



Biología Molecular Computacional CB309

Laboratorio 01a: MSA Estrella

Yván Jesús Túpac Valdivia
Universidad Católica San Pablo, Arequipa – Perú
05 de setiembre de 2024

1 Objetivos

- Implementar el algoritmo de alineamiento MSA Estrella [Gusfield, 1997] en un lenguaje de programación.
- Verificar su funcionamiento y comportamiento computacional ante diversos tamaños y cantidades de cadenas
- Analizar la calidad de los resultados obtenidos.
- Analizar el impacto de la longitud y cantidad de cadenas ¿Cuál factor impacta más?
- Valorizar la utilidad de este algoritmo.

2 Desarrollo

2.1 Algoritmo MSA Estrella: Implementación

- Conforme se vio en clase, se tiene el algoritmo de alineamiento MSA Estrella [Gusfield, 1997] que permite alinear k secuencias s_1, s_2, \dots, s_k .
- Implementar el algoritmo en C++ preferentemente.
- Evalúe qué recursos del lenguaje deberían aprovecharse, utilícelos. [SO1, SO2]

2.2 Algoritmo MSA Estrella: Pruebas y ajustes

- Probar su implementación en el ejemplo visto en clase: alinear las cadenas:

S_1	A	T	T	G	C	C	A	T	T
S_2	A	T	G	G	C	C	A	T	T
S_3	A	T	C	C	A	A	T	T	T
S_4	A	T	C	T	T	C	T	T	
S_5	A	C	T	G	A	C	C		

verificando que se obtenga todas las soluciones posibles e indicando el *score* final obtenido. [SO6]

- Revisar el archivo BRCA1.txt. Estas cadenas son 6 muestras de segmentos de BRCA1 (gen que influye en las incidencias de cáncer de mama y ovarios), obtenidas por amplificación LR-PCR.
- Se cuenta con muestras *forward* y *reverse* (es decir que provienen de uno u otro *strand* del ADN). Pruebe alineando solo las cadenas *forward* y luego solo las *reverse*, e interprete lo hallado.

2.3 Algoritmo MSA Estrella: Valorización

- Dé una revisión en la literatura buscando qué logros ha permitido alcanzar este algoritmo en el área de Biología Molecular Computacional y revise también las diversas mejoras que se han investigado y logrado al mismo algoritmo. Indique cuál de las mejoras le parece la más relevante. [S07]

3 Entrega

- Deberá entregar dentro del plazo que aparece en el Aula Virtual un informe (PDF) con lo realizado respondiendo las preguntas y mostrando los funcionamientos.
- Las implementaciones deberán estar en repositorio, cuyo URL debe indicarse en el informe.

References

[Gusfield, 1997] Gusfield, D. (1997). *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*. Cambridge University Press.