**Optimización del Algoritmo de Forrajeo Bacteriano en Procesamiento de Secuencias de ADN**

**Optimization of the Bacterial Foraging Algorithm in DNA Sequence Processing**

Autor Cecyl Aglae García Orta 1[[1]](#footnote-1)

|  |  |
| --- | --- |
| **Resumen** | **Abstract** |
| Este estudio propone mejoras al Algoritmo de Forrajeo de Bacterias (BFOA) para optimizar el alineamiento múltiple de secuencias genéticas, un desafío computacional relevante en bioinformática. A través de ajustes en la interacción bacteriana y en parámetros como atracción, repulsión, y tamaño de población, el BFOA logra una optimización más precisa del fitness y una reducción en el número de evaluaciones de funciones necesarias.  Estas modificaciones permiten identificar y eliminar columnas con gaps de manera más eficiente, mejorando la calidad del alineamiento y reduciendo los tiempos de procesamiento. Este enfoque contribuye significativamente a la bioinformática al facilitar análisis rápidos y precisos de datos genéticos complejos, lo cual es esencial para el avance en investigaciones de biología molecular y genética.  ***Palabras Clave:*** Forrajeo Bacteriano, Secuencias de ADN, Optimización de Algoritmos | This study proposes enhancements to the Bacterial Foraging Optimization Algorithm (BFOA) to improve the precision of multiple sequence alignment, a significant computational challenge in bioinformatics. By adjusting bacterial interactions and parameters such as attraction, repulsion, and population size, the BFOA achieves more precise fitness optimization and reduces the number of function evaluations required.  These modifications enable the efficient identification and removal of columns with gaps, enhancing alignment quality and reducing processing times. This approach contributes significantly to bioinformatics by facilitating rapid and accurate analysis of complex genetic data, which is essential for advancing research in molecular biology and genetics.  ***Keywords:*** Bacterial Foraging, DNA Sequences, Algorithm Optimization |

# Introducción.

La bioinformática aplica herramientas computacionales para interpretar datos biológicos complejos, y uno de sus desafíos clave es el alineamiento múltiple de secuencias genéticas, esencial para comparar ADN, ARN o proteínas en investigaciones moleculares. Este problema es intensivo en recursos computacionales, especialmente con grandes volúmenes de datos. Algoritmos de optimización, como los algoritmos genéticos y el Algoritmo de Optimización Aritmética (AOA), han mostrado avances en la precisión del alineamiento. Entre estos, el Algoritmo de Forrajeo de Bacterias (BFOA) destaca por su enfoque en inteligencia de enjambre para resolver problemas de optimización de múltiples objetivos. Este estudio propone mejoras al BFOA ajustando parámetros clave para optimizar el fitness y reducir evaluaciones de funciones, mejorando así la precisión y eficiencia en el procesamiento de secuencias genéticas.

# Fundamentos.

El alineamiento múltiple de secuencias es un proceso esencial en bioinformática, empleado para identificar similitudes, variaciones y patrones en secuencias de ADN, ARN o proteínas. Esta tarea resulta fundamental para el análisis comparativo, la evolución molecular y la identificación de funciones genéticas. Sin embargo, su complejidad aumenta exponencialmente con el tamaño y la cantidad de secuencias, lo que plantea un desafío computacional considerable.

Para abordar este problema, se han desarrollado diversos algoritmos de optimización y heurísticos. Los métodos tradicionales, como el alineamiento progresivo y el de optimización global, suelen ser efectivos, pero demandan altos recursos computacionales, especialmente en estudios a gran escala. En respuesta, la bioinformática ha incorporado métodos basados en inteligencia artificial e inteligencia de enjambre, como los algoritmos genéticos y el Algoritmo de Optimización Aritmética (AOA), que buscan mejorar la precisión y reducir el tiempo de procesamiento en el alineamiento de secuencias.

El Algoritmo de Forrajeo de Bacterias (BFOA) es un enfoque inspirado en el comportamiento de bacterias que buscan nutrientes, aplicado aquí como un modelo de inteligencia de enjambre para resolver problemas de optimización de múltiples objetivos. El BFOA ha mostrado ser efectivo en diversas aplicaciones bioinformáticas, ya que ajusta las interacciones entre las bacterias para alcanzar soluciones óptimas en problemas de alineamiento, especialmente cuando se presentan gaps o espacios en blanco en las secuencias.

