

On-line Assembling Mitochondrial DNA from de novo transcriptome

Juan David Arcila Moreno
Santiago Passos Patiño
Medellín, Mayo 15 - 2017

Estructuras de Datos utilizadas

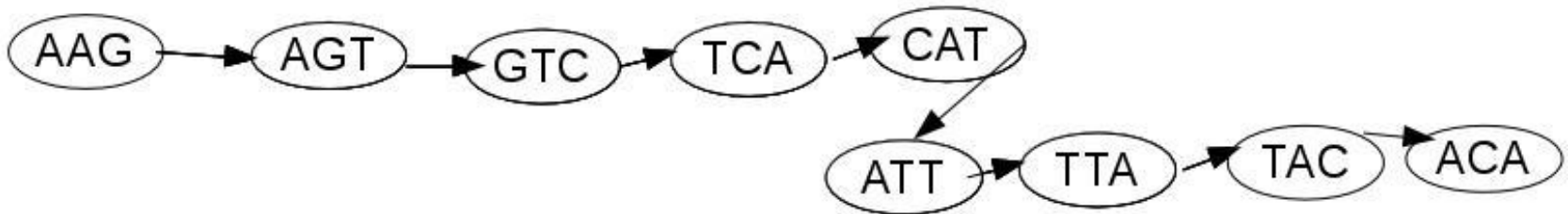
Para el caso de la cadenas:



