**Analisis BLAST punto 2.b:**

A continuación haremos un análisis de los resultados arrojados tras correr un BLAST remoto. Estos son las 8 secuencias que el algoritmo encontró más similares a la dada.

Score E Value

Q99497.2 RecName: Full=Protein/nucleic acid deglycase DJ-1 383 6e-135

Q95LI9.1 RecName: Full=Protein/nucleic acid deglycase DJ-1 383 1e-134

Q5E946.1 RecName: Full=Protein/nucleic acid deglycase DJ-1 368 7e-129

Q7TQ35.1 RecName: Full=Protein/nucleic acid deglycase DJ-1 358 5e-125

O88767.1 RecName: Full=Protein/nucleic acid deglycase DJ-1 357 2e-124

Q99LX0.1 RecName: Full=Protein/nucleic acid deglycase DJ-1 354 2e-123

Q8UW59.1 RecName: Full=Protein/nucleic acid deglycase DJ-1 347 1e-120

Q5XJ36.1 RecName: Full=Protein/nucleic acid deglycase DJ-1 320 8e-110

Como se puede observar dos valores destacan, el E value (E) y el Score (S), los cuales están relacionados.

El valor Expect (E) es un parámetro que describe el número de hits que uno puede "esperar" ver por casualidad al buscar en una base de datos de un tamaño particular. Disminuye exponencialmente a medida que aumenta la puntuación (S) del matching. Esencialmente, el valor E describe el ruido de fondo aleatorio. Por ejemplo, un valor E de 1, asignado a un hit, puede ser interpretado como que en una base de datos del tamaño actual, uno puede esperar ver 1 coincidencia con una puntuación similar simplemente por casualidad.

Cuanto más bajo sea el valor E, o cuanto más cerca esté de cero, más "significativo" será el matching. Sin embargo, tenga en cuenta que las alineaciones cortas prácticamente idénticas tienen valores de E relativamente altos. Esto se debe a que el cálculo del valor E tiene en cuenta la longitud de la secuencia de consulta.

El score (S) es otro indicador estadístico prominente, utilizado además del E-value, en una salida BLAST. (S) mide la similitud de la secuencia independientemente de la longitud de la secuencia de consulta y el tamaño de la base de datos, y se normaliza en base a la puntuación de alineación del par sin procesar. (S) es determinada por la siguiente fórmula: S = (λ × S − lnK)/ ln2 donde λ es la constante de distribución de Gumble, S es la puntuación de alineación cruda, y K es una constante asociada con la matriz de puntuación utilizada. Por lo tanto, cuanto más alta sea la puntuación, más significativo será el partido. La puntuación de bits proporciona un indicador estadístico constante para la búsqueda en diferentes bases de datos de diferentes tamaños o para buscar en la misma base de datos en diferentes momentos a medida que la base de datos se amplía.

Como se puede observar en nuestro casos, toda esta familia de proteínas (acid deglycase DJ-1) posee un valor de Score alto y un E value sumamente bajo, lo cual hace suponer una similitud en los alineamientos muy alto con respecto a nuestra secuencia de referencia.