08

Fecha de presentación: octubre, 2021
Fecha de aceptación: diciembre, 2021
Fecha de publicación: febrero, 2022

W-M2ALIGN: UNA HERRAMIENTA WEB

PARA EL ALINEAMIENTO MÚLTIPLE DE SECUENCIAS BASADA EN INFORMACIÓN ESTRUCTURAL DE LAS PROTEÍNAS

W-M2ALIGN: A WEB-SERVER TOOL FOR PROTEIN MULTIPLE SEQUENCE ALIGNMENT BASED ON STRUCTURAL INFORMATION

Cristian Zambrano Vega[1]
E-mail: czambrano@uteq.edu.ec
ORCID: https://orcid.org/0000-0001-8568-8024
Byron Oviedo[1]
E-mail: boviedo@uteq.edu.ec
ORCID: https://orcid.org/0000-0002-5366-5917
Emilio Zhuma[1]
E-mail: ezhuma@uteq.edu.ec
ORCID: https://orcid.org/0000-0002-3086-1413
[1] Universidad Técnica Estatal de Quevedo. Ecuador.

Cita sugerida (APA, séptima edición)

Zambrano Vega, C., Oviedo, B., & Zhuma, E. (2022). w-M2Align: una herramienta web para el alineamiento múltiple de secuencias basada en información estructural de las proteínas. *Revista Universidad y Sociedad*, 14(S1), 69-76.

RESUMEN

Este artículo presenta w-M2Align, una herramienta web fácil e intuitiva de usar para el alineamiento múltiple de secuencias de proteínas basada en la metaheurística multiobjetivo M2Align, cuyo criterio de calidad a optimizar son: STRIKE, basado en la información estructural de las proteínas, el Número de Columnas Totalmente Alineadas y el Porcentaje de Non-Gaps, los cuales están basados en la conservación de la estructura de las secuencias. El formato de las secuencias biológicas sin alinear es compatible con el formato FASTA y la información estructural de las proteínas puede ser obtenida automáticamente desde el sitio web del Protein Data Bank (PDB) o proporcionada por el usuario. Todos los cálculos se llevan a cabo en un clúster de computadora dedicado y los usuarios reciben los resultados a través de URL o correo electrónico. La plataforma permite descargar los ficheros con los resultados finales (Aproximaciones al Frente de Pareto y el conjunto de alineamientos generados en formato FASTA). El software W-M2Align está disponible en la dirección web http://revistas.uteq.edu.ec/m2align/, de forma gratuita y no requiere de inicio de sesión para su uso.

Palabras clave: Alineamiento Múltiple de Secuencias, Bioinformática, Optimización Multi-Objetivo, Herramienta Web BioInformática.

ABSTRACT

This article presents w-M2Align, an easy and intuitive web-server tool for Protein Multiple Sequence Alignment based on the multiobjective metaheuristic M2Align, which aims to jointly optimize three accuracy metrics: STRIKE score, based on the structural information of the proteins, the Number of Totally Aligned Columns and the Percentage of Non-Gaps, based on the conservation of the structure of the sequences. The format of the biological sequences is compatible with the FASTA format and the structural information of the proteins can be provided by user or obtained automatically from the website of the Protein Data Bank (PDB) or. All computations are carried out in a dedicated computer cluster and users receive the results via URL or email. The platform allows downloading the files with the final results (Approximations to the Pareto Front and the set of alignments generated in FASTA format). The w-M2Align software is available at the web address http://revistas.uteq.edu.ec/m2align/, free of charge and does not require login for its use.

Keywords: Multiple Sequence Alignment, Bioinformatics, Multi-Objective Optimization, BioInformatic Web-Tool.

INTRODUCCIÓN

El alineamiento múltiple de secuencias (MSA), es uno de los principales tópicos de interés dentro del campo de la Bioinformática. Su objetivo principal es la de encontrar un alineamiento óptimo para tres o más secuencias y resaltar la mayor cantidad de zonas conversadas y de similitud entre ellas. Para lograr esto, uno de los principales enfoques que se han empleado es justamente la Optimización Multiobjetivo, la cual ha demostrado brindar, incluso en algunos campos de la Bioinformática, beneficios y resultados significativos frente a los enfoques mono-objetivo (Zambrano-Vega, et al., 2016). Actualmente se han realizado análisis comparativos para conocer el rendimiento de las metaheurísticas multiobjetivo aplicadas al problema del Alineamiento Múltiple de Secuencia, el primero bajo un entorno bi-objetivo y el segundo bajo un enfoque tri-objetivo (Zambrano-Vega, et al., 2017c). En ambos trabajos se puede confirmar que el uso de la principal referencia multiobjetivo NSGAII, es la que mejor resultados brinda abordando el problema.