AAGTCAT



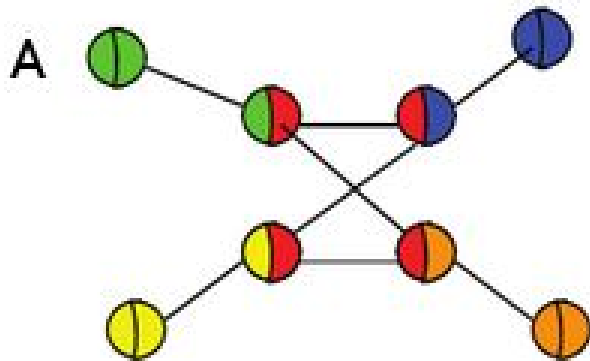
CATTACA



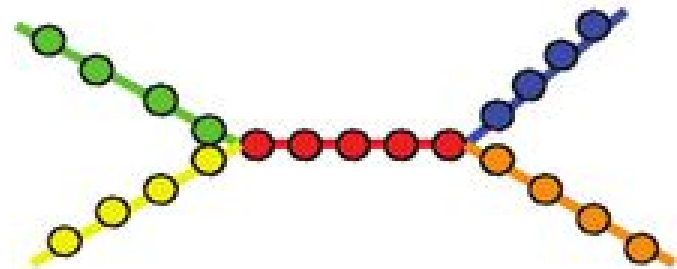
AAGTCATTACA

Estructuras de Datos utilizadas

En la figura A un k-overlap graph.
En la figura B Brujin graph



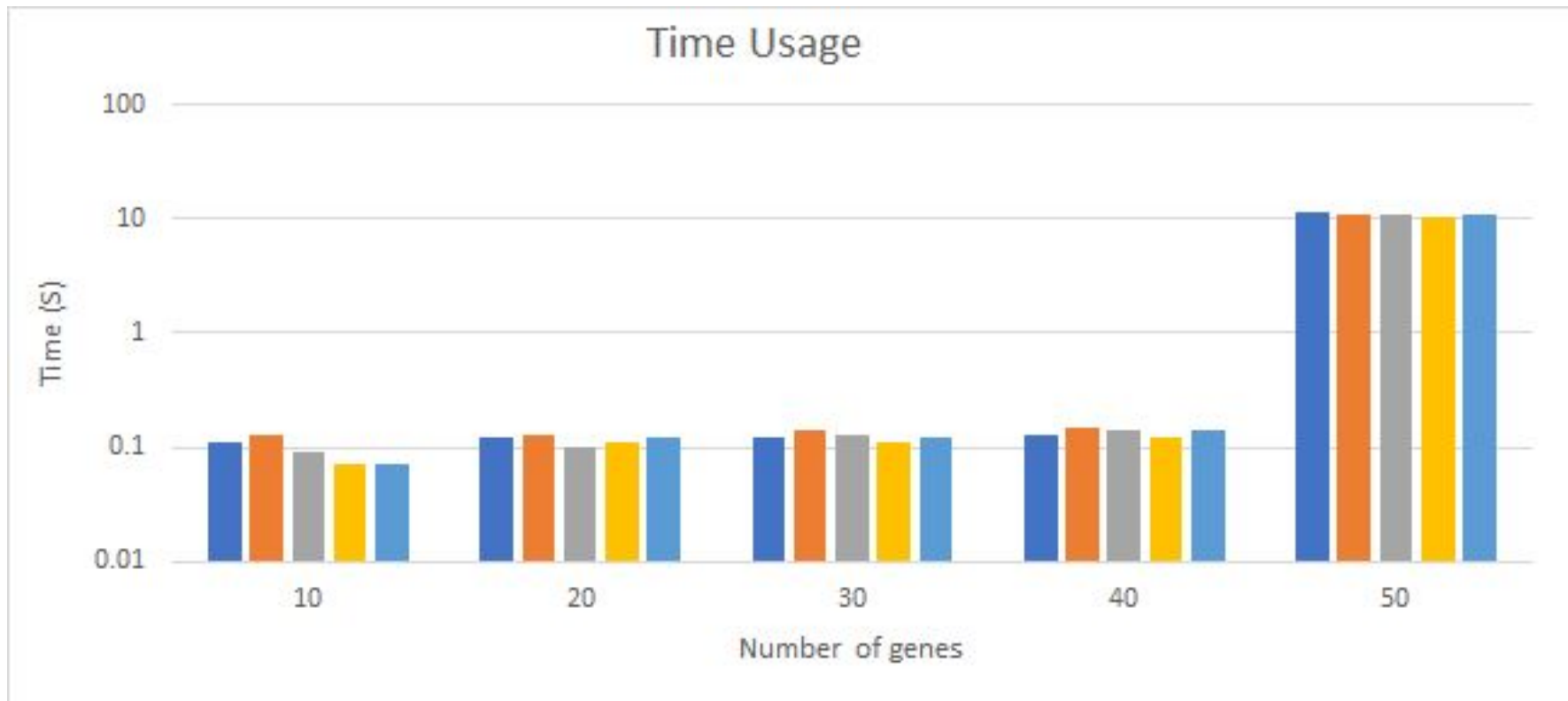
B



Algoritmo y Análisis de complejidad

Algoritmo	Variables	Complejidad
BooyerMoore	n: Cadena de búsqueda m: patrón a buscar	$O(n/m)$
División en k-mers	S: tamaño de la cadena n : tamaño del k-mer	$O(S-n)$
Brujin Graph	e: número de k-mers generados	$O(e)$

Algoritmo y Análisis de complejidad



Comparación de soluciones

Solution	Time complexity	Memory complexity
Substrings (Brute force)	$O(n!)$	$O(n!)$
k-Overlap graph (TSP)	$O(2^n * n^2)$	$O(2^n * n)$
Bruijn Graph (eurealian path)	$O(nm+e)$	$O(e)$

n = number un segments to read

m = time to get a segment into k-mers, always $m < n$ in terms of time

e = time to do the eurealian path or in other words the number of kmers, y the length of each segments minus 201

Aplicación de Software funcionando

Problemas durante el desarrollo y soluciones

- La abstracción del problema.
- El tamaño de las cadenas y la cantidad de segmentos.
- Decidir cuál era la solución más eficaz.
- Mejorar el rendimiento del algoritmo.
- El campo teórico del problema (Biología).

¿Qué hizo cada integrante del equipo?

Mar 12, 2017 – May 15, 2017

Contributions to master, excluding merge commits

Contributions: **Commits** ▾

