

ALINEAMIENTO DE PARES DE SECUENCIAS EJERCICIOS

Ejercicio 1

Las transcriptasas inversas virales, tales como son el producto del gen *pol*, codificado por el virus HIV-1, tienen homólogos humanos. El acceso para la transcriptasa inversa del virus HIV-1 es NP_057849 (usa Entrez para confirmar que es correcto)

Una búsqueda en Entrez revela muchos productos génicos relacionados con virus, incluyendo la polyproteína *Pol*, relacionada con los retrovirus, de 874 aminoácidos (P10266).

Realizad un alineamiento de pares entre ambas secuencias mediante el blastp del NCBI. Aunque su identidad es baja (28%), el alineamiento tiene un e-valor de alrededor de 1×10^{-67} , lo que indica que las proteínas están muy relacionadas.

Por defecto el alineamiento lo hemos realizado con la matriz BLOSUM62. Probad con las matrices BLOSUM45 y BLOSUM80.

¿Cuál es el efecto de cambiar la matriz de puntuación?

¿Puedes explicar de alguna manera los cambios que se observa?

¿Qué matriz crees que es más idónea?

Ejercicio 2

Ahora vamos a realizar el mismo alineamiento del ejercicio 1, pero usando las matrices PAM30 y PAM70

¿Cuáles son los valores E ahora?

¿Qué parte de la secuencia es alineada?

¿Se parecen a los resultados obtenidos utilizando BLOSUM?

Ejercicio 3

Realiza un alineamiento de pares entre RBP y beta-lactoglobulina usando BLAST (del NCBI o de cualquier otro recurso).

Repite el alineamiento usando unas penalizaciones a huecos menores. Construye una tabla indicando los distintos parámetros y los resultados, en término de número de fragmentos alineados, %identidad y E-valores

¿Puedes extraer alguna conclusión?

Ejercicio 4

Compara el ADN mitocondrial del Homo sapiens con el del Neanderthal. Primero, obtén la secuencia de nucleótidos de una de las regiones hipervariables de la mitocondria neanderthal (comienza por la sección de taxonomía del NCBI y selecciona "extinct organisms", para encontrar ADN Neanderthal)

Luego haz un alineamiento por pares con el mismo gene en Homo sapiens y reporta las identidades

Ejercicio 5

En la siguiente figura tienes algunas de las 34 familias de proteínas para las que Dayhoff et al. estimaron, en 1978, sus tasas de mutación (PAM)

Protein	PAMs per 100 million years
Immunoglobulin (Ig) kappa chain C region	37
Kappa casein	33
Epidermal growth factor	26
Serum albumin	19
Hemoglobin alpha chain	12
Myoglobin	8.9
Nerve growth factor	8.5
Trypsin	5.9
Insulin	4.4
Cytochrome <i>c</i>	2.2
Glutamate dehydrogenase	0.9
Histone H3	0.14
Histone H4	0.10

Corroborar mediante alineamientos de pares estos datos, grosso modo.

Por ejemplo, busca en NCBI por citocromo *c* (*cycs*), limitado a humano. En la columna de la derecha tienes un apartado "Homologs to CYCS gene" donde puedes ver sus homólogos en otras especies. Haz un alineamiento de pares de las secuencias de humano y rata, y compáralas con el mismo alineamiento para kappa casein

Ejercicio 6

Vamos a hacer cuentas.

¿Cuál es la puntuación log-odd en la matriz PAM250 para el alineamiento de una cisteína (C) a un glutamato (E)?

Recordad la fórmula de los scores en Dayhoff:

$$S_{ij} = 10 \times \log_{10} \frac{q_{ij}}{p_i}$$

Y las frecuencias normalizadas de aminoácidos

Gly	0.089	Arg	0.041
Ala	0.087	Asn	0.040
Leu	0.085	Phe	0.040
Lys	0.081	Gln	0.038
Ser	0.070	Ile	0.037
Val	0.065	His	0.034
Thr	0.058	Cys	0.033
Pro	0.051	Tyr	0.030
Glu	0.050	Met	0.015
Asp	0.047	Trp	0.010

Ejercicio 7

Las proteínas de enlace “oloroso” (odorant binding proteins) son un ejemplo de lipocalinas.

Obtén los identificadores de dos de estas proteínas para roedores distintos y alínelas. Reportad el porcentaje de identidad y similitud, E-valores, y bit score.

Discute la fiabilidad de la similitud según estos valores y cuáles de ellos te resultan más útiles para evaluar la relación entre dos secuencias.

Ejercicio 8

El celacanto (*Latimeria chalumnae*) es un pez que ha sido bautizado como “fósil viviente”. Se creyó extinto durante mucho tiempo, pero se han descubierto varios especímenes merodeando en el océano.

Contra todo pronóstico, algunos análisis filogenéticos indican que su ADN mitocondrial está más cercano al humano que al de otros peces como los arenques (en inglés, *herring*) (Lewin, 2001)

Vamos a ver si podemos corroborar esto con un análisis sencillo:

1. Descarga ADN mitocondrial (no necesariamente toda la secuencia, puedes elegir algún gen, como citocromo b -cytb-) de humano, celacanto y arenque.
2. Haz alineamiento de pares entre los tres para determinar los grados de identidad y si este pequeño estudio corrobora o no la cercanía entre humano, celacanto y arenque de Lewin, 2001.

Ejercicio 9

La matriz de probabilidades PAM1 no es simétrica. La probabilidad de cambiar Ala a Arg es distinta a la probabilidad de cambiar Arg a Ala
¿Por qué?

Ejercicio 10

¿El hipopótamo está más cerca del cerdo o de la ballena?

Para dar una respuesta parcial a esta pregunta, compara las secuencias de la hemoglobina (subunidad alfa) en estos tres organismos.

¿Puedes encontrar algún artículo que hable sobre el tema en PubMed?