Una de las más recientes propuestas algorítmicas multiobjetivo publicadas para el Alineamiento Múltiple de Secuencias es el software M2Align (Zambrano-Vega, et al., 2017b), el cual está basado en la clásica referencia multiobjetivo NSGA-II, debido a buenos resultados expuestos previamente un optimizador multi-objetivo para alineamientos de secuencias, que utiliza tres objetivos para evaluar la precisión de MSA: puntuación Strike, columnas totalmente conservadas (CT) y el porcentaje de no-gaps. Este algoritmo está desarrollado en el framework multiobjetivo jMetalMSA (Zambrano-Vega, et al., 2017a) el cual es una extensión del framework jMetal (Nebro, et al., 2015) de donde se toman gran parte de las clases esenciales y básicas.

El software M2Align es un proyecto de código abierto que facilita a los usuarios interesados (biólogos, docentes, etc.) acceder desde el repositorio en GitHub (https://github.com/KhaosResearch/M2Align) al código fuente del mismo, para descargarlo, compilarlo y ejecutarlo, siguiendo las instrucciones que se encuentran publicadas en su sitio web. Implementa una representación más compacta de las secuencias alineadas basadas solo en la información de los gaps.

Los objetivos a optimizar de M2Align son tres, uno basado en la información estructural de las secuencias el cual nos garantiza la obtención de alineamientos más precisos en los casos de secuencias menos relacionadas evolutivamente y dos basados en la conservación de la estructura de las secuencias. Considerando el alto coste computacional requerido para el proceso de inferencia de un alineamiento óptimo para un grandes conjuntos de secuencias y de gran tamaño, se han incluido funcionalidades de paralelismo que permiten aprovechar las ventajas que brindan los sistemas de hardware multi-core, permitiendo un importante incremento de los speed-ups del algoritmo usando un conjunto de 4, 10, y 20 cores (Zambrano-Vega ***et al***., 2017b). M2Align está desarrollado usando las funcionalidades del framework multiobjetivo jMetalMSA (Zambrano-Vega, et al., 2017a).

Las herramientas software para el alineamiento múltiple de secuencias en gran parte se basan en líneas de comandos para ponerlos en funcionamiento, incluyendo M2Align, éste es un motivo que provoca que la tarea de llevar acabo su ejecución se convierta en una labor complicada para la comunidad científica (biólogos, biotecnólogos y afines) quienes comúnmente poseen pocos conocimientos técnicos y de configuración de aplicaciones, además las exigencias necesarias para el alineamiento de un gran número de secuencias y de un gran tamaño son computacionalmente costosas y requieren de mucho tiempo para su ejecución completa.

Por este motivo, el objetivo del presente trabajo de investigación es brindar una interfaz web al software multiobjetivo M2Align, llamada w-M2Align, que permitirá el rápido y fácil uso del algoritmo, además de poner a disposición de la comunidad de usuarios un clúster de computación dedicado para la ejecución de procesos subidos por los usuarios de la plataforma.

En las siguientes secciones se detallarán, primero en la sección 2 una descripción del problema, complejidad y las funciones objetivo. Luego en la sección 3 se detallarán los algoritmos multiobjetivo aplicados al MSA, en la sección 4 se incluyen las funcionalidades de la herramienta web w-M2Align, y finalmente en la sección 5 se mostrará las conclusiones del software

DESARROLLO

El Alineamiento Múltiple de Secuencias define el dominio del problema del Alineamiento Múltiple de Secuencia en términos formales. Sea **Σ** un alfabeto finito, por ejemplo un conjunto finito de caracteres, y $\Sigma \neq 0$, y un conjunto de k secuencias biológicas $S = (s_1; s_2, \ldots, s_k)$ de longitudes finitas y variables denotadas como l_1 a l_k y compuestas de caracteres $s_i = s_{i1}s_{i2},\ldots, s_{ili}$ $(1 \leq i \leq k)$, S' es una matriz que representa el alineamiento óptimo de S, la cual está definida formalmente por la ecuación 1:

$$S' = (s'_{ij}),\ con\ 1 \leq i \leq k, 1 \leq j \leq l, max(l_i) \leq l \leq \sum_{i=1}^{k} l_i \quad (1)$$

Y cumple con:

1. $S'_{ij} \in \Sigma \cup \{-\}$, donde "-" denota el carácter de espacios o "gaps";
2. Cada fila $s'_i = s'_{i1}, s'_{i2}, \ldots, s'_{il}$ $(1 \leq i \leq k)$ de S' es exactamente igual a la secuencia correspondiente s_i si eliminamos todos los gaps;
3. La longitud de todas las k secuencias es exactamente la misma;
4. S' no tiene columnas conformada solo por gaps.

