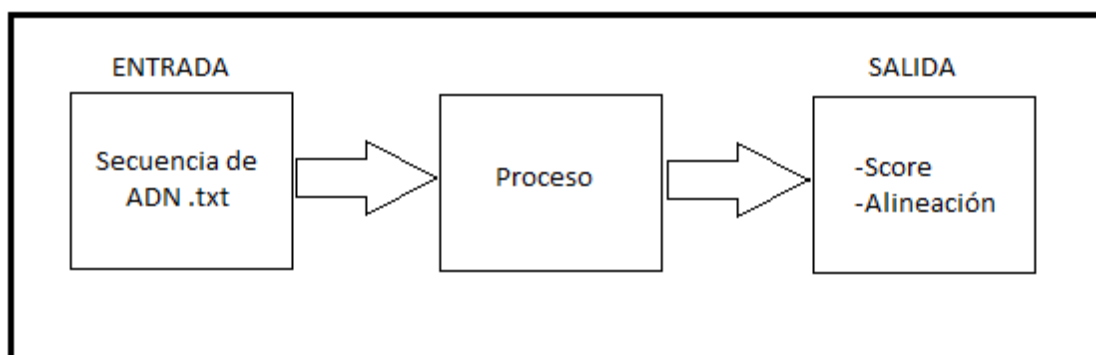


## Control de Avance

### Correspondencia entre secuencias de ADN

1. Objetivos de Problema: Analizar 2 cadenas de ADN para poder encontrar segmentos similares entre ellas.
2. Alcances del problema: El alineamiento de secuencias se realiza para determinar posibles relaciones evolutivas (segmentos similares), mutaciones (segmentos distintos) o indels (2 o más mutaciones genéticas que se consideran como una a causa de su similitud).
3. Desarrollo del problema: Dada 2 secuencias de ADN en formato fasta, se desea encontrar el grado de concurrencia y resultado de alineamiento entre ambas, haciendo uso de un algoritmo basado en la programación dinámica de **Needleman-Wush**. Nuestro programa tendrá como parámetro de entrada un fichero en formato .txt, este contendrá 2 secuencias de ADN en formato fasta. El Programa deberá reconocer el segmento correspondiente a las 2 secuencias de caracteres, para luego realizar un conjunto de operaciones (las detallaremos más adelante), que darán como resultado una puntuación equivalente al grado de concurrencia, junto a la alineación de la secuencia con el objetivo de encontrar segmentos similares. Este resultado, será un material de apoyo para el análisis científico.



..... (Me falta).....

# Plan B

1. Objetivos del problema: Buscar similitudes entre una cadena de ADN dada y algún registro en una base de datos.
2. Alcances del problema:

# 1. DEFINICIÓN

En bioinformática, el formato FASTA es un formato de fichero informático basado en texto, utilizado para representar secuencias bien de ácidos nucleicos, bien de péptido, y en el que los pares de bases o los aminoácidos se representan usando códigos de una única letra. El formato también permite incluir nombres de secuencias y comentarios que preceden a las secuencias en sí.

La simplicidad del formato FASTA hace fácil el manipular y analizar secuencias usando herramientas de procesador de textos y lenguajes de guion como Python y PERL.

## 2. CONSIDERACIONES PARA IMPLEMENTACIÓN

### Formato FASTA

Una secuencia bajo formato FASTA comienza con una descripción en una única línea (línea de cabecera), seguida por líneas de datos de secuencia. La línea de descripción se distingue de los datos de secuencia por un símbolo '>' (mayor que) en la primera columna. La palabra siguiente a este símbolo es el identificador de la secuencia, y el resto de la línea es la descripción (ambos son opcionales). No debería existir espacio entre el '>' y la primera letra del identificador. Se recomienda que todas las líneas de texto sean menores de 80 caracteres. La secuencia termina si aparece otra línea comenzando con el símbolo '>'; esto indica el comienzo de otra secuencia. Un ejemplo simple de una secuencia en el formato FASTA puede ser:

```
>gi|5524211|gb|AAD44166.1| cytochrome b [Elephasmaximusmaximus]
LCLYTHIGRNIYYGSYLYSETWNTGIMLLITMATAFMGYVLPWQGMSFWGATVITNLFSAIPYIGTNLV
EWIWGGFSVDKATLNRFFAFHFILPFTMVALAGVHLTFLHETGSNNPLGLTSDSDKIPFHPYYTIKDFLG
LLILILLLLLLALLSPDMLGDPDNHMPADPLNTPHLIKPEWYFLFAYAILRSVPNKLGGVLALFLSIVIL
GLMPFLHTSKHRSMMLRPLSQALFWTLTMDLLTLTWIGSQPVEYPYTIIGQMASILYFSIILAFPLIAGX
IENY
```