Este estudio fundamenta el uso del BFOA en el alineamiento múltiple de secuencias genéticas, proponiendo ajustes en parámetros como atracción, repulsión y tamaño de la población de bacterias para mejorar el fitness de las soluciones y reducir las evaluaciones de funciones. Estas mejoras buscan optimizar la precisión en la identificación de patrones y minimizar los gaps en el alineamiento, contribuyendo a análisis más eficientes en estudios genéticos y moleculares.

# Definiciones.

Estas definiciones brindan una base teórica sólida y facilitan la comprensión de los conceptos empleados en el estudio y análisis de algoritmos aplicados al alineamiento de secuencias.

**Bioinformática**: Disciplina que integra herramientas computacionales y métodos estadísticos para analizar datos biológicos, como secuencias de ADN, ARN y proteínas. Su objetivo es interpretar y organizar grandes volúmenes de datos biológicos para avanzar en la investigación en biología molecular, genética y medicina.

**Algoritmo de Forrajeo de Bacterias (BFOA)**: Algoritmo de optimización inspirado en el comportamiento natural de las bacterias al buscar nutrientes. Utilizado en problemas de optimización, el BFOA simula cómo una colonia de bacterias interactúa y busca "mejores soluciones" mediante procesos de quimiotaxis, reproducción y eliminación.

**Quimiotaxis**: Comportamiento de las bacterias de moverse en respuesta a estímulos químicos, desplazándose hacia concentraciones de nutrientes. En el contexto del BFOA, la quimiotaxis representa el proceso de exploración y explotación de soluciones en el espacio de búsqueda para optimización.

**Fitness**: Medida de la calidad de una solución en un algoritmo de optimización. En el BFOA, el fitness determina qué tan bien se alinea una secuencia genética respecto a las demás, considerando la penalización por gaps y la similitud en las posiciones alineadas.

**Gaps**: Espacios en blanco o huecos introducidos en el alineamiento de secuencias para maximizar las coincidencias entre las secuencias. Aunque necesarios para algunos alineamientos, los gaps reducen la precisión de la comparación y suelen penalizarse en el cálculo de fitness.

# Formulación de objetivos y establecimiento de hipótesis.

Definimos los objetivos que el artículo busca alcanzar, así como las hipótesis que guiarán el análisis y evaluación del estudio.

**Objetivo General:**

Optimizar el Algoritmo de Forrajeo de Bacterias (BFOA) para mejorar la precisión y eficiencia en el alineamiento múltiple de secuencias genéticas, reduciendo el número de evaluaciones de funciones y minimizando la presencia de gaps en el alineamiento.

**Objetivos Específicos:**

1. Implementar ajustes en parámetros clave del BFOA, como atracción, repulsión y tamaño de la población bacteriana, para mejorar el fitness y reducir el tiempo de procesamiento.
2. Evaluar el impacto de estos ajustes en la precisión del alineamiento múltiple de secuencias genéticas.
3. Validar el modelo mediante pruebas en secuencias genéticas reales para medir la reducción de gaps y el aumento de coincidencias en el alineamiento.

**Hipótesis**

El ajuste de parámetros de interacción bacteriana en el BFOA mejorará significativamente el fitness y la precisión del alineamiento múltiple de secuencias genéticas, reduciendo la cantidad de gaps sin aumentar el tiempo de procesamiento.

# Materiales

El siguiente link contiene el código:

<https://github.com/CecylAglae/AdminDeProyectos.git>

* 1. **Herramientas y Entorno de Desarrollo**

El proyecto emplea Python como lenguaje de programación, aprovechando librerías como NumPy para operaciones matemáticas y manejo de datos y módulos personalizados como fastaReader y evaluadorBlosum para el procesamiento de secuencias genéticas y la evaluación de similitudes.

