

Un enfoque Multi-Objetivo a la optimización del Alineamiento Múltiple de Secuencias (MSA)

Cristian Zambrano-Vega, Miriam Cárdenas-Zea y Ricardo Aguirre-Pérez

Resumen—El Alineamiento Múltiple de Secuencias (MSA por sus siglas en inglés) es uno de los principales tópicos de interés en el campo de la Bioinformática, consiste en encontrar un alineamiento óptimo para tres o más secuencias biológicas en el que exista la mayor cantidad de zonas conservadas o columnas de caracteres totalmente alineadas. Diferentes métricas para evaluar la calidad de los alineamientos han sido definidas en la literatura, lo que hace preciso que el problema MSA sea formulado y resuelto como un Problema de Optimización MultiObjetivo (MOP). Es por ello que hemos realizado un estudio de la resolución del problema bajo un enfoque MultiObjetivo considerando, como objetivos a optimizar simultáneamente, dos de los criterios de calidad más usados: la Suma Ponderada de Pares (the weighted Sum-Of-Pairs with affine gap penalties -wSOP) y el Porcentaje de Columnas Totalmente Alineadas (TC), realizando un análisis de rendimiento de tres algoritmos de mayor referencia en el área de la Optimización MultiObjetivo: NSGAIL, SPEA2 y MOCeL, resolviendo un conjunto de problemas del benchmark BALiBASE (V3.0). Los resultados revelan el alto grado de competitividad del algoritmo NSGAIL generando los mejores resultados, tanto desde una perspectiva multi-objetivo como bajo un aspecto biológico.

Index Terms—Alineamiento Múltiple de Secuencias, Metaheurísticas de Optimización MultiObjetivo, Optimización, Bioinformática.

Abstract—Multiple Sequence Alignment (MSA) is one of the main topics in the bioinformatics domain, consists in finding an optimal alignment for three or more biological sequences with the number maximum of conserved zones or totally aligned columns. Different scores to assess the quality of the alignments have been proposed, so the problem can be formulated and resolved as a Multi-Objective Optimization Problem (MOP). For this reason we have carried out a performed study resolving the MSA problem under a multi-objective approach, considering two popular metrics as objectives to be optimized: The weighted Sum-Of-Pairs with affine gap penalties (wSOP) and the Totally Aligned Columns (TC), with three algorithms from the state-of-the-art of Multi-Objective Optimization: NSGAIL, SPEA2 and MOCeL. Our experiments reveals that the classic metaheuristic NSGA-II provides the best overall performance resolving some problems provided by the benchmark BALiBASE (v3.0), under a multi-objective and biological approach.

Index Terms—Multiple Sequence Alignment, MultiObjective Optimization Metaheuristics, optimization, Bioinformatics.

I. INTRODUCCIÓN

Los autores Cristian Zambrano-Vega (czambrano@uteq.edu.ec), Miriam Cárdenas-Zea (mcardenas@uteq.edu.ec) y Ricardo Aguirre-Pérez (gaguirre@uteq.edu.ec), son docentes de la Carrera de Ingeniería en Sistemas de la Unidad de Estudios a Distancia de la Universidad Técnica Estatal de Quevedo. Quevedo - Los Ríos - Ecuador.

C. Zambrano-Vega se encuentra cursando sus estudios de Doctorado en Ingeniería del Software e Inteligencia Artificial de la Universidad de Málaga - España gracias al programa de Becas Doctorales de la Secretaría Nacional de Educación Superior, Ciencia y Tecnología - SENESCYT.

EL alineamiento múltiple de secuencias biológicas, sea ADN, ARN o estructuras primarias proteicas (proteínas), es uno de los principales tópicos de interés dentro del campo de la Bioinformática [1]. Su objetivo principal es la representar y comparar más de dos secuencias de aminoácidos o nucleótidos para resaltar la mayor cantidad de zonas de similitud entre ellas, las cuales podrían indicar relaciones funcionales o evolutivas entre los genes o proteínas consultadas. Su importancia radica en que de la calidad de los alineamientos depende la exactitud y precisión de otros procesos bioinformáticos que se realizan a partir de tales secuencias alineadas, como son la Inferencia Filogenética y la predicción estructural y funcional de proteínas.