En biología molecular, para las secuencias de ADN, el alfabeto Σ consiste de cuatro nucleótidos representados por los caracteres {A, T, G, C} y para las secuencias de proteínas, el alfabeto Σ que consiste de 20 amino-ácidos representados por los caracteres {A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y}.

El problema MSA es considerado como un problema de Complejidad NP-Completo (NP-Hard), ya que la exploración del espacio de búsqueda se incrementa exponencialmente; según el número de secuencias a alinear k y a su longitud máxima L, definida como $O(k2^kL^k)$ (Waterman, et al., 1976). Para tenerlo un poco más claro, en un grupo de solo 5 secuencias con un máximo de 10 residuos (amino-ácidos o nucleótidos) existen 1038 posibles combinaciones de alineamientos que se pueden generar.

Inicialmente el alineamiento de un par de secuencias se realizaba mediante el uso de técnicas de Programación Dinámica (Needleman & Wunsch, 1970). Aunque el uso de estas estrategias garantizan alineamientos matemáticamente óptimos, no pueden ser aplicadas cuando se consideran más de dos secuencias en el proceso, debido a la complejidad antes mencionada. Por estas razones, cada vez más, se considera importante y necesario el uso de metaheurísticas de optimización en la resolución del problema.

Una función objetivo mide la calidad del alineamiento y refleja cuan cerca está dicho del alineamiento óptimo biológico. En esta sección, definimos algunas de las funciones objetivo que fueron consideradas en el desarrollo del software M2Align, todas ellas están destinadas a ser maximizadas. Para la formulación de estas funciones, hemos considerado al alineamiento a evaluar como S, con un conjunto de k secuencias alineadas representadas como:

$S = s_1; s_2, \ldots, s_k$

todas ellas con la misma longitud L.

STRIKE (Kemena, et al., 2011) representa una nueva métrica para evaluar la calidad de los alineamientos basada en información estructural, de al menos, una de las secuencias del alineamiento. La información estructural de las secuencias de proteínas es comúnmente obtenida desde el sitio web del ***Protein Data Bank*** (PDB) (Berman, et al., 2000).

STRIKE calcula los contactos de la secuencia conteniendo información estructural mediante las distancias entre sus aminoácidos. Específicamente, se dice que dos átomos están en contacto intra-molecular cuando una solvente molécula no puede ser insertada entre sus superficies moleculares. La distancia entre dos aminoácidos para determinar si están en contacto se calcula a partir de la posición espacial de los átomos en los aminoácidos información que viene proporcionada en la estructura del archivo PDB. Con el fin de evitar contactos producidos por la estructura secundaria, STRIKE sólo considera aquellos contactos que involucran aminoácidos separados por al menos cinco aminoácidos en la secuencia.

Después de estimar los contactos de una secuencia, de acuerdo con su estructura terciaria, los pares de aminoácidos alineados en las mismas posiciones de las demás secuencias, son también retribuidos como contactos. Tales pares de aminoácidos son puntuados de acuerdo a una nueva matriz de puntuación denominada STRIKE Matrix, provista por los mismos autores

Esta matriz STRIKE está basada en información estructural según los contactos encontrados en un conjunto de alineamientos extraídos en varias base de datos. Específicamente se genera calculando el radio entre las frecuencias de cada posible contacto y de su expansión, dado la frecuencia de fondo de cada uno de los aminoácidos. Dado cualquier par de aminoácidos i y j, el puntaje de sus contactos es estimada según la Ecuación 2:

$$M_{ij} = 10 \times \ln\left(\frac{f_{ij}}{f_i f_j}\right) \quad (2)$$

Donde f_{ij} es la frecuencia de los contactos que implican los aminoácidos i y j a través de todos los contactos ***residue-residue*** observados, f_i y f_j son las frecuencias de aminoácidos únicas en el conjunto de datos considerado. Entonces, dado una secuencia s con su información estructural, la puntuación STRIKE del alineamiento calculada según la Ecuación 3:

$$\sum_{i=1;i\neq s}^{k} \sum_{j,c=1}^{L} M(x_{ij}, x_{ic}) \times EsContacto(x_{ij}, x_{ic}) \quad (3)$$

En la que x_{ij} y x_{ic} representan a cada par de aminoácidos en a i-ésima secuencia a excepción de la secuencia s. La función ***EsContacto*** es definida según la Ecuación 4

$$EsContacto(x_{ij}, x_{ic}) = \begin{cases} 1 & Si\ x_{ij}\ y\ x_{ic}\ son\ contactos \\ 0 & caso\ contrario \end{cases} \quad (4)$$