* 1. **Algoritmo de Forrajeo de Bacterias (BFOA)**

El Algoritmo de Forrajeo de Bacterias (BFOA) es un método de optimización basado en inteligencia de enjambre. En este proyecto, el BFOA se ha adaptado para realizar alineamientos múltiples de secuencias genéticas, simulando interacciones bacterianas para encontrar la mejor configuración de alineamiento. Las bacterias representan soluciones potenciales al problema de alineamiento, y a través de los procesos de quimiotaxis, interacción, y clonación, el BFOA optimiza el fitness del alineamiento.

1. **Clases Principales**:
   * bacteria: Representa cada solución bacteriana en el BFOA, con métodos para insertar gaps, clonar, evaluar el fitness, y eliminar columnas con gaps.
   * chemiotaxis: Administra las interacciones entre bacterias, calculando la atracción y repulsión, y aplica el proceso de quimiotaxis para optimizar el fitness de cada bacteria.
   * fastaReader: Lee archivos en formato FASTA para importar secuencias genéticas y almacenarlas como arreglos manipulables.
   * evaluadorBlosum: Evalúa el puntaje de similitud entre aminoácidos en las secuencias alineadas, utilizando matrices de sustitución BLOSUM.
   1. **Estrategia de Modificación de Parámetros**

Se realizaron ajustes en los parámetros de atracción (d\_attr, w\_attr), repulsión (h\_rep, w\_rep) y tamaño de la población bacteriana para reducir las evaluaciones de funciones (NFE) y mejorar la precisión del alineamiento. Los cambios clave incluyen:

* **Ajuste de Atracción y Repulsión**: La atracción favorece la proximidad entre soluciones similares (mejores alineamientos), mientras que la repulsión reduce la exploración en áreas de bajo fitness.
* **Modificación de Población**: Se aplicó un esquema de eliminación y clonación, en el cual las bacterias menos aptas se eliminan y las más aptas generan clones para diversificar la población.
  1. **Procesamiento de Secuencias y Penalización de Gaps**

El proceso de alineamiento introduce gaps en las secuencias para maximizar las coincidencias. Sin embargo, la presencia de gaps en columnas completas reduce la calidad del alineamiento. Para abordar esto:

* Se emplea el método tumboNado, que inserta gaps en posiciones aleatorias, y limpiaColumnas, que elimina las columnas con gaps en todas las secuencias.
* La función de penalización considera tanto el número de gaps como la variabilidad en los puntajes BLOSUM de las columnas alineadas, optimizando el fitness.

# Resultados.

La simulación del Algoritmo de Optimización por Forrajeo Bacteriano (BFOA) fue realizada en dos etapas: una versión inicial (antes de las mejoras) y una versión optimizada (después de las mejoras) para evaluar su impacto en la eficacia del proceso de optimización.

**6.1 Comparación de Parámetros y Configuración:**

* En ambas versiones se mantuvo una población de 8 bacterias, con 2 seleccionadas aleatoriamente en cada iteración. Sin embargo, se realizaron ajustes en los coeficientes de atracción y repulsión en la versión mejorada para optimizar la interacción entre bacterias.
* La versión mejorada utilizó un enfoque más robusto en los movimientos de nado para maximizar la eficiencia de búsqueda y minimizar el riesgo de estancamiento.

**6.2 Resultados del Fitness:**

* **Antes de la mejora:** Los valores de fitness obtenidos fluctuaban dentro de un rango limitado, con una ligera mejora en las primeras iteraciones, pero sin grandes saltos hacia un óptimo global. Se observó que algunas bacterias lograban mejorar su rendimiento de manera marginal, pero en general, el proceso era ineficaz para alcanzar soluciones de alta calidad.
* **Después de la mejora:** Se registraron variaciones más pronunciadas en el fitness, con algunas bacterias alcanzando valores extremadamente altos o anómalos, como -1.94 × 10^17 y -8.22 × 10^55. Aunque estas fluctuaciones sugieren una mejora en la capacidad de exploración, también indican posibles inestabilidades numéricas o desafíos en la precisión.

