

Laboratorio 01a: MSA Estrella

Frank Roger Salas Ticona

11 de septiembre de 2024

1. Introducción

El alineamiento múltiple de secuencias (MSA) es una técnica fundamental en bioinformática que permite comparar y analizar múltiples secuencias biológicas simultáneamente. En este informe, nos centraremos en la implementación y análisis del algoritmo MSA Estrella, propuesto por Gusfield en 1997, como una aproximación eficiente para realizar alineamientos múltiples.

2. Implementación

Basándose en la implementación anterior de Needleman-Wunsch se hizo el código para MSA Estrella. El código está en el [repositorio de GitHub](#).

3. Resultados

3.1. Experimentos

La tabla de puntaje producto del alineamiento por pares es la siguiente:

0	7	-2	0	-3
7	0	-2	0	-4
-2	-2	0	0	-7
0	0	0	0	-3
-3	-4	-7	-3	0

Así mismo, la secuencia final de alineamiento es la siguiente:

```
ATTGCCATT
ATGGCCATT
A-TCCAATTTT
ATCTTC-TT
ACTG--ACC
```

Se obtuvo el resultado esperado. Posteriormente, la ejecución con el archivo BRCA1.txt fue la siguiente:

```

daerean 雨 Lab 02 main ▶ ./output BRCA1.txt
 0      -15     -9     -8     -2     -5
-15      0     -11     4     -14     -1
-9      -11     0     -7     -12     -7
-8       4     -7     0     -14     2
-2      -14    -12    -14     0     -9
-5      -1     -7     2     -9     0

Sequence with maximum sum: 5 -TGACGTGTCTGCTCCACTTCCA

Alignment between -TGACGTGTCTGCTCCACTTCCA and -TGCCGGCAGGGATGTGCTTG:
-TGACGTGTCTGCTCCACTTCCA
-TGCCGGCAGGGATGTGCTT--G
Alignment between -TGACGTGTCTGCTCCACTTCCA and -GTTTAGGTTTTGCTTATGCAGCATCCA:
-TGACGTGTCTGCTCCACTTCCA
-GTTTAGGTTTTGCTTATGCAGCATCCA
Alignment between -TGACGTGTCTGCTCCACTTCCA and -GGAAAAGCACAGAACTGGCCAACA:
-TGACGTGTCTGCTCCACTTCCA
-GGAAAAGCACAGAACTGGCCAACA
Alignment between -TGACGTGTCTGCTCCACTTCCA and -GCCAGTTGGTTGATTTCACCTCCA:
-TGACGTGTCTGCTCCACTTCCA
--GCCAGTTGGTTGATTTCACCTCCA
Alignment between -TGACGTGTCTGCTCCACTTCCA and -ACCCCCGACATGCAGAAAGCTG:
-TGACGTGTCTGCTCCACTTCCA
-ACCCCCGACATGC-AGA-AGCTG

-TGACGTGTCTGCTCCACTTCCA
-TGCCGGCAGGGATGTGCTT--G
-GTTTAGGTTTTGCTTATGCAGCATCCA
-GGAAAAGCACAGAACTGGCCAACA
--GCCAGTTGGTTGATTTCACCTCCA
-ACCCCCGACATGC-AGA-AGCTG
 0      -3     -6     -2     -2     -8
-3      0     -12     -9     -7     -6
-6     -12     0     -5     -4     -3
-2     -9     -5     0     -12     -6
-2     -7     -4     -12     0     -6
-8     -6     -3     -6     -6     0

Sequence with maximum sum: 0 -TGCTTGCAAGTTGCTTTCACTGATGGA

Alignment between -TGCTTGCAAGTTGCTTTCACTGATGGA and -TCAGGTACCCTGACCTTCTCTGAAC:
-TGCTTGCAAGTTGCTTTCACTGATGGA
-T-CAGGTACCCTGACCTTCTCTGA--AC
Alignment between -TGCTTGCAAGTTGCTTTCACTGATGGA and -GTGGTTGTAAAGGTCCCAAATGGT:
-TGCTTGCAAGTTGCTTTCACTGATGGA
--G-TGGTTGTAAAGGTCCCAAATGGT
Alignment between -TGCTTGCAAGTTGCTTTCACTGATGGA and -TGCTTGGGTCCTCTGACTGG:
-TGCTTGCAAGTTGCTTTCACTGATGGA
-TGCTTGG--GGTCCCTCTGAC---TGG-
Alignment between -TGCTTGCAAGTTGCTTTCACTGATGGA and -GTGGTGCATTGATGGAAGGAAGCA:
-TGCTTGCAAGTTGCTTTCACTGATGGA
-GTG-GTGCA--TTG-ATGGAAGGAAGCA
Alignment between -TGCTTGCAAGTTGCTTTCACTGATGGA and -AGTGAGAGGAGCTCCAGGGC:
-TGCTTGCAAGTTGCTTTCACTGATGGA
-AG--TG-AG-AGGAGCTC-C-CAGGGC

-TGCTTGCAAGTTGCTTTCACTGATGGA
-T-CAGGTACCCTGACCTTCTCTGA--AC
--G-TGGTTGTAAAGGTCCCAAATGGT
-TGCTTGG--GGTCCCTCTGAC---TGG-
-GTG-GTGCA--TTG-ATGGAAGGAAGCA
-AG--TG-AG-AGGAGCTC-C-CAGGGC

```

Figura 1: Ejecución del código con secuencias propuestas.

4. Análisis y Discusión

El rendimiento del algoritmo MSA Estrella está directamente influenciado por la longitud y cantidad de secuencias. La longitud de las secuencias aumenta la complejidad temporal y el uso de memoria, mientras que el número de secuencias afecta más significativamente el rendimiento global, debido a las múltiples comparaciones necesarias entre ellas.

5. Conclusiones

El algoritmo MSA Estrella es eficaz para conjuntos de datos moderados, aunque presenta limitaciones con secuencias largas o numerosas. Con mejoras como el alineamiento progresivo y la paralelización, sigue siendo una herramienta útil en bioinformática, especialmente para análisis preliminares.