

PRÁCTICA 2: Alineamientos con Clustal Omega

Elvira Mayordomo

21/Feb/2018

1. Introducción

Esta segunda sesión, de aproximadamente dos horas, tratará de la creación y visualización de multialineamientos. Al principio de la sesión dos estudiantes harán una introducción a Clustal Omega y MSAAviewer.

2. Conceptos básicos

2.1. Alineamiento

Como se ha podido observar al explorar GenBank, existen diferencias en la longitud de la misma secuencia de distintas especies. Esto es debido, principalmente, a las mutaciones que han sufrido a lo largo de los años, base de la evolución. Las mutaciones se suelen clasificar en 3 tipos: sustitución, inserción y deleción, siendo más frecuentes las dos últimas (conocidas más comúnmente como *indels*). Por supuesto, nunca se puede descartar que estos errores de longitud sean debidos a un fallo en el proceso de secuenciación, aunque dichos errores son inevitables se asume la buena praxis en esta fase.

Hay muchos estudios biológicos que requieren que el conjunto de secuencias con el que van a trabajar esté alineado, es decir, que todas las secuencias tengan la misma longitud. Para ello, se introduce en el alfabeto de la secuencia un nuevo elemento: el gap (-). Este símbolo se interpreta como posiciones que deberían existir y han sido eliminadas por una mutación, o como ajuste

por inserciones en otras secuencias. El proceso de obtener el alineamiento óptimo tiene un coste asociado de $O(m^n)$, donde m es la longitud de las secuencias y n el número total de secuencias. Este coste es extremadamente elevado, incluso en los casos donde m y n son pequeños (dentro de los parámetros de datos reales biológicos). Por ello, las herramientas y algoritmos de alineamiento aplican heurísticas para resolver este problema.

La siguiente tarea por parte del alumno será utilizar una herramienta muy extendida para generar alineamientos, Clustal Omega (<http://clustal.org/omega/>), aplicando distintas configuraciones a conjuntos de las secuencias descargadas en la práctica anterior. La idea es familiarizarse con este proceso y evaluar cómo los costes temporales varían según cambian los tamaños de entrada, así como las variaciones en los alineamientos resultantes según la configuración y herramienta utilizada. En la siguiente sesión introduciremos la herramienta de multialineamiento alternativa mafft.

Una vez hemos obtenido el alineamiento correspondiente, es recomendable realizar una revisión para confirmar que todo está correcto y el proceso ha finalizado satisfactoriamente. Vamos a utilizar una herramienta web para poder visualizar el alineamiento. Para ello, iremos a la web <http://msa.biojs.net/app/> y haremos **Import** → **From file (Fasta, ...)** y cargaremos el fichero con nuestro alineamiento. Tras unos segundos podremos observar el aspecto que tiene nuestro alineamiento. La segunda tarea del alumno será familiarizarse con esta herramienta y revisar el alineamiento que ha generado.

Tareas a realizar:

1. Escoger 10 secuencias de cada uno de los paquetes descargados en la práctica anterior y alinearlas usando Clustal Omega con distintas configuraciones.
2. Estudiar algunas de las diferencias obtenidas en los alineamientos anteriores.
3. Descargar un pequeño conjunto de secuencias (no más de 10) de proteínas y repetir los puntos 1 y 2.
4. Buscar alternativas de alineamiento para mejorar la eficiencia en tiempo y/o memoria de los procesos anteriores.

5. A partir de las secuencias del siguiente artículo ya recuperadas en la práctica 1, intentar hacer un alineamiento de todas ellas (son 174) o bien de una parte. Esta tarea se completará en siguientes sesiones utilizando la herramienta mafft.

Metzker M.L., Mindell D.P., Liu X.M., et al. (2002). Molecular evidence of HIV-1 transmission in a criminal case. *Proc Natl Acad Sci U S A*. 99(22):14292-7.

2.2. Entrega

Debe entregarse una breve memoria (formato preferido PDF) sobre esta práctica incluyendo los comandos de Clustal Omega utilizados así como los ficheros FASTA necesarios y las visualizaciones de los alineamientos obtenidos (todo ello incluido dentro de un zip). Debe enviarse por correo electrónico a elvira@unizar.es antes de la siguiente sesión (7 de marzo).