

ALINEAMIENTO DE SECUENCIAS USANDO CLUSTALX: UNA INTRODUCCIÓN

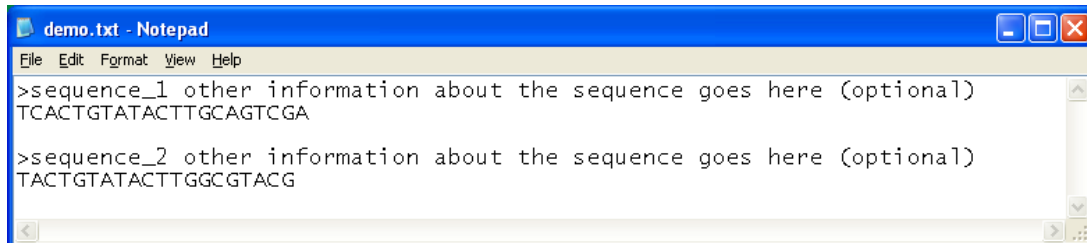
Este documento puede usarse para introducir a los estudiantes a los conceptos básicos de alineamiento de secuencias. El alineamiento es necesario antes de que las secuencias de ADN puedan ser comparadas y utilizadas en análisis filogenéticos.

REPRESENTACIÓN DE SECUENCIAS DE ADN

Hay diferentes formatos comúnmente utilizados para representar secuencias de ADN. A continuación se muestra la secuencia parcial del gen de la subunidad I de la *citocromo oxidasa* del perro (*COI*) en formato FASTA. El formato FASTA empieza con un ">," seguido de la información acerca de la secuencia en la primera línea, seguido de la secuencia de ADN.

```
>gi|377685879|gb|JN850779.1| Canis lupus familiaris isolate dog_3 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial"
TACTTTATACTTACTATTTGGAGCATGAGCCGGTATAGTAGGCACTGCCTTGAGCCTCCTCATCCGAGCC
GAACTAGGTCAGCCCGGTACTTTACTAGGTGACGATCAAATTTATAATGTCATYGTAAACGCCCATGCTT...
```

Un archivo con secuencias en formato FASTA puede contener múltiples secuencias, una después de la otra. Por ejemplo:



```
demo.txt - Notepad
File Edit Format View Help
>sequence_1 other information about the sequence goes here (optional)
TCACTGTATACTTGCGATCGA
>sequence_2 other information about the sequence goes here (optional)
TACTGTATACTTGGCGTACG
```

¿QUÉ SECUENCIAS SE ESCOGEN PARA COMPARAR?

En taxonomía moderna los científicos rutinariamente analizan el ADN de las especies que recolectan para obtener un "código de barras de ADN" que es una secuencia única para cada especie, y se usa para identificar a dicha especie. Para animales y muchos otros eucariontes, diferentes genes han sido usados con este propósito. Un ejemplo es el gen mitocondrial de la subunidad I de la citocromo oxidasa (*COI*), el cual codifica parte de una enzima que es importante para la respiración celular, y otro es la NADH deshidrogenasa subunidad 2 (*ND2*). Secuencias como éstas están disponibles para un rango amplio de especies, haciendo posible su uso para explorar relaciones filogenéticas entre especies.

COI o *ND2* son útiles como "códigos de barras de ADN", porque en general muestran poca variación entre los miembros de la misma especie, y al mismo tiempo la variación es suficiente como para distinguir a miembros de especies diferentes. Por lo tanto, estas secuencias proporcionan un "sello" único para cada especie y son adecuadas para determinar relaciones filogenéticas.

Debido a que estas secuencias son muy similares dentro de la misma especie, estos genes no son adecuados para estudiar variaciones intra-específicas, o diferencias entre especies que se han separado muy recientemente. Las secuencias *COI* también tienen índices bajos de mutación en muchas especies de plantas, y no son generalmente usadas para hacer códigos de barras o para comparaciones filogenéticas entre estas especies.

La secuencia usada en la actividad *Usando secuencias de ADN para construir árboles filogenéticos* incluye el gen de la subunidad 2 de la NADH deshidrogenasa junto con secuencias adyacentes de varios genes de ARN de transferencia. El gen ND2 es comúnmente utilizado para perfiles genéticos en animales. Es apropiado para este propósito porque se conserva lo suficiente como para estar presente en un diverso grupo de animales, y además es suficientemente variable como para examinar relaciones evolutivas.

¿QUÉ PROGRAMA USAR?

Para enseñar a los estudiantes acerca de los principios de alineamiento de secuencias, ClustalX es una buena opción, si es que los estudiantes pueden instalar el software en su computadora. Para generar filogenias, el sitio www.phylogeny.fr es la herramienta más fácil de usar.