El alineamiento de un par de secuencias se puede realizar mediante el uso de técnicas de Programación Dinámica [2]. Sin embargo, estas estrategias no pueden ser aplicadas cuando se consideran más de dos secuencias en el proceso, debido a que el espacio de búsqueda crece de forma exponencial según el número y el tamaño de las secuencias consideradas [3]. Por estas razones, cada vez más, se considera importante y necesario el uso de metaheurísticas de optimización en la resolución del problema.

El procedimiento de alineación básica se basa principalmente en la inserción de espacios o huecos (**gaps**) representados por el carácter “-” dentro del conjunto de caracteres de las secuencias, para hacer que todas ellas tengan la misma longitud y para lograr la alineación del mayor número de sus columnas. Es importante llevar a cabo la manipulación de las operaciones con los gaps (inserción, eliminación, desplazamiento, agrupamiento, etc.) con el fin de ir generando nuevas alternativas de alineaciones para mejorar la precisión y calidad del alineamiento final, ya que el número de gaps y sus ubicaciones determinan finalmente la calidad del mismo.

Se han propuesto una serie de métricas diferentes para medir la precisión y calidad de los alineamientos, tales como: el porcentaje de columnas totalmente alineadas (TC), el porcentaje de caracteres -No espacios- (NonGapsP), la Suma de pares (Sum-of-Pairs, SOP), la suma ponderada de pares con penalidad de gaps afines (weighted Sum-of-Pairs, wSOP), Strike [4], Entropy [5], BALiScore [5] o MetAl [6]. Sin embargo, todavía no existe un consenso acerca de qué métrica es la más apropiada o la más precisa para medir la calidad de los alineamientos. Por esta razón, es necesario considerar un enfoque MultiObjetivo para optimizar el problema, que permita obtener de forma simultánea alineamientos optimizados bajo dos o más criterios de evaluación, a fin de que los biólogos puedan disponer, no de una, sino de un conjunto de soluciones que les brinde la posibilidad de escoger una mejor

Tabla II: Mediana y Rangos Intercuartílicos de los valores del indicador I_{E+}.

Table with 4 columns: Algorithm, NSGAI, MOCe, SPEA2. Rows list various problem instances (BB11001 to BB11038) and their performance values for each algorithm.

se han aplicado los indicadores de calidad Épsilon y Spread. Según los resultados del indicador Épsilon (I_{E+}) en la tabla II, existe una alta relación de competitividad entre NSGA-II y SPEA2, ambos generan frentes de Pareto con un mejor nivel de convergencia para casi el mismo número de problemas (14 y 16 respectivamente), mas del doble de los problemas en los que MOCe logra superarlos (8 de 38). Finalmente los resultados del indicador Spread (I_Δ) en la tabla III indican que los frentes generados por el algoritmo MOCe tienen una mayor diversidad de soluciones para el 100 % de los problemas del grupo RV11 del benchmark.

V-B. Test estadístico de Wilcoxon Rank-Sum

Dado la alta competitividad que existe entre los tres algoritmos, para confirmar estos resultados, hemos realizado el test estadístico de Wilcoxon Rank-Sum para conocer si existen o no diferencias significativas entre ellos. Se consideran 7 problemas seleccionados del BAlIBASE (BB11007, BB11013, BB11019, BB11031, BB11032, BB11033 y BB11038) y los resultados se detallan en las tablas IV, V y VI para los indicadores Hypervolumen (I_{HV}), Épsilon (I_{E+}) y Spread (I_Δ) respectivamente.

Tabla IV: Resultado del Test de Wilcoxon rank-sum del indicador I_{HV}

Small table comparing MOCe and SPEA2 results for I_{HV} using symbols like ▲, ▼, -, +.

Tabla III: Mediana y Rangos Intercuartílicos de los valores del indicador I_Δ.

Table with 4 columns: Algorithm, NSGAI, MOCe, SPEA2. Rows list various problem instances and their performance values.

Tabla V: Resultado del Test de Wilcoxon rank-sum del indicador I_{E+}

Small table comparing MOCe and SPEA2 results for I_{E+} using symbols like ▲, ▼, -, +.

Tabla VI: Resultado del Test de Wilcoxon rank-sum del indicador I_Δ.

