

Patrones en Conjuntos de Señales Discretas

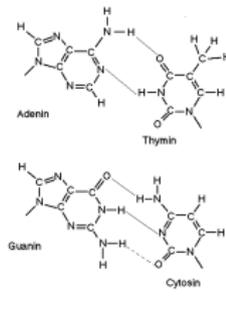
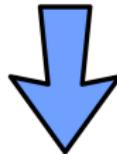
I Jornada de Tesistas de Grado de la ComElec

Melisa Halsband

11 de agosto de 2010



EI ADN



El Input

Tenemos algunos o varios GB de

El Input

Tenemos algunos o varios GB de

tgcatagttgttaaaatttgttgttagtggggga
tctataatcttgctctgccctctgcctgatttctatt
tgtgtttggtatcattgggtatgtgcatgagcagatga
ttgggtgaagagttctttagatgtctactaggtccat
tcttgaatatctttgttagctttctccccagtgatct
aatctctactattactatgcagttatctaagtgttt
ttatgaatctgagtggttctgagtttgggtgcatatata
gtcgaattgaaccctttaccattttgtaatacccttct
gtttaaagtctgtttggctgaaattagaatagcaacc
cttggtaggggtctctccatccctttactttgaggcta
gggtctcttgaagacagcatactattgagtcttgcttc
gctttttaattggaaaattttgactgtttatattcaag
ttgatcccgttattatattgtagctggttattattca
ataatgacaatggtctatgtaattatgtgtattttgt
ttcatatttagcactccctttaggaccttttgaagg
tctctcaacatttacttgtccgaaaaggatattatttc
ttggctggatatgaaattttggttggaaatttcttta

El Input

Tenemos algunos o varios GB de

```

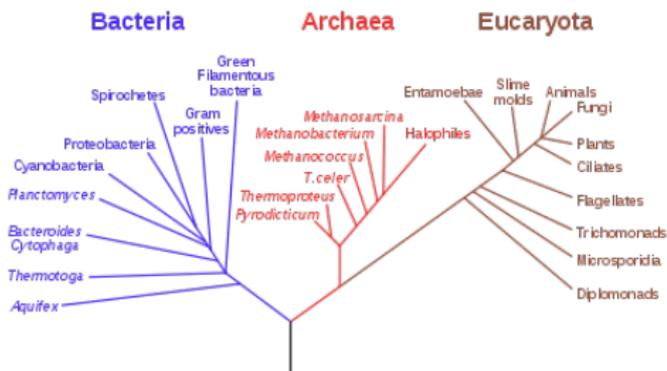
tgcagatagttgttaaattttgttgttagtggggga
tctataatcttgctctgccctctgcctgatttctatt
tgtgtttggtatcattgggtatgtgcatgagcagatga
ttgggtgaagagttctttagatgtctactaggtccat
tcttgaatatctttgttagctttctccccagtgatct
aatctctactattactatgcagttatctaagtgtttt
ttatgaatctgagtggttctgagtttgggtgcatatata
gtcgaattgaaccctttaccattttgtaatacccttct
gtttaaagtctgtttggctctgaaattagaatagcaacc
cttggtaggggtctctccatccctttactttgaggcta
gggtctcttgaagacagcatactattgagtcttgcttc
gcttttaattggaaaattttgactgtttatattcaag
ttgatcccgttattatattgtagctggttattattca
ataatgacaatgggtctatgtaattatgtgtattttgt
ttccatatttagcactccctttaggaccttttgtaagg
tctctcaacatttacttgtccgaaaaggatatttttc
ttggctggatatgaaattttggttggaaatttctttta
    
```

Ejemplos

Especie	Tamaño
Virus HIV	10 KB
Escherichia Colli	4 MB
Mosca	130 MB
Álamo	480 MB
Humano	3.1 GB
Ameba	670 GB

