

# Laboratorio 03a: Ensamblaje de Fragmentos de ADN

Frank Roger Salas Ticona

10 de octubre de 2024

## 1. Introducción

El ensamblaje de fragmentos de ADN es una tarea fundamental que permite reconstruir secuencias genéticas completas a partir de lecturas cortas. Utilizando un C para el procesamiento eficiente de datos y Python para la visualización de resultados. Se analizarán los resultados obtenidos.

## 2. Implementación

Se utilizó C puro para implementar los algoritmos de ensamblaje de ADN, lo que permitió una manipulación eficiente de las estructuras de datos y un control total sobre el proceso de cálculo. La implementación se optimizó para manejar grandes volúmenes de datos de secuencias, asegurando que las ejecuciones fueran rápidas y efectivas. En el [link](#) del repositorio de GitHub se encuentran los archivos con la implementación completa.

## 3. Resultados

Se consiguió verificar el funcionamiento del algoritmo de manera correcta obteniendo una secuencia de consenso con una longitud de 55.

### 3.1. Experimentos

Se pueden observar en la figura [2](#) la salida del programa en C para el ensamblaje de secuencias, en el que se puede evidenciar también como sería ensamblada cada secuencia. En la figura ?? se observa el camino hamiltoniano obtenido a partir de la salida del programa en C.

```
daerean [1] Lab 04 main ▶ 11
total 240
-rw-r--r-- 1 daerean daerean 132k Oct 10 12:57 03a_Assemblyfragments_2024-11.pdf
-rw-r--r-- 1 daerean daerean  24 Oct 10 13:46 best_path.txt
-rw-r--r-- 1 daerean daerean  59k Oct 10 13:46 dna_assembly_graph.png
-rw-r--r-- 1 daerean daerean  36 Oct 10 13:46 graph_structure.txt
daerean [1] Lab 04 main ▶ 12
-rw-r--r-- 1 daerean daerean  43k Oct 10 12:59 latex
-rw-r--r-- 1 daerean daerean  7.2k Oct 10 13:46 main.c
-rw-r--r-- 1 daerean daerean  1.9k Oct 10 13:46 main.py
-rwxr-xr-x 1 daerean daerean  20k Oct 10 13:46 output
daerean [1] Lab 04 main ▶ 13
Best Consensus Sequence: ATCCGTTGAAGCCGCGGGCCTCGAGTTAGTACTGCCCGACACAGATTGTGTG
Length: 55
Aligned Sequences:
-----
ATCCGTTGAAGCCGCGGGC-----
-----CCTCGAGTTAA-----
-----TTAGTTACTGCCCG-----
-----CCGACACAGAT-----
CGATTCGACAGCA-----
CACAGATCCTTGAAGCCGCGGG-----
-----CTCGAGTTAGTA-----
CGCGGCGAGTCTT-----
Consensus: ATCCGTTGAAGCCGCGGGCCTCGAGTTAGTACTGCCCGACACAGATTGTGTG
Difference Matrix:
0 0 0 0 0 0 1
0 0 0 0 0 0 0
0 0 0 0 1 0 0
0 0 0 0 0 0 0
0 0 0 0 1 0 0
1 0 0 0 0 0 1
0 0 1 0 0 0 0
0 1 1 0 0 0 0
Graph structure output to graph_structure.txt
Debug: Best path found:
0 -> 2 (edge exists: Yes)
2 -> 4 (edge exists: Yes)
4 -> 5 (edge exists: Yes)
5 -> 0 (edge exists: Yes)
0 -> 7 (edge exists: Yes)
7 -> 1 (edge exists: Yes)
Best path output to best_path.txt
daerean [1] Lab 04 main ▶ 14
```

Figura 1: Algoritmo para el ensamblaje para las secuencias de ejemplo.

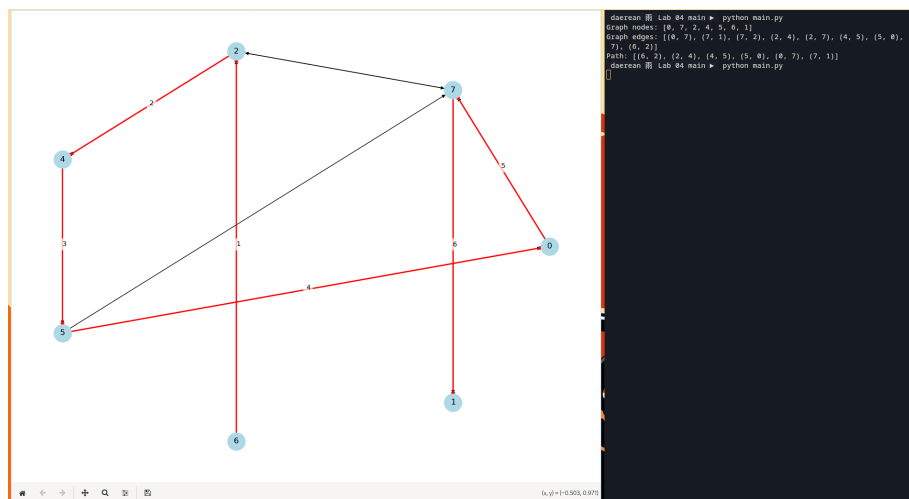


Figura 2: Visualización del camino hamiltoniano.

## 4. Análisis y Discusión

En esta implementación del algoritmo de ensamblaje de ADN, se obtuvo una secuencia consensuada de 55 nucleótidos, destacando la capacidad del algoritmo para alinear efectivamente múltiples secuencias. La matriz de adyacencia generada muestra cómo se relacionan las diferentes secuencias, lo que es importante para entender las conexiones filogenéticas entre ellas. Estos resultados indican que el algoritmo es útil para procesar datos biológicos, proporcionando información valiosa sobre las relaciones entre las secuencias.

## 5. Conclusiones

Los resultados demuestran que la implementación del algoritmo de ensamblaje de ADN es eficaz para generar secuencias consensuadas a partir de datos de secuencias alineadas. La eficiencia del algoritmo permite su aplicación en investigaciones genómicas, facilitando el análisis de relaciones evolutivas entre secuencias. Esto resalta la relevancia de utilizar enfoques optimizados en bioinformática para obtener conclusiones significativas a partir de datos complejos.