Small table comparing MOCe and SPEA2 results for I_Δ using symbols like ▲, ▼, -, +.

Los resultados de los indicadores I_{HV} y I_{E+}, revelan diferencias estadísticamente significativas a favor del algoritmo NSGAI frente al algoritmo MOCe, en la mayoría de los problemas seleccionados evaluando el Hipervolumen, y en todos los problemas seleccionados evaluando la convergencia de sus frentes. Además ambos análisis estadísticos indican que entre los resultados generados por los algoritmos NSGAI y SPEA2, no existe mayor número de diferencias significativas entre ellos; aunque para ciertos problemas, cada uno logra tener resultados significativos a su favor. Y finalmente, confirmando los resultados por el indicador I_Δ en la tabla III, se resaltan las diferencias significativas favorables de los resultados del algoritmo MOCe frente a NSGAI y SPEA2.

V-C. Frentes de Pareto aproximados

Los resultados de los indicadores de calidad multiobjetivo, son gráficamente confirmados por las figuras 1 y 2, en la que se ilustran los frentes de Pareto aproximados generados por los tres algoritmos junto a un frente de Pareto referencial generado en base a todas las ejecuciones independientes de los tres algoritmos, para los problemas BB11003, BB11010, BB11015, BB11021, BB11017, BB11019, BB11031 y BB11037 del BALiBASE.

Podemos observar que los cuatros frentes de Pareto de referencia de la figura 1 están en su mayoría conformados por las soluciones (alineamientos) de los frentes aproximados generados por el algoritmo NSGAI, y con pocas soluciones (alineamientos) de los algoritmos MOCe y SPEA2 en sus bordes. Para el caso de los problemas BB11031 y BB11037, cuyos resultados se ilustran en las figuras 1c y 1d, respectivamente, donde todo el frente de referencia está construido únicamente por el algoritmo NSGAI, en uno de los casos MOCe no brinda ninguna solución representativa en relación al frente de Pareto de referencia, y en el otro, sucede lo mismo con SPEA2.

Así mismo en la figura 2 se muestran los frentes de Pareto de referencia y los frentes de Pareto de los tres algoritmos sobre los problemas BB11003, BB11010, BB11015 y BB11021, ilustrados en las figuras 2a, 2b, 2c y 2d, respectivamente. En estos cuatros problemas seleccionados podemos darnos cuenta la alta competitividad que existe entre los tres algoritmos en estudio, ya que los tres generan soluciones no-dominadas de referencia, pero que al igual que en la figura 1, las soluciones de los frentes de Pareto aproximados generados por NSGAI representan la mayoría de las soluciones del frente de Referencia en cada problema. Finalmente podemos resaltar que de manera similar al grupo de frentes de Pareto de la figura 1, hay problemas en que los algoritmos MOCe o SPEA2 no generan ninguna solución no-dominada frente a las soluciones generadas por el algoritmo NSGAI.

V-D. Resultados Biológicos

En esta subsección de resultados biológicos hemos agregado la métrica BaliScore [22] (*Q*), la cual evalúa la precisión que existe entre un alineamiento candidato frente a otro alineamiento de referencia, para nuestro estudio hemos usado como referencia los alineamientos generados para tal efecto por el benchmark BALiBASE (v3.0) para cada uno de los 38 problemas del grupo RV11.

Las puntuaciones de las métricas *wSOP*, *TC* y *Q* de los alineamientos obtenidos del experimento se muestran en la tabla VII, se detallan los mejores scores (máximo valores en 10 ejecuciones independientes) de los tres algoritmos resolviendo cada uno de los 38 problemas del grupo RV11 del BALiBASE.

Según estos resultados el algoritmo NSGAI logra obtener los mejores resultados de la métrica *wSOP* en el 42 % de los problemas del RV11 (16 de los 38 problemas) frente al 24 % y 34 % de MOCe y SPEA2 respectivamente; lo que indica que NSGAI presenta ser ligeramente el mejor de los tres algoritmos en estudio con respecto a este indicador de calidad biológico.

