**Análisis ejercicio 3: BLAST**

**Fundamentos en biología computacional**

**Universidad EAFIT**

**2022-1**

**Realizado por:**

**Sebastian Patiño Baena**

**¿Cuántos genes parálogos hay en el genoma? cumplirán la misma función? con cual trabajaría para hacer filogenética? ¿cómo extraer las secuencias de estos genes?.**

**Imagen que contiene Interfaz de usuario gráfica

Descripción generada automáticamente**

Si filtramos por el 90% de identidad, podemos encontrar nueve secuencias que cumplen este requisito, sin embargo, se trata de secuencias muy cortas que no podrían darnos información clara acerca de su relación con la secuencia que estamos comparando, es por esto que muestro además la lista original sin el filtro del 90%, pues allí podemos encontrar resultados mucho más significativos. Si nos fijamos en las tres primeras secuencias que arroja la búsqueda, encontramos el mismo código (LATX01002120.1) que a pesar de tener valores de identidad del 85% como promedio, estos tienen un E value bastante cercano al cero y sus bit score son bastante cercanos a la longitud de las secuencias ( que en este caso podemos ver que son secuencias mucho más largas y por lo tanto, pueden ser semejantes a la secuencia que estamos comparando), es por esto que considero que estas tres secuencias pueden ser los genes parálogos.

Ahora bien, que sean genes parálogos no nos garantiza que tengan una función similar, en especial las secuencias que tienen menos longitud, si queremos hacer una filogenia pensando que son genes con una función similar debemos entonces utilizar la secuencia de mayor longitud. Ahora bien, si queremos extraer estas secuencias deberíamos hacer un script que nos compare nuestra referencia con estos genes seleccionados y así, que nos extraiga las secciones de cada gen con mayor similitud o con el mismo patrón de nucleótidos que posee la referencia.