En el caso de existir varias estructuras proporcionadas, el puntaje STRIKE es calculado de forma separada para cada estructura y finalmente promediado. Esta métrica de evaluación permite identificar de mejor manera la exactitud en los alineamientos mejor que otras puntuaciones clásicas como BLOSUM62 (Henikoff & Henikoff, 1992) and PAM250. Además, supera claramente a las otras métricas clásicas cuando las secuencias son evolutivamente más distantes (Kemena, et al., 2011). STRIKE también muestra un fuerte efecto de correlación no-paramétrico con los valores BAliscore. Es decir, en una comparativa entre dos diferentes alineamientos, tanto BAliscore como STRIKE, generalmente identifican al mismo alineamiento como el mejor (alrededor del 79% de los casos) (Kemena, et al., 2011).

El porcentaje de columnas totalmente alineadas (TC) se refiere al número de columnas que están compuestas totalmente del mismo carácter en cada una de sus filas (amino-ácidos o nucleótidos). Esta función objetivo necesita ser maximizada para asegurar la mayor cantidad de regiones conservadas dentro del alineamiento. TC puede ser definida como (Ecuación 5):

$$TC(S) = 100 \sum_{l=1}^{L} \frac{ColumnaAlineada(S_l)}{L} \quad (5)$$

Donde S_l representa la l-ésima columna del alineamiento S, tal que $S_l = s_{il}\ \forall\ i = 1, \ldots, k$, y la función $ColumnaAlineada(S_l)$ está definida como (Ecuación 6):

$$ColumnaAlineada(S_l) = \begin{cases} 1 & Si\ s_{il} = s_{1l}\ \forall i = 2, ..., k \\ 0 & caso\ contrario \end{cases} \quad (6)$$

El porcentaje de no-gaps mide el número de residuos con respecto al número de gaps dentro del alineamiento, está definido en la Ecuación 7:

$$NonGaps(S) = 100 \sum_{i=1}^{k} \sum_{j=1}^{L} \frac{EsNonGap(s_{ij})}{k * L} \quad (7)$$

Donde s_{ij} representa el símbolo en la j-ésima posición de la i-ésima secuencia en el alineamiento S. La función *EsNonGap* para un determinado residuo del alineamiento está definido en la siguiente Ecuación 8:

$$EsNonGap(residuo) = \begin{cases} 1 & si\ residuo = \text{“}-\text{”}(gap) \\ 0 & caso\ contrario \end{cases} \quad (8)$$

Recientemente ha habido un creciente interés en la formulación multiobjetivo de los problemas de optimización que surgen en el campo de la Bioinformática. Handl, et al. (2007), muestran los beneficios de la Optimización Multiobjetivo aplicada específicamente en el campo de la Bioinformática en comparación con los enfoques monoobjetivo. A continuación se detallan algunos trabajos:

En los últimos tiempos, se han publicado varias propuestas multiobjetivo para resolver el problema del MSA usando técnicas metaheurísticas. La primera aproximación multiobjetivo fue presentada por Seeluangsawat & Chongstitvatana (2005), quienes publicaron MOMSA (Multiple Objective Multiple Sequence Alignment) un algoritmo evolutivo que optimiza dos objetivos de calidad implementados en una sola función, con el fin de mejorar las soluciones obtenidas desde el software Clustal X (Thompson, et al., 1994).

Considera dos objetivos a optimizar, la Suma de Pares y La Penalidad de Gaps, empleando como matriz de distancia Blosum45. Este algoritmo propuesto fue probado con nueve conjuntos de datos del benchmark BAliBASE 2.0 (Thompson, et al., 1999).

Zhu, et al. (2016), presentaron una propuesta basada en el algoritmo evolutivo multiobjetivo basado en la descomposición (MOEA/D) aplicado a resolver el problema del MSA, llamado MOMSA. También resaltan dos nuevas aportaciones dentro de su algoritmo: la generación de la población inicial y un nuevo operador de mutación. El rendimiento de esta técnica se comparó con varios métodos de alineamientos basados en algoritmos evolutivos, y también con técnicas basadas en métodos progresivos, usaron el conjunto de datos del BAliBASE 2.0 y BAliBASE 3.0 para evaluar su rendimiento.