Alineando totalmente el mayor número de columnas de las secuencias (métrica *TC*), NSGAI presenta ser el mejor algoritmo en el 50 % de los problemas del RV11, a diferencia del 24 % y 26 % de los problemas resueltos de mejor manera por MOCe y SPEA2 respectivamente;

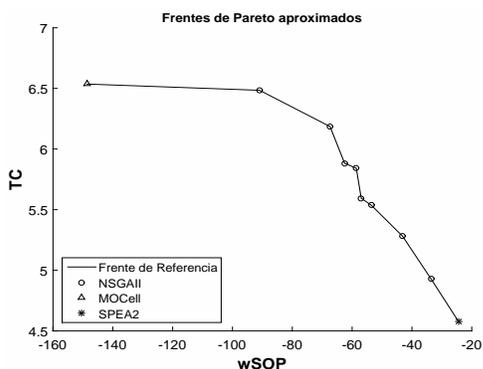
Y en el caso de la métrica *Q* existe un rendimiento muy competitivo y similar entre los tres algoritmos, NSGAI logra obtener los mejores puntajes en el 34 % de los problemas del RV11, MOCe en el 32 % y SPEA2 en el 34 %, lo que nos indica que los tres algoritmos generan alineamientos muy similares entre sí comparándolos con los alineamientos de referencia generados por BALiBASE.

VI. CONCLUSIONES Y TRABAJO FUTURO

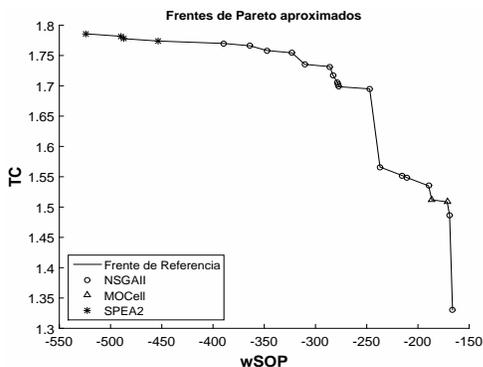
El objetivo del presente estudio fue optimizar el problema del alineamiento múltiple de secuencias (MSA) bajo un enfoque multiobjetivo empleando metaheurísticas clásicas como NSGAI, MOCe y SPEA2. Se realizaron dos tipos de análisis en la experimentación, el primero bajo una perspectiva de optimización multiobjetivo, en el que se emplearon tres indicadores de calidad multiobjetivo, y el otro bajo la perspectiva biológica, en el que se definieron tres métricas para evaluar la calidad de los alineamientos *wSOP*, *TC* y BaliScore; este último para medir la precisión de los resultados obtenidos frente a los alineamientos de referencia generados por el benchmark BALiBASE (v3.0). En base a los resultados obtenidos podemos concluir que, a pesar de la alta competitividad que existe entre los tres algoritmos en estudio, NSGAI y SPEA2 resultaron ser mejor que MOCe, pero que al competir entre ellos, NSGAI presentó generar un mejor rendimiento bajo ambas perspectivas (enfoque multi-objetivo y biológico). Sus frentes de Pareto aproximados presentan un mejor Hypervolumen (mayor convergencia y diversidad) que los generados por SPEA2, además sus puntajes de *wSOP* y la generación de zonas conservadas (*TC*) son mayores a los resultados de SPEA2. Esto nos permite considerar a NSGAI como un algoritmo base para implementar una futura propuesta mucho más competitiva e incluso definir otros objetivos a optimizar, como STRIKE que considera información estructural de las proteínas, para mejorar la calidad de los alineamientos.

AGRADECIMIENTO

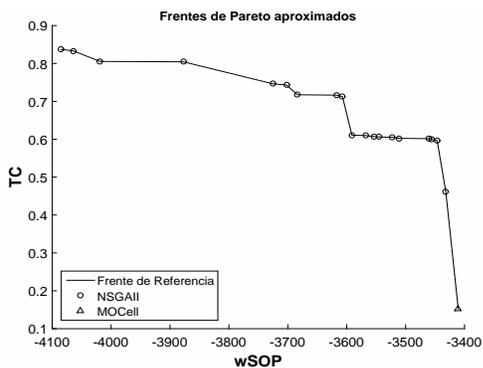
Este trabajo ha sido financiado por la Secretaría Nacional de Educación Superior Ciencia y Tecnología SENESCYT. Los autores agradecen el apoyo de las autoridades de la Universidad Técnica Estatal de Quevedo por incrementar la calidad de la producción Científica en la Unidad de Estudios a Distancia.



