### **UNA PALABRA NO DICE NADA Y AL MISMO TIEMPO LO DICE TODO**

**¿Qué tipo de información se puede extraer de la comparación de secuencias? ¿Cómo esperás que se vea en una comparación?**

Podemos encontrar similitudes que nos permiten identificar si la secuencias tienen algun tipo de relacion. Generalmente las vemos de acuerdo al score y el e-value que tienen.

**¿Por qué crees que es mejor evaluar las relaciones evolutivas lejanas comparando proteínas?**

Porque nos permite armar un arbol filogenetico mas interesante que si todas se encuentran a un mismo nivel. Además de entender como fue evolucionando esa proteina.

**RETO I: Intentemos, entonces alinear estas dos palabras, para comprender mejor el problema. Alineá en la** [**tabla interactiva**](https://flbulgarelli.github.io/umi/#parecido-no-es-lo-mismo) **las palabras "BANANA" y "MANZANA".**

Dependiendo de como los aliniemos obtendremos diferentes resultados. La idea es conseguir la mayor similitud posible entre las distintas cadenas y para ello a veces es necesario utilizar diferentes metodos para poder lograrlo

️ PREGUNTAS DISPARADORAS: ¿Existe una única forma de alinearlas? ¿Es alguno de los posibles alineamientos mejor que otro? Si así fuera ¿Por qué?

**Hay varias formas de alinear las cadenas.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **B** | **A** | **N** | **A** | **N** | **A** |  |
| **M** | **A** | **N** | **Z** | **A** | **N** | **A** |

En esta forma encontramos 3 match.

Y si la cadena la alineamos de otra forma:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **B** | **A** | **N** | **-** | **A** | **N** | **A** |
| **M** | **A** | **N** | **Z** | **A** | **N** | **A** |

**Obtenemos 6 match, duplicando el valor anterior.**

**En la siguiente** [**tabla interactiva**](https://flbulgarelli.github.io/umi/#parecido-no-es-lo-mismo) **distintos alineamientos para las palabras "ANA" y "ANANA". Verás que en el margen superior derecho aparece un valor de identidad calculado para cada alineamiento que intentes.**

**¡Tomá nota de los valores de identidad observados y de las conclusiones que se desprendan de estas observaciones!**

**☑ ️ PREGUNTAS DISPARADORAS: ¿Son todos los valores iguales? ¿Qué consideraciones deberían tenerse en cuenta a la hora de realizar el cálculo? ¿Se te ocurre, distintas formas de calcularlo? ¿Serán todas ellas igualmente válidas en Biología?**

Hemos definido la identidad y hemos comenzado a entender las implicancias de introducir esos guiones, que de ahora en más llamaremos "gaps". La presencia de gaps, que introducen huecos en el alineamiento, representan las inserciones y deleciones. Y cómo pueden intuir, la apertura de un gap en una u otra posición o la persistencia de más de un gap en el alineamiento, tiene sus implicancias.

**Probá en** [**tabla interactiva**](https://flbulgarelli.github.io/umi/#parecido-no-es-lo-mismo) **distintos alineamientos para las palabras "ANA" y "ANANA". Verás que en el margen superior derecho aparece un valor de identidad calculado para cada alineamiento que intentes y un botón para cambiar la penalidad que se le otorga a dicho para el cálculo de identidad.**

**Probá varias combinaciones, tomá nota de los valores de identidad observados y de las conclusiones que se desprendan de estas observaciones.**

**☑ ️ PREGUNTAS DISPARADORAS: ¿Cómo se relacionan los valores de identidad obtenidos con las penalizaciones que se imponen al gap? ¿Qué implicancias crees que tiene una mayor penalización de gaps? ¿Se te ocurre alguna otra forma de penalización que no haya sido tenido en cuenta en este ejemplo?**