Y por último, Ranjani & Ramyachitra (2016), propusieron dos algoritmos: Hybrid Genetic Algorithm with Articial Bee Colony Algorithm (GA-ABC) y Bacterial Foraging Optimization Algorithm (MO-BFO), pero su trabajo se centra principalmente en el rendimiento del algoritmo MO-BFO ya que obtiene un mejor rendimiento y porque identica mayormente bloques conservados dentro de los alineamientos Rani & Ramyachitra (2016), incorporaron en su trabajo cuatro objetivos a optimizar: la maximización de la Similitud, el porcentaje de no-gaps y Bloques Conservados, y la Minimización de la penalidad por gaps. Los algoritmos propuestos fueron evaluados resolviendo el benchmark BAliBASE v3.0 y comparados con

otros métodos MSA clásicos muy usados como: ClustalW y Clustal!, KAlign, MUSCLE, MAFFT y con varios algoritmos genéticos

En los últimos años se han desarrollado algunos métodos basados en información estructural para obtener alineamientos precisos. Básicamente, estos enfoques utilizan la información estructural del Protein Data Bank (PDB) (Berman, et al., 2000) como plantilla para guiar el alineamiento de un conjunto de secuencias no alineadas. Dos ejemplos de estas metodologías son: el software 3D-Coffee (Armougom, et al., 2006), una versión especial del método progresivo T-Coffee (Notredame, et al., 2000) que utiliza alineamientos estructural sobre las secuencias y el algoritmo evolutivo multiobjetivo MO-SAStrE (Ortuño, et al., 2013), que entre una de sus funciones objetivo a optimizar está el score basado en información estructural STRIKE (Kemena, et al., 2011), un índice para calcular las exactitudes de los alineamientos utilizando al menos la información estructural 3-D de una de las proteínas del conjunto de secuencias a alinear. Para más detalles del funcionamiento del score, ver la subsección 2.2.

El uso de software libre y open source es cada vez más grande, la plataforma con la que se contó para el desarrollo de la interfaz web contaba con el sistema operativo CentOS y con el servidor web Apache, estos catalogados como software libre. Basándonos en esto se procedió a determinar el lenguaje que proporcione todas las características necesarias y que sea compatible con la plataforma ya existente. Para el almacenamiento y manejo de los datos se utilizó el motor de base de datos MySql debido a que su licencia es software libre lo cual brinda gran compatibilidad con la plataforma ya instalada y cuenta además con una buena conectividad con el lenguaje de programación PHP.

En la sección de “Ejecución de Análisis” se presentan las opciones para el envío de un nuevo proceso de alineamiento de secuencias, en la Figura 1 se ilustra los campos requeridos, entre ellos:

- Archivo con las secuencias sin alinear en formato FASTA.
- Número de evaluaciones máximas del algoritmo.
- Tamaño de la población del algoritmo.
- Archivos PDB (Opciones: Generar automáticamente _ Subir manualmente).
- Archivos con los alineamientos iniciales.
- Email donde recibirá los resultados.

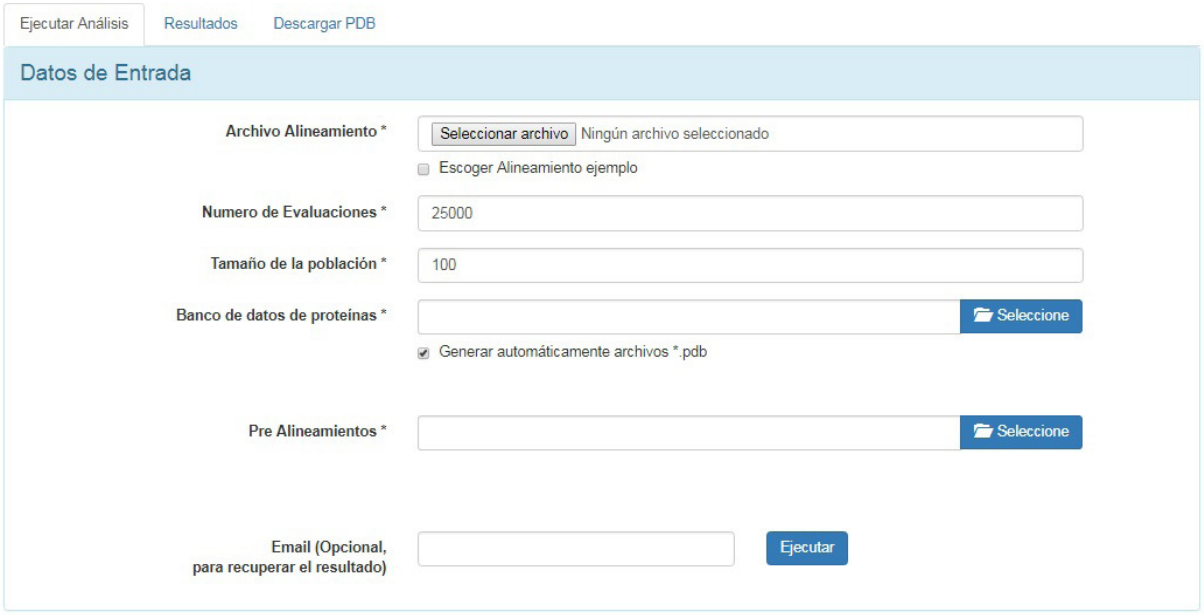

Figura 1. Interfaz para la creación de un nuevo proceso en w-M2Align.