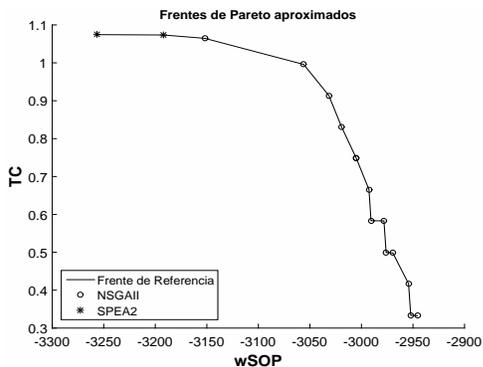
(a) Dataset BALiBASE BB11017



(b) Dataset BALiBASE BB11019

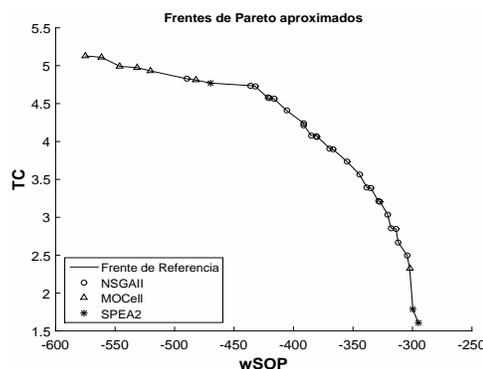


(c) Dataset BALiBASE BB11031

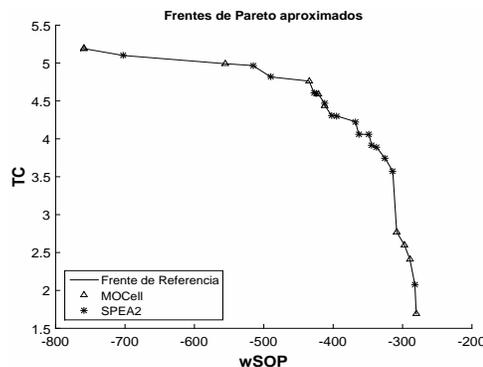


(d) Dataset BALiBASE BB11037

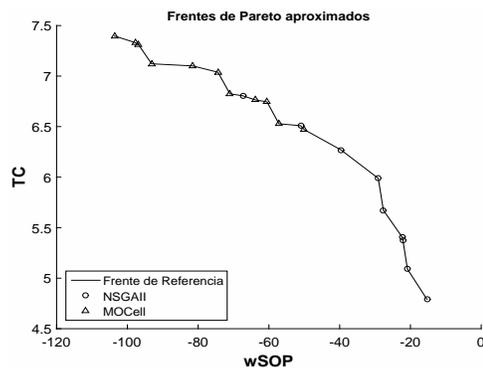
Figure 1: Frentes de Pareto aproximados de los algoritmos NSGAIi, MOCeII y SPEA2 a partir de todas sus 10 ejecuciones independientes y Frente de Pareto de Referencia generado en base a todas las ejecuciones independientes de los tres algoritmos, para los problemas del BALiBASE (v3.0): a) BB11017, b) BB11019, c) BB11031 y d) BB11037.



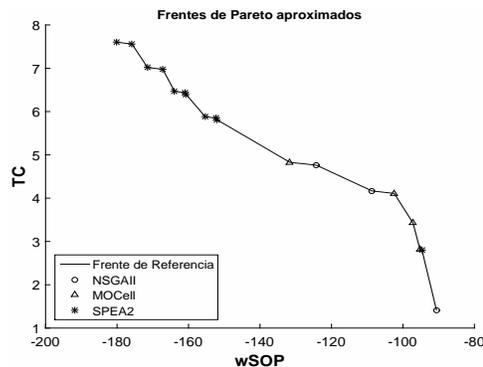
(a) Dataset BALiBASE BB11003



(b) Dataset BALiBASE BB11010



(c) Dataset BALiBASE BB11015



(d) Dataset BALiBASE BB11021

Figure 2: Frentes de Pareto aproximados de los algoritmos NSGAIi, MOCeII y SPEA2 a partir de todas sus 10 ejecuciones independientes y Frente de Pareto de Referencia generado en base a todas las ejecuciones independientes de los tres algoritmos, para los problemas del BALiBASE (v3.0): a) BB11003, b) BB11010, c) BB11015 y d) BB11021.

