**Ejercicio de Alineamiento local**

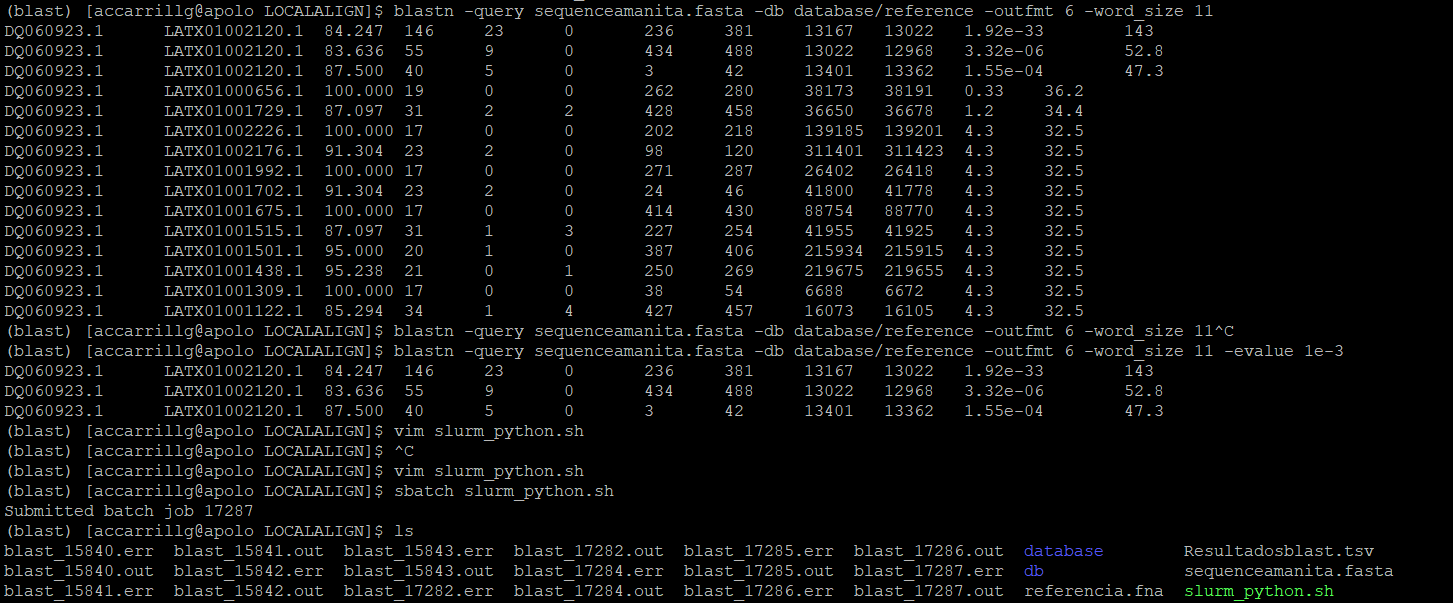
Estudiante: Andrea Carolina Carrillo

Biología computacional

Código hecho en el slurm que creé.

Texto

Descripción generada automáticamente





**Interpretación del resultado:** Estos resultados corresponden a un alineamiento local realizado por BLASTn entre la secuencia DQ060923.1 y la secuencia LATX01002120.1.

Cada línea de los resultados muestra la siguiente información:

1. La identidad de las dos secuencias alineadas (DQ060923.1 y LATX01002120.1)
2. El porcentaje de identidad entre las dos secuencias alineadas (84.247% en la primera línea, 83.636% en la segunda línea y 87.500% en la tercera línea).
3. La longitud de la secuencia alineada en pares de bases (146 en la primera línea, 55 en la segunda línea y 40 en la tercera línea).
4. La cantidad de gaps en la secuencia alineada (23 en la primera línea, 9 en la segunda línea y 5 en la tercera línea).
5. La ubicación del comienzo y del final de la secuencia alineada en ambas secuencias.
6. El valor de e o valor de Expectativa, que es una medida de la probabilidad de que una coincidencia entre dos secuencias sea el resultado del azar. En otras palabras, cuanto menor es el valor de E, mayor es la probabilidad de que la similitud observada entre las dos secuencias sea significativa.
7. El puntaje de bits, que es una medida de la similitud entre las dos secuencias alineadas.

En general, estos resultados sugieren que hay una similitud significativa entre las dos secuencias, sin embargo, creo que no sería confirmar que sean parálogos, o que tengan alguna relación evolutiva. Los hits obtenidos tienen menos nucleótidos que los necesarios para determinar si es parálogo. Estos genes podrían en realidad tener funciones similares o diferentes. La duplicación de genes puede generar nuevas funciones a través de cambios en la secuencia o la regulación de la expresión génica, a veces, pueden incluso tener funciones redundantes y actuar de manera sinérgica, mientras que en otros pueden haber divergido y tener funciones distintas. Sin embargo, la similitud entre las secuencias, especialmente la primera podría sugerir que podrían tener una función similar u homóloga.

Para trabajar con filogenia, usaría la secuencia que tenga más similitud, o sea la que tiene un valor de 84.247% de identidad y un e-value de 1.92e-33.

Para extraer las secuencias de estos genes, podría usar el comando “blastdbcmd”.

**Ejemplo:**

“blastdbcmd -db my\_blast\_db -entry DQ060923.1 -range 13022-13167 -strand plus > secuencia1.fasta

blastdbcmd -db my\_blast\_db -entry LATX01002120.1 -range 12968-13022 -strand minus > secuencia2.fasta

blastdbcmd -db my\_blast\_db -entry DQ060923.1 -range 13362-13401 -strand plus > secuencia3.fasta”

Primero tendría que asegurarme de tener una base de datos de blast con las secuencias de referencia correspondientes a las identificaciones que aparecen en mis resultados de blast. Luego, ejecutaría el comando anterior reemplazando “my\_blast\_db” con el nombre de mi base de datos y las identificaciones con las que quiero trabajar.

Al final tendría tres archivos FASTA con las secuencias correspondientes a las regiones alineadas.