### Codificación

- Python puede implementar librerías para trabajar problemas de Bioinformática y he aquí un tutorial.
- <http://biopython.org/DIST/docs/tutorial/Tutorial.html>
- Código de parseo de momento secuencial.
- <http://biopython.org/DIST/docs/tutorial/Tutorial.html#htoc11>
- Solución a un problema particular de codificación
- <http://stackoverflow.com/questions/13132215/using-biopython-python-to-extra-ct-sequence-from-fasta-file>
- Codificación para conectarse a bases de datos
- <http://biopython.org/DIST/docs/tutorial/Tutorial.html#htoc14>
- <http://biopython.org/DIST/docs/tutorial/Tutorial.html#chapter:entrez>

### 3. WEB de Cómputo

- <http://www.genome.jp/tools/fasta/>
- <http://www.ebi.ac.uk/Tools/sss/fasta/>
- <http://ebi.edu.au/tools/ssearch-protein-similarity-search>
- <http://molbiol-tools.ca/Convert.htm>

### 4. Bases de Datos Fidedignas

Archivos Nombre.fasta

- <http://www.uniprot.org/uniparc/>
- <https://sites.google.com/a/agro.uba.ar/bioinformatica/Home/clase-6-primers-para-pcr>

### 5. Bases de Datos Bibliográficas para la “Cháchara” (Referencias fidedignas y reales)

- Datos Generales <http://www.scielo.cl/>
- Tesis por universidades latinas <http://www.cybertesis.info/>
- Revistas Científicas <http://www.redalyc.org/home.oa>

NOTA:

Para descargar Papers o artículos que se deben pagar acudir a esta página solo con el link (url) del documento <http://www.scihub.org/index.html>. Descarga Gratis!!!

### 6. Referencias Bibliográficas

Mount, David W. *Bioinformatics: Sequence and Genome Analysis*, 2ª ed. Spring Harbor Press, 2004. [ISBN 0-87969-712-1](#). Págs. 45 y siguientes. (Aquí se explica lo que significa cada letra en las secuencias)

### 7. Libros Virtuales

<https://books.google.cl/books?id=8UV5jxkuBZIC&pg=PA14&lpg=PA14&dq=PDF+Mount,+David+W.+Bioinformatics:+Sequence+and+Genome+Analysis,+2%C2%AA+ed.+Spring+Harbor+Press,+2004.+ISBN+0-87969-712-1.&source=bl&ots=wJQoYMIJGO&sig=HzUmhia0Nc7BXWB7WW6mp96kLnE&hl=es&sa=X&ei=gfRPVfKnDIL7gwSzm4CICw&ved=0CCgQ6AEwAg#v=onepage&q=fasta&f=false>

## Extras

- Codificación en Ruby (Está en ingles entendible, y está súper bueno para la chachara y le codificación)
- [http://ac.els-cdn.com/S1877050911001037/1-s2.0-S1877050911001037-main.pdf?\\_tid=b2ee2fa4-f775-11e4-ad11-00000aacb362&acdnat=1431304741\\_913d283a654a4e0633dcd8b5bcab2397](http://ac.els-cdn.com/S1877050911001037/1-s2.0-S1877050911001037-main.pdf?_tid=b2ee2fa4-f775-11e4-ad11-00000aacb362&acdnat=1431304741_913d283a654a4e0633dcd8b5bcab2397)
- 
- Explicación de un software de cálculo.
- [http://fernando.gonzalez.unileon.es/web\\_mex14/guias\\_rapidas/guia\\_rapida\\_COMPARACION\\_SECUENCIAS.pdf](http://fernando.gonzalez.unileon.es/web_mex14/guias_rapidas/guia_rapida_COMPARACION_SECUENCIAS.pdf)