Tabla VII: Mejores Scores (suma ponderada de pares wSOP, columnas totalmente alineadas TC y BaliScore Q) generados por los algoritmos NSGAI, MOCell y SPEA2 en 10 ejecuciones independientes resolviendo todo el dataset de 38 problemas del grupo RV11 del BALiBASE (V3.0).

Instance	NSGAI			MOCell			SPEA2		
	wSOP	TC	Q	wSOP	TC	Q	wSOP	TC	Q
BB11001	29.88	5.88	0.92	28.66	6.48	0.92	29.88	6.67	0.92
BB11002	-803.23	1.48	0.49	-783.00	1.48	0.41	-756.28	1.48	0.49
BB11003	-304.43	4.83	0.67	-302.35	5.13	0.69	-295.09	4.77	0.68
BB11004	-399.73	3.49	0.60	-389.04	3.65	0.60	-410.03	3.59	0.59
BB11005	-2715.11	0.54	0.50	-2867.00	0.52	0.49	-2695.12	0.66	0.50
BB11006	-824.19	1.14	0.38	-866.19	1.11	0.39	-853.54	1.15	0.42
BB11007	-96.25	1.79	0.66	-107.69	1.72	0.68	-130.89	1.96	0.68
BB11008	-853.10	1.79	0.59	-850.69	1.82	0.61	-849.63	1.79	0.61
BB11009	-474.27	2.20	0.37	-475.31	2.41	0.35	-482.95	2.39	0.35
BB11010	-293.21	5.07	0.31	-280.25	5.20	0.30	-282.23	5.14	0.32
BB11011	-358.64	2.54	0.27	-368.39	2.28	0.28	-355.22	2.24	0.28
BB11012	165.03	6.25	0.88	162.79	6.10	0.89	166.59	6.25	0.88
BB11013	-179.86	1.96	0.11	-189.27	1.71	0.12	-179.83	1.85	0.15
BB11014	116.61	2.91	0.75	139.12	2.91	0.75	123.94	3.06	0.74
BB11015	-15.32	7.10	0.73	-24.22	7.40	0.74	-28.87	6.85	0.75
BB11016	-2321.15	0.97	0.45	-2287.56	1.10	0.43	-2305.29	1.01	0.43
BB11017	-32.79	6.48	0.73	-36.16	6.54	0.75	-24.42	6.48	0.73
BB11018	-4500.18	1.05	0.50	-4572.96	1.01	0.52	-4564.68	1.05	0.52
BB11019	-166.27	1.77	0.65	-171.50	1.72	0.63	-170.17	1.79	0.64
BB11020	-288.92	2.47	0.70	-310.87	2.46	0.69	-309.12	2.43	0.72
BB11021	-90.63	7.06	0.60	-91.67	7.56	0.62	-94.67	7.60	0.59
BB11022	-330.49	3.81	0.12	-329.05	3.81	0.12	-330.45	3.81	0.12
BB11023	-703.25	2.22	0.45	-681.06	2.10	0.40	-676.98	2.11	0.43
BB11024	-472.09	2.91	0.21	-455.26	2.72	0.19	-461.68	2.88	0.22
BB11025	-122.97	4.92	0.22	-126.78	4.92	0.18	-130.26	4.92	0.23
BB11026	-3992.68	0.11	0.20	-3989.91	0.11	0.19	-3989.97	0.11	0.17
BB11027	-1186.64	1.14	0.36	-1180.95	1.14	0.36	-1180.70	1.14	0.37
BB11028	-606.22	0.44	0.48	-613.32	0.44	0.49	-616.31	0.44	0.48
BB11029	-132.16	7.89	0.50	-134.09	7.33	0.50	-132.13	7.33	0.51
BB11030	-2111.00	0.42	0.58	-2043.73	0.42	0.57	-2092.90	0.42	0.58
BB11031	-3431.83	0.84	0.46	-3410.62	0.79	0.46	-3472.21	0.80	0.44
BB11032	-1339.50	1.18	0.64	-1359.47	1.31	0.63	-1338.28	1.15	0.61
BB11033	-1093.50	0.37	0.50	-1099.85	0.37	0.48	-1103.30	0.39	0.51
BB11034	-2044.50	1.54	0.44	-2081.76	1.50	0.42	-2010.19	1.49	0.44
BB11035	-143.86	5.48	0.54	-147.05	5.48	0.54	-147.64	5.44	0.54
BB11036	-772.76	1.68	0.60	-812.35	1.56	0.60	-783.83	1.71	0.59
BB11037	-2945.67	1.08	0.48	-2948.39	1.07	0.46	-2946.70	1.07	0.42
BB11038	-1533.82	1.93	0.71	-1607.61	1.92	0.70	-1558.03	1.91	0.70

