha implementado de manera coherente dentro del entorno unificado del software.

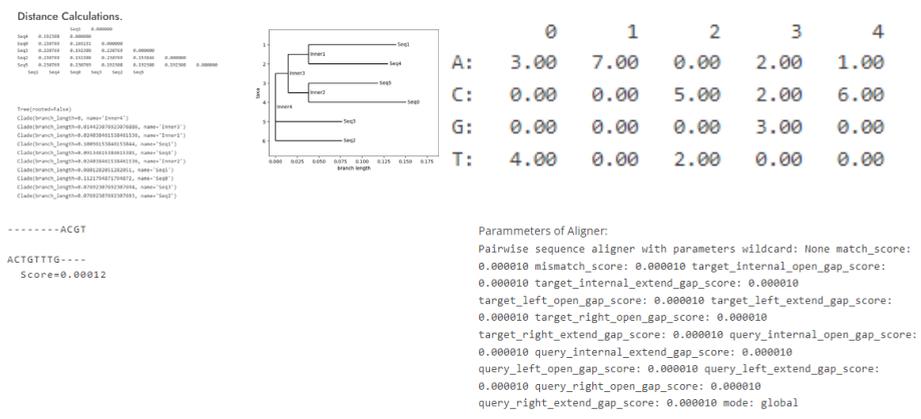


Figura 10 – Resultados cuantitativos.

La herramienta provee de resultados cuantitativos de los resultados que entrega, por ejemplo utilizando el generador de árboles filogenéticos muestra el cálculo de las

distancias entre las secuencias alineadas (ver Figura 10 esquina superior izquierda). Otro resultado cuantitativo que presenta la plataforma es al usar el alineador de secuencias por pares, ya que presenta el puntaje de alineamiento y una sección de estadísticas del alineador (ver Figura 10 sección inferior). Otro ejemplo de datos cuantitativos es al usar la herramienta de MOTIFS, ya que nos entrega una matriz de conteo de proteínas (ver Figura 10 esquina superior derecha).

4.2. Eficiencia y Precisión

El uso de librerías especializadas ha garantizado la eficiencia y precisión en el análisis de las secuencias biológicas. Por ejemplo, la implementación de BLAST ha permitido realizar búsquedas rápidas y precisas en la base de datos de NCBI, identificando secuencias similares con alta fiabilidad, funcionalidad que ha sido validada en el artículo publicado por P. Cock (Cock, 2009) que aborda la funcionalidad de BioPython. Asimismo, la integración de algoritmos de alineación de secuencias ha facilitado la comparación y el análisis de similitudes entre secuencias de ADN y proteínas que al igual esta funcionalidad ha sido validada por expertos en el campo que desarrollaron cada componente que se integró a esta herramienta.

4.3. Facilidad de uso

Gracias a la utilización de librerías bien establecidas y documentadas, unido al framework de DJANGO con sus integraciones la herramienta desarrollada ofrece una interfaz amigable y fácil de usar para los usuarios, destacando que se hizo más accesible el uso de herramientas bioinformáticas sin la necesidad de tener conocimientos en programación. Las funciones se presentan de manera clara, accesible y bien documentadas, lo que permite a los usuarios realizar análisis mediante el uso de los componentes de la herramienta de manera intuitiva y eficiente (Ver Figura 9).

4.4. Escalabilidad y adaptabilidad

La arquitectura del sistema y la forma en la que fue desarrollada la aplicación garantiza su escalabilidad y adaptabilidad a futuras actualizaciones y ampliaciones de funcionalidades ya que cada componente es independiente uno del otro, por tanto se pueden desarrollar más módulos mismos que se pueden integrar fácilmente a la herramienta sin comprometer otros componentes o funcionalidades ya desarrolladas, de esa forma se garantiza una constante actualización y adaptabilidad a futuras tecnologías, que es otra de las problemáticas identificadas por los expertos ante la disponibilidad de muchas herramientas que no reciben actualización.

5. Conclusiones

En esta artículo se destaca la diversidad de herramientas bioinformáticas disponibles que presentan tanto ventajas como desafíos significativos respaldados por la opinión de expertos en el campo. Si bien esta variedad ofrece una amplia gama de funcionalidades, también plantea problemas de eficiencia y efectividad para los profesionales, como la dificultad para encontrar software adecuado y la falta de unificación en las herramientas disponibles.

En respuesta a estas problemáticas identificadas, se propone una solución innovadora: el desarrollo de una aplicación web que unifica diversas funcionalidades para análisis de secuencias biológicas. Esta propuesta surge como una respuesta directa a la necesidad de simplificar y optimizar el proceso de análisis, ofreciendo a los usuarios un entorno integrado que abarca múltiples funcionalidades sin requerir conocimientos de programación.