Al usuario se le brinda tres diferentes tipos de ejecuciones, las que se detallan a continuación:

- Escoger un alineamiento de ejemplo y ejecutar.
- Cargar el alineamiento, los prealineamientos y ejecutar.
- Cargar el alineamiento, los pdb's, los prealineamientos y ejecutar.

Descargar información estructural (PDB)

Esta utilidad facilitará la descarga de los archivos *.pdb de manera conjunta y automática. Los archivos son descargados directamente del repositorio público Protein Data Bank (https://www.rcsb.org/) el cual brinda un servicio Https para la descarga de las estructuras de las proteínas de forma individual y en diferentes formatos (PDB, PDBx, /mmCIF, XML). El usuario debe indicar en forma de lista los "PDB ID" de las proteínas registradas en el repositorio y la plataforma W-M2Align procederá a la descarga de los ficheros PDB comprimidos en formato Zip.

RESULTADOS

En esta sección se encontrarán listados todos los trabajos (ejecuciones realizadas) del usuario con información correspondiente a su fecha y hora de ejecución (Figura 2), y su correspondiente estado (Ejecutándose, Finalizado, Error) según el avance de la ejecución. También se brinda la posibilidad de que los resultados producto de la ejecución puedan ser descargados en formato Zip mediante el ícono descarga de color rojo. Además se detallan los resultados generados por la herramienta web, en la Figura 3 se ilustra la manera en que se presentan, la aproximación del Frente de Pareto (FUN), los alineamientos en formatos FASTA (VAR), un log de resultados (Log) y la visualización de los alineamientos

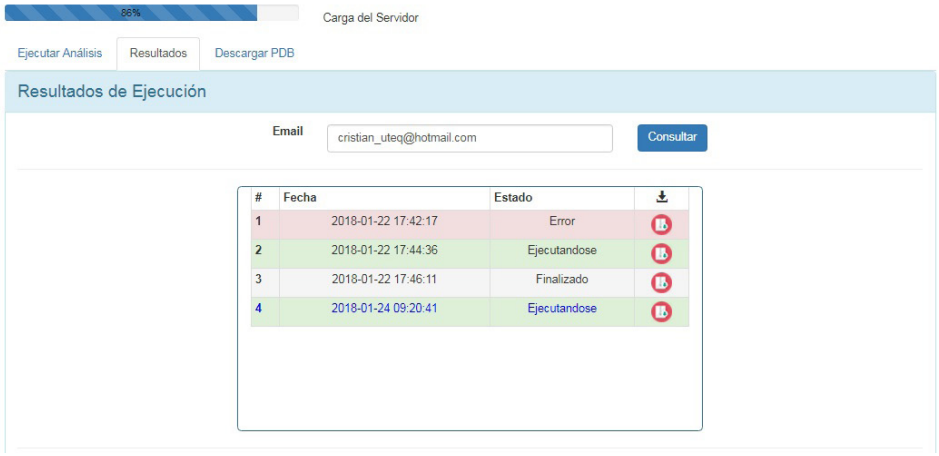

Figura 2. Lista de procesos en la sección de Resultados de la herramienta Web W-M2Align.

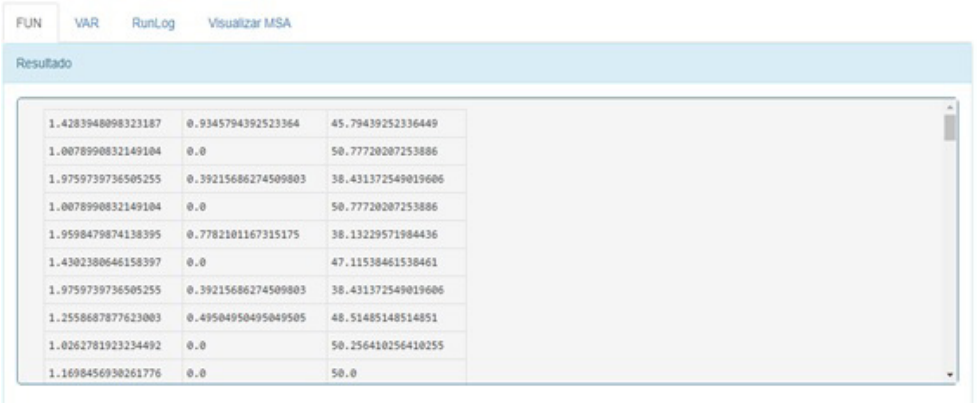

Figura 3. Detalle de resultados (Frente de Pareto, Alineamientos, Logs y Visualización de los MSA) generados por la herramienta Web W-M2Align.