**6.3 Evolución y Desempeño de la Población:**

* **Antes de la mejora:** La población de bacterias mostró un patrón de eliminación lento, donde las bacterias menos aptas permanecían activas durante más tiempo, afectando la eficiencia global del algoritmo.
* **Después de la mejora:** Se observó una eliminación más rápida y efectiva de las bacterias menos aptas, con una reducción significativa en la población activa desde 6 bacterias al inicio (NFE: 16) a solo 4 hacia las etapas finales. Esto refleja un comportamiento más adaptativo y eficiente.

**6.4 Análisis del Impacto:**

* **Antes de la mejora:** El algoritmo presentaba limitaciones en la capacidad de exploración y explotación del espacio de búsqueda, con mejoras en el fitness que se estancaban rápidamente.

A screenshot of a computer

Description automatically generated

Figura 1. Corrida antes de la mejora.

* **Después de la mejora:** Se observó una mayor diversidad de soluciones y una búsqueda más agresiva, aunque esto llevó a situaciones de inestabilidad numérica en algunas iteraciones.

A screenshot of a computer program

Description automatically generated

Figura 2. Corrida después de la mejora.

# Conclusiones.

El estudio destaca la efectividad de las mejoras propuestas al Algoritmo de Forrajeo de Bacterias (BFOA) para el alineamiento múltiple de secuencias genéticas. Las modificaciones en los parámetros de interacción bacteriana, como atracción, repulsión y tamaño de la población permitieron optimizar el fitness de las soluciones y reducir el número de evaluaciones de funciones necesarias. Esto resultó en una mayor precisión del alineamiento y una gestión más eficiente de los gaps, contribuyendo a un procesamiento más rápido de las secuencias.

Sin embargo, aunque las mejoras incrementaron la capacidad de exploración y aceleraron la convergencia hacia soluciones óptimas, se observaron inestabilidades numéricas en ciertos casos, lo que sugiere la necesidad de un equilibrio cuidadoso entre precisión y estabilidad. Este avance proporciona un enfoque más robusto para el análisis de datos genéticos en bioinformática, facilitando estudios más eficientes en biología molecular y genética, y destaca la relevancia de la optimización de algoritmos basados en inteligencia de enjambre para abordar problemas computacionales complejos en este campo.

Esta investigación establece una base sólida para futuras mejoras en el BFOA, enfocándose en la estabilidad numérica y la adaptabilidad del algoritmo para procesar secuencias genéticas de gran escala.

# Referencias.

*Artículos de revistas:*

[1] A. Bayat, “Bioinformatics,” *BMJ: British Medical Journal*, vol. 324, pp. 1018-1022, 2002. <https://doi.org/10.1136/bmj.324.7344.1018>.

[2] J. Gauthier, A. Vincent, S. Charette, y N. Derôme, “A brief history of bioinformatics,” *Briefings in Bioinformatics*, 2018. <https://doi.org/10.1093/bib/bby063>.

[3] B. Chowdhury y G. Garai, “A review on multiple sequence alignment from the perspective of genetic algorithm,” *Genomics*, vol. 109, no. 5-6, pp. 419-431, 2017. <https://doi.org/10.1016/j.ygeno.2017.06.007>.

[4] F. Naznin, R. Sarker, y D. Essam, “Progressive Alignment Method Using Genetic Algorithm for Multiple Sequence Alignment,” *IEEE Transactions on Evolutionary Computation*, vol. 16, pp. 615-631, 2012. <https://doi.org/10.1109/TEVC.2011.2162849>.

[5] L. Abualigah, A. Diabat, S. Mirjalili, M. Elaziz, y A. Gandomi, “The Arithmetic Optimization Algorithm,” *Computer Methods in Applied Mechanics and Engineering*, 2021. <https://doi.org/10.1016/J.CMA.2020.113609>.

[6] J. Zhang, Z. Cui, Y. Wang, H. Wang, X. Cai, J. Chen, y W. Li, “A Coupling Approach With GSO-BFOA for Many-Objective Optimization,” *IEEE Access*, vol. 7, pp. 120248-120261, 2019. <https://doi.org/10.1109/ACCESS.2019.2937538>.

1. Ing. en sistemas computacionales, Facultas de sistemas, Universidad Autónoma de Coahuila, México. [↑](#footnote-ref-1)