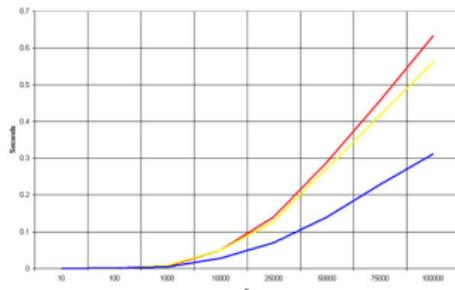
Preguntas Biológicas

- ¿Qué fragmentos se repiten más frecuentemente?
- ¿Qué fragmentos se conservan entre distintas especies?
- ¿Qué secuencias son exclusivas de una especie o conjunto de especies?
- ¿Qué especies tienen genomas más “simples” o “complejos”?
- ¿Cuán parecidas son entre sí distintas especies?



Preguntas Técnicas

- ¿Cuándo una secuencia incluye a otra o es igual a ella?
- ¿Qué medida de similitud o semejanza utilizamos?
- ¿Cómo definir una métrica de “complejidad” de una secuencia?
- ¿Cómo procesar los datos en forma exacta y exhaustiva, pero eficiente?



Definiciones

Repeticiones Maximales: Son aquellas que se repiten más de una vez, y cualquiera de sus extensiones tiene menos ocurrencias.

Repeticiones Supermaximales: Son aquellas que se repiten más de una vez, y cualquiera de sus extensiones tiene a lo sumo una ocurrencia.

Repeticiones de un Conjunto: Son aquellas que pertenecen a todos los elementos del conjunto, y cualquiera de sus extensiones no pertenece a alguno de los elementos.

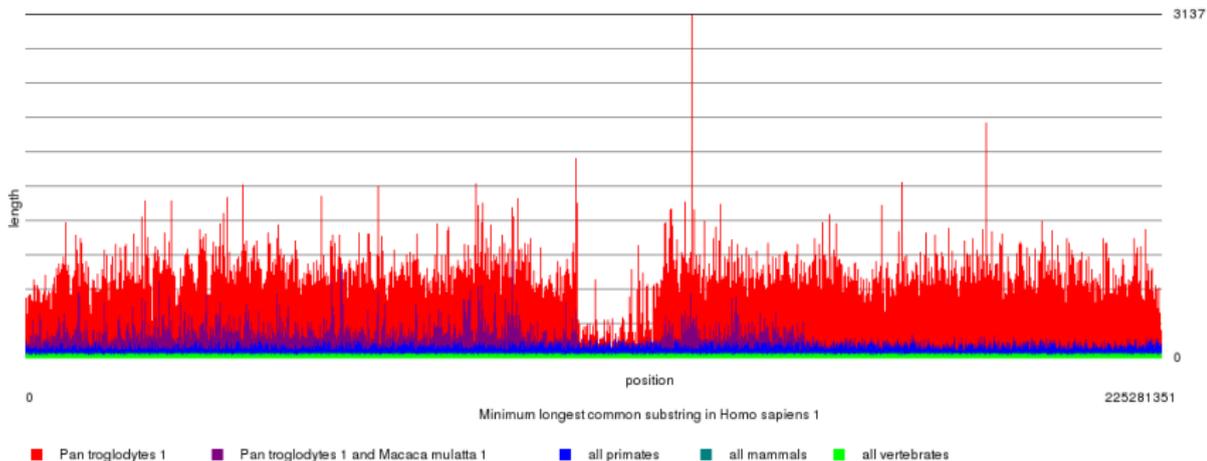
Repeticiones Exclusivas: Son aquellas que se repiten en una secuencia o conjunto de secuencias y no aparecen en ningún elemento de otro conjunto.

Complejidad de una Secuencia: Fórmula no computable, aproximación basada en las longitudes de las repeticiones en cada posición.

Distancia entre secuencias: Definición construida sobre la definición de complejidad.