REFERENCIAS

- [1] J. Pei, "Multiple protein sequence alignment," *Current Opinion in Structural Biology*, vol. 18, no. 3, pp. 382 – 386, 2008, nucleic acids / Sequences and topology. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0959440X08000407>
- [2] S. Needleman and C. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins," *Journal of Molecular Biology*, vol. 48, no. 3, pp. 443 – 453, 1970. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0022283670900574>
- [3] I. Elias, "Settling the intractability of multiple alignment," *Journal of Computational Biology*, vol. 13, no. 7, pp. 1323 – 1339, 2016.
- [4] C. Kemea, J. Taly, J. Kleijnung, and C. Notredame, "Strike: evaluation of protein msas using a single 3d structure," *Bioinformatics*, vol. 27, no. 24, pp. 3385–3391, 2011.
- [5] W. Soto and D. Becerra, "A multi-objective evolutionary algorithm for improving multiple sequence alignments," in *Advances in Bioinformatics and Computational Biology*, ser. Lecture Notes in Computer Science, S. Campos, Ed. Springer International Publishing, 2014, vol. 8826, pp. 73–82.
- [6] B. Blackburne and S. Whelan, "Measuring the distance between multiple sequence alignments," *Bioinformatics*, vol. 28, no. 4, pp. 495–502, 2012. [Online]. Available: <http://bioinformatics.oxfordjournals.org/content/28/4/495.abstract>
- [7] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.
- [8] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength pareto evolutionary algorithm," in *EUROGEN 2001. Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems*, K. Giannakoglou, D. Tsahalis, J. Periaux, P. Papailou, and T. Fogarty, Eds., Athens, Greece, 2002, pp. 95–100.
- [9] A. Nebro, J. Durillo, F. Luna, B. Dorronsoro, and E. Alba, "Design issues in a multiobjective cellular genetic algorithm," in *Evolutionary Multi-Criterion Optimization. 4th International Conference, EMO 2007*, ser. Lecture Notes in Computer Science, S. Obayashi, K. Deb, C. Poloni, T. Hiroyasu, and T. Murata, Eds., vol. 4403. Springer, 2007, pp. 126–140.
- [10] F. Ortuño, O. Valenzuela, F. Rojas, H. Pomares, J. Florido, J. Urquiza, and I. Rojas, "Optimizing multiple sequence alignments using a genetic algorithm based on three objectives: structural information, non-gaps percentage and totally conserved columns," *Bioinformatics (Oxford, England)*, vol. 29, no. 17, pp. 2112–21, Sep. 2013. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23793754>
- [11] F. J. M. da Silva, J. M. S. Pérez, J. A. G. Pulido, and M. a. V. Rodríguez, "Parallel Niche Pareto AlineaGA—an evolutionary multiobjective approach on multiple sequence alignment," *Journal of integrative bioinformatics*, vol. 8, no. 3, p. 174, 2011. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21926437>
- [12] M. Kaya, A. Sarhan, and R. Abdullah, "Multiple sequence alignment with affine gap by using multi-objective genetic algorithm," *Computer methods and programs in biomedicine*, vol. 114, no. 1, pp. 38–49, Apr. 2014. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/24534604>
- [13] F. P. M. Abbasi, L. Paquete, "Local search for multiobjective multiple sequence alignment," in *Bioinformatics and Biomedical Engineering*, ser. Lecture Notes in Computer Science, F. Ortuño and I. Rojas, Eds. Springer International Publishing, 2015, vol. 9044, pp. 175–182.
- [14] M. Dayhoff, R. Schwartz, and B. B.C. Orcutt, "A model of evolutionary change in proteins," in *Atlas of Protein Sequences and Structure*, vol. 5, pp. 345–352, 1978.
- [15] S. Henikoff and J. Henikoff, "Amino acid substitution matrices from protein blocks," *Proceedings of the National Academy of Sciences*, vol. 89, no. 22, pp. 10915–10919, 1992.
- [16] A. Nebro, J. Durillo, F. Luna, B. Dorronsoro, and E. Alba, "Mocell: A cellular genetic algorithm for multiobjective optimization," *International Journal of Intelligent Systems*, vol. 24, no. 7, pp. 723 – 725, 2009.
- [17] L. Bradstreet, *The hypervolume indicator for multi-objective optimisation: calculation and use*. University of Western Australia, 2011.
- [18] E. Zitzler, L. Thiele, M. L. an C.M. Fonseca, and V. da Fonseca, "Performance assessment of multiobjective optimizers: An analysis and review," *IEEE Trans. on Evolutionary Computation*, vol. 7, no. 2, pp. 117 – 132, 2003.
- [19] G. Raghava, S. M. Searle, P. C. Audley, J. D. Barber, and G. J. Barton, "Oxbench: A benchmark for evaluation of protein multiple sequence alignment accuracy," *BMC Bioinformatics*, vol. 4, no. 1, pp. 1–23, 2003. [Online]. Available: <http://dx.doi.org/10.1186/1471-2105-4-47>
- [20] P. I. W. de Bakker, A. Bateman, D. F. Burke, R. N. Miguel, K. Mizuguchi, J. Shi, H. Shirai, and T. L. Blundell, "Homstrad: adding sequence information to structure-based alignments of homologous protein families," *Bioinformatics*, vol. 17, no. 8, pp. 748–749, 2001. [Online]. Available: <http://bioinformatics.oxfordjournals.org/content/17/8/748.abstract>
- [21] R. Edgar, "Muscle: multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Research*, vol. 32, no. 5, pp. 1792–1797, 2004. [Online]. Available: <http://nar.oxfordjournals.org/content/32/5/1792.abstract>
- [22] J. Thompson, P. Koehl, and O. Poch, "Balibase 3.0: latest developments of the multiple sequence alignment benchmark," *Proteins*, vol. 61, pp. 127–136, 2005.
- [23] H. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. Shindyalov, and P. Bourne, "The protein data bank," *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, 2000. [Online]. Available: <http://nar.oxfordjournals.org/content/28/1/235>
- [24] A. Nebro, J. J. Durillo, and M. Vergne, "Redesigning the jmetal multi-objective optimization framework," in *Proceedings of the Companion Publication of the 2015 Annual Conference on Genetic and Evolutionary Computation*, ser. GECCO Companion '15. New York, NY, USA: ACM, 2015, pp. 1093–1100. [Online]. Available: <http://doi.acm.org/10.1145/2739482.2768462>