En este caso existe varias formas de alinearlas donde sin importar donde agreguemos el gap la identidad se mantiene.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **A** | **N** | **A** |  |  |
| **A** | **N** | **A** | **N** | **A** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **-** | **-** | **A** | **N** | **A** |
| **A** | **N** | **A** | **N** | **A** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **A** | **N** | **-** | **-** | **A** |
| **A** | **N** | **A** | **N** | **A** |

**PARA PENSAR: Entonces, pensando en un alineamiento de ácidos nucleicos ¿Cuáles te parece que son las implicancias de abrir un gap en el alineamiento? ¿Qué implicaría la inserción o deleción de una región de más de un residuo?**

Un gap puede generar un costo en la identidad pero a su vez puede mantener toda la continuidad en la alineacion de la cadena posterior al gap.

**PARA PENSAR: Ingresá al servidor del NCBI y mirá los distintos programas derivados del** [**BLAST**](https://blast.ncbi.nlm.nih.gov/Blast.cgi) **que se ofrecen ¿Para qué sirve cada uno? ¿En qué casos usarías cada uno?**

Blastn buqueda de nucleotidos usando una query de nucleotidos

Blastp buqueda de proteinas usando una query de proteinas

Blastx buqueda de proteinas usando una query de nucleotidos

tblastn busca bases de datos de nucleótidos traducidas mediante una consulta de proteínas.

TBLASTX busca bases de datos de nucleótidos traducidas utilizando una consulta de nucleótidos traducida

¡Vamos a explorar esta herramienta!

**👇 RETO VII: calculá el E-value y porcentaje de identidad utilizando el programa BLAST de la siguiente secuencia input usando 5000 hits, un e-value de 100 y tomando aquellos hits con un mínimo de 70% cobertura. Observe y discuta el comportamiento de : E-value vs. % id, Score vs % id, Score vs E-value**

VVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPVILLISFLIFLIVG

unnamed protein product [Homo sapiens] Homo sapiens 281 281 100% 5e-95 100.00% 163 BAG52189.1

**👇 RETO VIII: Realizá nuevas búsquedas usando la mitad de la secuencia problema y para un cuarto de la secuencia original. Compará los gráficos obtenidos. ¿Qué conclusiones puede sacas?**

Que al tener la cadena incompleta las homologas que aparecen son mas inexactas en comparacion con la cadena completa.

A partir de los resultados de una búsqueda con BLAST se pueden inferir relaciones funcionales o estructurales entre secuencias homólogas. Ya que esta búsqueda asume una relación evolutiva, es posible de este modo identificar nuevos miembros de una familia de genes o de proteínas o encontrar secuencias idénticas, con una significancia estadística.

**👇 RETO IX: Utilizando** [**BLAST**](http://www.ncbi.nlm.nih.gov) **utilice búsquedas de similitud secuencial para identificar a la siguiente proteína:**

MIDKSAFVHPTAIVEEGASIGANAHIGPFCIVGPHVEIGEGTVLKSHVVVNGHTKIGRDNEIYQFASIGEVNQDLKYAGEPTRVEIGDRNRIRESVTIHRGTVQGGGLTKVGSDNLLMINAHIAHDCTVGNRCILANNATLAGHVSVDDFAIIGGMTAVHQFCIIGAHVMVGGCSGVAQDVPPYVIAQGNHATPFGVNIEGLKRRGFSREAITAIRNAYKLIYRSGKTLDEVKPEIAELAETYPEVKAFTDFFARSTRGLIR

Escherichia coli strain 493/89 UDP-N-acetylglucosamine acetyltransferase (ECs0183) gene, complete cds

**👇 RETO X: Realizá una nueva corrida del BLASTp, utilizando la misma secuencia , pero ahora contra la base de datos PDB. ¿Se obtienen los mismo resultados? ¿Qué tipo de resultados(hits) se recuperan? ¿Cuándo nos podría ser útil este modo de corrida?**

Sigue siendo la Escherichia coli pero cambia el score y el e-value.