Las ejecuciones que se ejecutan mediante la Interfaz Web W-M2Align corren procesos en el servidor por cada ejecución realizada. Estos procesos no dejan de correr mientras el software M2Align haya generado los resultados al finalizar su ejecución. Por esto en la sección de Administración, se puede Detener, Reanudar o Finalizar los procesos que se encuentren activos. Además se pueden conocer un histórico de todos los procesos ejecutados en la plataforma, mostrando la fecha, el usuario, el estado y el directorio donde se encuentran los resultados de la ejecución

Con el objetivo de demostrar la funcionalidad de la herramienta web, hemos realizado varios ejemplos sobre varios datasets del benchmark del BAliBASE 3.0.

En la Tabla 1 se muestran los detalles de los problemas y los tiempos de uso, así como también se ilustra en la Figura 4 un ejemplo de la visualización de los alineamientos que pueden ser obtenidos con la herramienta.

Tabla 1. Problemas del BAliBASE 3.0 resueltos con ayuda del W-M2Align.

Dataset	Num_Evals	Población	Email	Tiempo Ejecución (ms)
BB11001	25000	100	je_er_brito@hotmail.com.ar	14676
BB11002	25000	100	je_er_brito@hotmail.com.ar	48374
BB11003	25000	100	r.m.p.v94@gmail.com	482776
BB11004	25000	100	r.m.p.v94@gmail.com	431774.
BB11005	25000	100	czambrano@uteq.edu.ec	1330523
BB11006	25000	100	czambrano@uteq.edu.ec	156396.

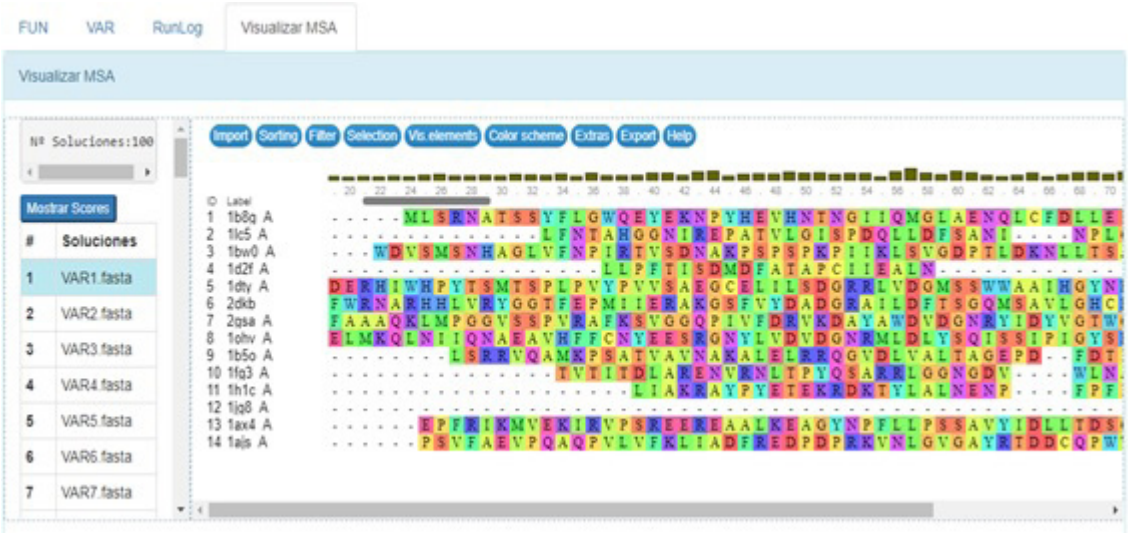

Figura 4. Visualización de los alineamientos generados por W-M2Align resolviendo el dataset BB11005 del BAliBASE 3.0.