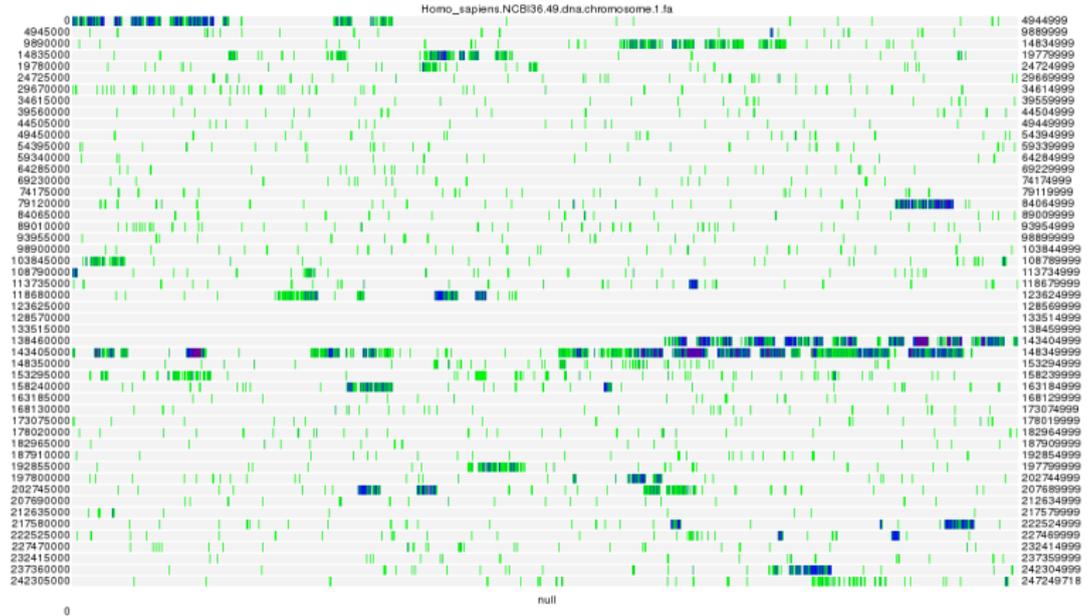
Algunos Resultados

Longitud del patrón del conjunto más largo por posición



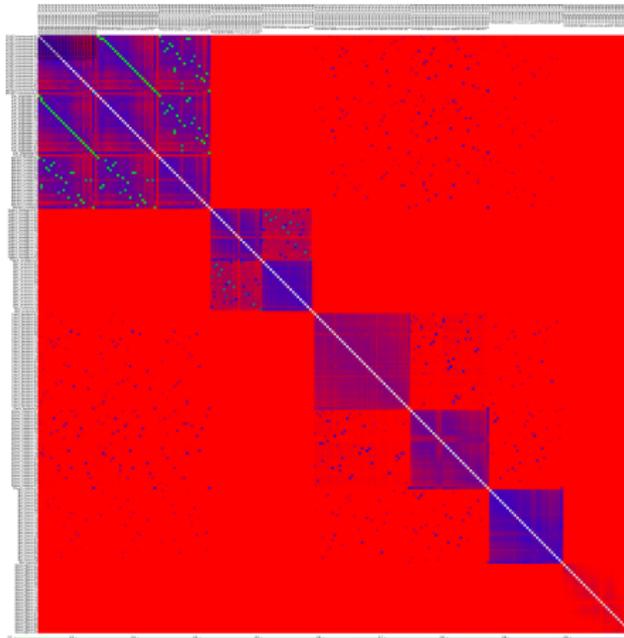
Algunos Resultados

Cubrimiento de un cromosoma con patrones nuevos



Algunos Resultados

“Distancia” entre especies



¡Muchas gracias!

¿Preguntas?

Melisa Halsband
mhalsband@fi.uba.ar

Grupo



<http://kapow.dc.uba.ar>

Comisión de Estudiantes de Electrónica (ComElec)
<http://groups.google.com/group/comelec>