



Universidad Católica
San Pablo

CIENCIA DE LA COMPUTACIÓN

Biología Molecular Computacional

Laboratorio - 03 Ensamblaje de Fragmentos de ADN

Rodrigo Alonso Torres Sotomayor

CCOMP 9-1

"El alumno declara haber realizado el presente trabajo de acuerdo a las normas de la Universidad Católica San Pablo"

Laboratorio 03: Ensamblaje de Fragmentos de ADN

Desarrollo

Para este trabajo se utilizó el lenguaje de programación C++.

0.1 Secuencia de Consenso

Implemente un algoritmo que, para un conjunto de cadenas encuentre:

- 1. La secuencia de consenso.
- 2. Entre las posibles secuencias de consenso aquella que se aproxime más a un tamaño determinado l .
- 3. Que considere tanto secuencias directas como complemento-reversas

Verificar su funcionamiento para el siguiente conjunto de secuencias considerando que la longitud l de la molécula destino es próxima a 55 pares, y existen secuencias que pueden ser del otro strand (usar complementos reversos).

```
 $f_1$ : ATCCGTTGAAGCCGCGGGC  
 $f_2$ : TTAAGTCGAGG  
 $f_3$ : TTAAGTACTGCCCG  
 $f_4$ : ATCTGTGTCGGG  
 $f_5$ : CGACTCCCGACACA  
 $f_6$ : CACAGATCCGTTGAAGCCGCGGG  
 $f_7$ : CTCGAGTTAAGTA  
 $f_8$ : CGCGGGCAGTACTT
```

0.2 Subgrafos Acíclicos

Implementar el algoritmo de búsqueda del camino hamiltoniano con linkage t (parámetro), y aplíquelo a las cadenas anteriores definiendo un valor t de acuerdo al nivel de linkage que se pudo evidenciar al buscar la secuencia de mejor consenso. Visualizar el grafo y las secuencias enlazadas junto al string encontrado

0.3 Resultados

La mejor secuencia para obtener el consenso fue una cadena de 48 caracteres. El valor de linkage tuvo que ser pequeño para obtener un mejor consenso, a partir de 3 ya comenzaba a disminuir. Los resultados se pueden observar en las Figure 1 y 2.

Link del repositorio

<https://github.com/RodATS/Molecular.git>

```

Secuencia de Consenso: CTCAGTCCAGGCCAAGCCGCGGG
-----
Cadena 0:
ATCCGTTGAAGCCGCGGG -> CGACTCCCGACACA -> CACAGATCCGTTGAAGCCGCGGG -> CTCGAGTTAAGTA -> CGCGGGCAGTACTT -> CCTCGAGTTAA-
> CGGGCAGTACTTAA -> CCCGACACAGAT -> CCCGCGGCTTCAACGGATCTGTG ->
Cadena 1:
TTAACTCGAGG -> GCCCGCGGCTTCAACGGAT ->
Cadena 2:
TTAAGTACTGCCCG -> CGACTCCCGACACA -> CGCGGGCAGTACTT -> GCCCGCGGCTTCAACGGAT -> CGGGCAGTACTTAA ->
Cadena 3:
ATCTGTGTCGGG -> GCCCGCGGCTTCAACGGAT ->
Cadena 4:
CGACTCCCGACACA -> ATCCGTTGAAGCCGCGGGC -> ATCTGTGTCGGG -> AAGTACTGCCCGG ->
Cadena 5:
CACAGATCCGTTGAAGCCGCGGG -> ATCCGTTGAAGCCGCGGGC -> CGCGGGCAGTACTT -> GCCCGCGGCTTCAACGGAT ->
Cadena 6:
CTCGAGTTAAGTA -> ATCCGTTGAAGCCGCGGGC -> ATCTGTGTCGGG -> AAGTACTGCCCGG ->
Cadena 7:
CGCGGGCAGTACTT -> TTAAGTACTGCCCG -> TTAAGTACTGCCCG -> TGTGTCGGGAGTCG -> TACTTAACTCGAG ->
Cadena 8:
GCCCGCGGCTTCAACGGAT -> ATCCGTTGAAGCCGCGGGC -> TTAAGTACTGCCCG -> TTAAGTACTGCCCG -> TGTGTCGGGAGTCG -> TACTTAACTCGAG ->
Cadena 9:
CCTCGAGTTAA -> ATCCGTTGAAGCCGCGGGC -> ATCTGTGTCGGG -> AAGTACTGCCCGG ->
Cadena 10:
CGGGCAGTACTTAA -> ATCCGTTGAAGCCGCGGGC -> ATCTGTGTCGGG -> AAGTACTGCCCGG ->
Cadena 11:
CCCGACACAGAT -> ATCCGTTGAAGCCGCGGGC -> TTAAGTACTGCCCG -> TTAAGTACTGCCCG -> TGTGTCGGGAGTCG -> TACTTAACTCGAG ->
Cadena 12:
TGTGTCGGGAGTCG -> CGACTCCCGACACA -> CGCGGGCAGTACTT -> GCCCGCGGCTTCAACGGAT -> CGGGCAGTACTTAA ->
Cadena 13:
CCCGCGGCTTCAACGGATCTGTG -> ATCTGTGTCGGG -> GCCCGCGGCTTCAACGGAT ->
Cadena 14:
TACTTAACTCGAG -> TTAAGTACTGCCCG -> CTCGAGTTAAGTA -> GCCCGCGGCTTCAACGGAT ->
Cadena 15:
AAGTACTGCCCGG -> CGACTCCCGACACA -> CGCGGGCAGTACTT -> GCCCGCGGCTTCAACGGAT -> CGGGCAGTACTTAA ->

```

Figure 1: Todas las combinaciones posibles para encontrar la secuencia de consenso.

```

-----
Secuencia proxima a longitud 55: CACAGATCCGTTGAAGCCGCGGGATCCGTTGAAGCCGCGGGCGCGGGG
Tamaño: 48
Camino Hamiltoniano:
CACAGATCCGTTGAAGCCGCGGG -> ATCCGTTGAAGCCGCGGGC -> CGCGGGCAGTACTT -> GCCCGCGGCTTCAACGGAT -> CACAGATCCGTTGAAGCCGCGGGATCCG
TTGAAGCCGCGGGCGCGGGG ->

```

Figure 2: Secuencia más proxima a la longitud.