La elección de tecnologías sólidas, como Python y el framework Django, junto con la integración de librerías especializadas como BioPython, BioPandas, 3Dmol.js, BLAST, entre otras, proporciona una base robusta para el desarrollo de la herramienta propuesta. Aunado a la elección de una arquitectura modular, misma que asegura la escalabilidad y adaptabilidad de la aplicación a futuras actualizaciones y ampliaciones de funcionalidades sin comprometer funcionalidades ya implementadas.

La interfaz de usuario intuitiva y la documentación exhaustiva son aspectos clave de la propuesta, ya que facilitan la adopción y el uso adecuado de la herramienta por parte de usuarios de diversos niveles de experiencia. La capacidad de proporcionar resultados cuantitativos, respaldada por la validación de la eficacia de la herramienta, fortalece su utilidad y confiabilidad en el ámbito científico.

Esta propuesta de herramienta representa un avance significativo en la simplificación y mejora de los procesos de análisis de secuencias biológicas en el campo de la bioinformática. Su implementación promete no solo mejorar la eficiencia y efectividad de los profesionales en este campo, sino también impulsar el progreso científico y tecnológico en áreas afines.

6. Trabajo Futuro

A pesar de los avances significativos logrados con nuestra propuesta de herramienta, existen áreas que requieren atención adicional y mejoras para maximizar su utilidad y efectividad en el campo de la bioinformática. En este sentido, se identifican varias líneas de trabajo futuro que podrían enriquecer y fortalecer aún más nuestra herramienta: a) incluir la integración de algoritmos avanzados de análisis de secuencias, herramientas de visualización de datos mejoradas y métodos de análisis predictivo-basados en aprendizaje automático; b) expandir la herramienta para abarcar una gama más amplia de organismos, desde la medicina hasta la agricultura y la conservación de la biodiversidad.

Referencias

Armstrong, J., Hickey, G., Diekhans, M., Fiddes, I. T., Novak, A. M., Deran, A., Fang, Q., Xie, D., Feng, S., Stiller, J., Genereux, D., Johnson, J., Marinescu, V. D., Alföldi, J., Harris, R. S., Lindblad-Toh, K., Haussler, D., Karlsson, E., Jarvis, E. D., Zhang, G. & Paten, B. (2020). Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature*, 246-251. <https://doi.org/10.1038/s41586-020-2871-y>

- Azza, E., Ahmed, J. M. (2021). Design considerations for workflow management systems use in production genomics research and the clinic. *Applications in Plant Sciences*, 16-27. <https://doi.org/10.1002/aps3.11538>
- Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. (2017). National Center for Biotechnology Information (NCBI).
- Brown, T. A. (2002). *Genomes 2*. Garland Science.
- Butterfield, A. (2004). PyEvolve: a toolkit for statistical modelling of molecular evolution. *BMC Bioinformatics*. <https://doi.org/10.1186/1471-2105-5-1>
- Cazier, J.-B. (2013). Is there too many bioinformatics tools? *Encuesta*. Universidad de Birmingham. Obtenido de https://www.researchgate.net/post/Is_there_too_many_bioinformatics_tools2/5
- Cock, P. J. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 1422-1423. <https://doi.org/10.1093/bioinformatics/btp163>
- Cogent3 Team. (s.f.). Cogent3. Obtenido de <https://cogent3.org/>
- Emelianova, K., Riaño-Pachón, D. M., Jimenez, M. F. T. (2023). Making sense of complexity: Advances in bioinformatics for plant biology. *Applications in Plant Sciences*. <https://doi.org/10.1002/aps3.11538>
- Gasteiger, E. (2003). ExpASY: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkg563>
- JGI. (2022). About Us. DOE Joint Genome Institute. Obtenido de <https://jgi.doe.gov/about-us/>
- Maino, R. (2000). Proteómica. El desafío de la post-genómica. *Naturee*, 403, 601.
- Raschka, S. (2017). BioPandas: Working with molecular structures in pandas DataFrames. *The Open Journal*. <https://doi.org/10.21105/joss.00279>
- Rego, N. A. (2014). 3Dmol.js: molecular visualization with WebGL. *Bioinformatics*, 1322-1324. <https://doi.org/10.1093/bioinformatics/btu829>
- Rob Knight, P. M. (2007). PyCogent: a toolkit for making sense from sequence. *genome Biology*. <https://doi.org/10.1186/gb-2007-8-8-r171>
- Vasileiou, D., Karapiperis, C., Baltasavia, I., Chasapi, A., Ahrén, D., Janssen, P. J., Iliopoulos, I., Promponas, V. J., Enright, A. J., Ouzounis, C. A. (2023). CGG toolkit: Software components for a computational genomics. *PLOS Computational Biology*, 1-10. <https://doi.org/10.1371/journal.pcbi.1011498>