CONCLUSIONES

En el presente trabajo de investigación se desarrolló una Interfaz Web fácil y rápida de usar al software M2Align, gracias a su correcta implementación y al óptimo rendimiento del clúster computacional donde se encuentra alojada la plataforma Los mecanismos implementados en la InterfazWeb para la carga y ajuste de los parámetros de entrada del software M2Align permiten ahorrar tiempo y evitan que existan errores en virtud a que la forma de ejecutar el software por parte del usuario ya no se la realiza ingresando manualmente comandos de consola. Se estableció tres estados que reflejan el progreso de la ejecución del software M2Align: “En ejecución”, “Finalizado” y “Error”, los cuales pueden ser visualizados en tiempo real en la sección de “Resultados” de la interfaz, a la que se puede acceder por medio del enlace enviado al correo del usuario.

La herramienta MSAViewer se utilizó para la presentación numérica, textual y visual de los alineamientos, y permitió la navegación interactiva a través de los resultados que genera el software M2Align. Es una herramienta compatible

con cualquier navegador y no requiere la instalación de complemento alguno

REFERENCIAS BIBLIOGRÁFICAS

Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., Keduas, V., Notredame, C., (2006). Expresso: automatic incorporation of structural information in multiple sequence alignments using 3d-coffee. Nucleic Acids Research, 34(suppl 2), 604-608-

Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I., & Bourne, P. (2000). The protein data bank. Nucleic Acids Research 28(1), 235-242.

Handl, J., Kell, D.B., & Knowles, J. (2007). Multiobjective optimization in bioinformatics and computational biology. IEEE/ACM transactions on computational biology and bioinformatics. IEEE, ACM 4(2), 279-92.

Henikoff, S., & Henikoff, J. (1992). Amino acid substitution matrices from protein blocks. Proceedings of the National Academy of Sciences, 89(22), 10915-10919.

Kemena, C., Taly, J., Kleinjung, J., Notredame, C. (2011). Strike: evaluation of protein msas using a single 3d structure. Bioinformatics 27(24), 3385-3391.

Nebro, A., Durillo, J.J., & Vergne, M. (2015). Redesigning the jmetal multi-objective optimization framework. (Ponencia). Annual Conference on Genetic and Evolutionary Computation. New York, USA.

Needleman, S., & Wunsch, C. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of Molecular Biology 48(3), 443-453.

Notredame, C., Higgins, D., & Heringa, J. (2000). T-coffee: a novel method for fast and accurate multiple sequence alignment. Journal of Molecular Biology, 302(1), 205-217.

Ortuño, F., Valenzuela, O., Rojas, F., Pomares, H., Florido, J., Urquiza, J., & Rojas, I. (2013). Optimizing multiple sequence alignments using a genetic algorithm based on three objectives: structural information, non-gaps percentage and totally conserved columns. Bioinformatics, 29(17).

Rani, R.R., & Ramyachitra, D. (2016). Multiple sequence alignment using multi-objective based bacterial foraging optimization algorithm. Biosystems 150, 177-189,

Seeluangsawat, P., & Chongstitvatana, P. (2005). A multiple objective evolutionary algorithm for multiple sequence alignment. (Ponencia). 7th Annual Conference on Genetic and Evolutionary Computation. New York, USA-

Thompson, J., Higgins, D.G., &n Gibson, T. (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.

Thompson, J.D., Plewniak, F., & Poch, O. (1999). A comprehensive comparison of multiple sequence alignment programs. Nucleic acids research, 27(13), 2682-2690.

Waterman, M., Smith, T., & Beyer, W. (1976). Some biological sequence metrics. Advances in Mathematics 20(3), 367- 387.

Zambrano-Vega, C., Cárdenas-Zea, M., & Aguirre-Pérez, R. (2016). Un enfoque multi-objetivo a la optimización del alineamiento múltiple de secuencias (msa). Latin American Journal of Computing, 3(1), 43-51.

Zambrano-Vega, C., Nebro, A.J., García-Nieto, J., & Aldana-Montes, J. (2017a). A Multi-objective Optimization Framework for Multiple Sequence Alignment with Metaheuristics, (pp. 245-256). Springer International Publishing.

Zambrano-Vega, C., Nebro, A.J., García-Nieto, J., & Aldana-Montes, J.F. (2017b). M2align: parallel multiple sequence alignment with a multi-objective metaheuristic. Bioinformatics, 33(19), 3011-3017.

Zambrano-Vega, C., Nebro, A.J., García-Nieto, J., & Aldana-Montes, J. (2017c). Comparing multi-objective metaheuristics for solving a three-objective formulation of multiple sequence alignment. Progress in Artificial Intelligence, 1-16.

Zhu, H., He, Z., & Jia, Y. (2016). A novel approach to multiple sequence alignment using multiobjective evolutionary algorithm based on decomposition. IEEE Journal of Biomedical and Health Informatics, 20(2), 717-727.