

Ejercicio 2.b)

Blast es un algoritmo de alineamiento local que busca coincidencias a partir de secuencias .FASTA en relación otras especies. Devuelve las subregiones que se parecen entre la secuencia Query y las que se encuentran en las bases de datos. Utilizamos algoritmos de alineamiento local cuando las secuencias no son muy similares entre si pero se sospecha que poseen regiones de similitud, dentro de secuencias más grandes.

Trabajar con secuencias completas requiere mucho procesamiento y dificulta la búsqueda, es por eso que se trabaja con fragmentos de las mismas. Cada coincidencia trae asociado ciertos valores estadísticos que indican que tan fuerte es la asociación. Los valores estadísticos son los siguientes:

- ✓ **Score:** Valor calculado a partir de las coincidencias encontradas en la secuencia, es penalizado por la cantidad de huecos y mismatches. Existen diferentes estrategias para penalizar los cambios de un aminoácido por otro, a través de matrices de sustitución.
- ✓ **Expect (o E-value):** Número de alineaciones correctas esperadas a ser encontradas por azar, es decir, que cuanto menor sea el valor, menor será la posibilidad de que la secuencia haya sido correctamente alineada por azar, y por ende tendrá más chances de ser la secuencia correcta.
- ✓ **Identities:** Mayor número de coincidencias obtenidas para un conjunto de secuencias alineadas.
- ✓ **Positives:** Representa la cantidad de aminoácidos que son exactamente iguales o funcionalmente similares.
- ✓ **Gaps:** Es el porcentaje de huecos que se generan al hacer el alineamiento de secuencias.

Ejemplo obtenido a partir del archivo 0-blast.out

```
Score = 65.1 bits (157), Expect = 6e-10, Method: Compositional matrix adjust.  
Identities = 33/42 (79%), Positives = 36/42 (86%), Gaps = 2/42 (5%)  
  
Query    1735   AQSESAQSKMLSGVGGFVLGLIFLGLGLIIRQRSQKGEEPQG   1776  
||| ||| |  
        AQSESAQSKMLSG+GG VLG+IFLGLGL IR + QKG PQG  
Sbjct    186    AQSESAQSKMLSGIGGLVLGVIFLGLGLFIRHKRQKG--PQG   225
```

La línea indicada con *Query*, corresponde a la secuencia por la que se realizó la consulta, mientras que *Subjct* es la secuencia encontrada en la búsqueda. En la línea del medio pueden observarse los elementos que coinciden entre ambos fragmentos de secuencias: si son iguales entonces se escribe el elemento que tienen en común, si no lo son, ese espacio queda en blanco, por último, el signo más (+) indica que los aminoácidos no son exactamente iguales, pero sí funcionalmente equivalentes.

Otro ejemplo, teniendo la cuenta el resultado obtenido de **2-blast.out**:

```

>P01920.2 RecName: Full=HLA class II histocompatibility antigen, DQ beta
1 chain; AltName: Full=MHC class II antigen DQB1; Flags: Precursor
[Homo sapiens]
Length=261

Score = 167 bits (423), Expect = 6e-45, Method: Compositional matrix adjust.
Identities = 78/92 (85%), Positives = 81/92 (88%), Gaps = 0/92 (0%)

Query   530  EDFVFQFKGMCYFTNGTERVRLVTRYIYNREEYARFDSVGVYRAVTPQGRPDAEYWNSQ   589
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
      EDFV+QFK MCYFTNGTERVR VTRYIYNREEYARFDSV VYRAVTP G PDAEYWNSQ
Sbjct   37   EDFVYQFKAMCYFTNGTERVRYVTRYIYNREEYARFDSDEVYRAVTPLGPPDAEYWNSQ   96

Query   590  KEVLEGTRAELDTVCRHNYEVAFRGILQRRGE   621
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
      KEVLE TRAELDTVCRHNY++ R LQRR E
Sbjct   97   KEVLERTRAELDTVCRHNYQLELRRTLQRRVE   128

```

A la izquierda de cada fragmento de secuencia podemos encontrar el orden que posee el primer elemento del en la secuencia original de la cual proviene (en este ejemplo, 530 indica el comienzo del segmento Query), mientras que a la derecha se puede ver la posición del último elemento del fragmento de secuencia dentro de la secuencia original que lo contiene (589 indica el último segmento de la Query).