Cristian Zambrano-Vega Es estudiante de Doctorado en Ingeniería del Software e Inteligencia Artificial de la Universidad de Málaga. Su línea de investigación abarca las técnicas de optimización multiobjetivo aplicadas a la Inferencia Filogenética y al Alineamiento Múltiple de Secuencias. Docente en la carrera de Ingeniería en Sistemas de la Unidad de Estudios a Distancia de la Universidad Técnica Estatal de Quevedo - Ecuador. Email czambrano@uteq.edu.ec



Miriam Cárdenas-Zea Es Coordinadora de la Carrera de Ingeniería en Sistemas de la Unidad de Estudios a Distancia de la Universidad Técnica Estatal de Quevedo - Ecuador. Es estudiante de Doctorado en Ciencias Informáticas de la Universidad de Gramma - Cuba. Licenciada en Ciencias de la Educación en la Especialidad de Informática Educativa y Magister en Educación a Distancia y Abierta. Email mcardenas@uteq.edu.ec



Ricardo Aguirre-Pérez Es docente en la carrera de Ingeniería en Sistemas de la Unidad de Estudios a Distancia de la Universidad Técnica Estatal de Quevedo - Ecuador. Ingeniero en Sistemas e Informática, Magister en Educación a Distancia y Abierta y Especialista en Redes de Comunicación de Datos. Email gaguirre@uteq.edu.ec