- ClustalX es intuitivo, y a su vez es una herramienta excelente para ilustrar el concepto y el proceso del alineamiento de secuencias. ClustalX es gratuito y se puede instalar en el disco duro, lo que es una ventaja (no depende del internet) y también una desventaja (requiere instalación del programa) para su uso en el salón de clases. El algoritmo es un poco anticuado, hay otros programas que son mejores para generar filogenias; sin embargo ClustalX es apropiado para demostrar cómo generar árboles filogenéticos en base a secuencias de ADN. La filogenia generada requiere de otro programa que también está disponible de forma gratuita, NJplot, para imprimir o para ver los árboles.
- www.phylogeny.fr es una herramienta en internet para generar filogenias. Usando los parámetros estándar, “phylogeny.fr” es simple de usar, y utiliza un generador de alineamientos llamado MUSCLE. El sitio genera una filogenia que puede ser guardada en diferentes formatos. Sin embargo, la forma en que se muestra el alineamiento de las secuencias no es tan intuitivo como lo es en ClustalX.

INSTRUCCIONES PARA EL ALINEAMIENTO Y LA GENERACIÓN DE ÁRBOLES FILOGENÉTICOS USANDO CLUSTALX

Software y Archivos

Instale ClustalX, el cual está disponible en el sitio <http://www.clustal.org/clustal2/>. (Para Windows, baje clustalx-2.1-win.msi; para Mac OS, baje clustalx-2.1-macosx.dmg.). Luego instale NJplot, el cual está disponible en <http://pbil.univ-lyon1.fr/software/njplot.html>.

Cómo interpretar las secuencias alineadas

Utilice ClustalX para comparar las secuencias de ADN. Para este ejercicio, use las secuencias de prueba del archivo “test.txt”, el cual contiene secuencias cortas de ADN (test1, test2, y test3) como se muestra a continuación:

```

">test1
AAGGAAGGAAGGAAGGAAGGAAGG
>test2
AAGGAAGGAATGGAAGGAAGGAAGG
>test3
AAGGAACGGAATGGTAGGAAGGAAGG"
    
```



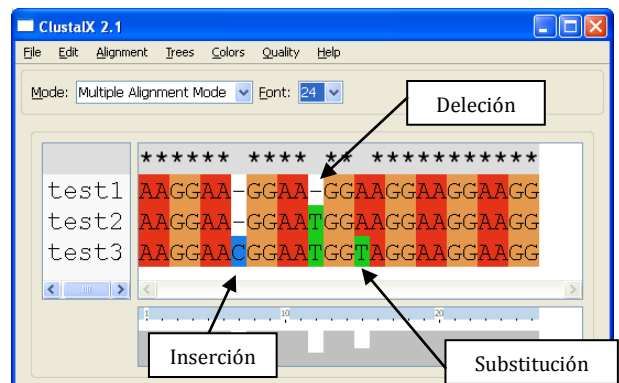
Cargue las secuencias en ClustalX usando el menú, "File -> Load Sequences", y seleccionando "test.txt". ClustalX le mostrará las secuencias como se muestra a continuación (en formato PC).

Antes de poder comparar las secuencias, usted debe de "alinearlas", esto significa organizar las secuencias y deslizar las bases hasta encontrar el patrón que maximice la similitud entre todas ellas. Este proceso de alineamiento permite examinar diferencias que reflejan relaciones evolutivas.



En el menú, escoja "Alignment -> Do Complete Alignment". Cuando se le pida que dé nombres a los archivos que se generarán en el análisis, use los nombres sugeridos por el programa y oprima "OK." La pantalla cambiará y se verá similar a la ilustración de la derecha.

Note que es mucho más fácil ver las diferencias entre las secuencias de ADN una vez que están alineadas. Se pueden ver los tipos de mutaciones que han ocurrido en cada secuencia (como se muestra en la ilustración). El número de diferencias entre las secuencias determinan qué tan cercanos o lejanos son los organismos que se están comparando.



Basados solamente en esta información ¿Cuáles son las dos secuencias que usted cree están más cercanamente relacionadas entre sí? Para ver si su respuesta es la correcta podemos usar ClustalX para generar un árbol filogenético.

En el menú, escoja “Trees -> Draw Tree”. Esto crea un archivo con el árbol filogenético llamado “test1.ph”, el cual se puede abrir usando “Njplot.exe.” Abra Njplot y escoja “File -> Open”, y seleccione “test1.ph.”

El resultado muestra que “test1 y test2” están en la misma rama del árbol, indicando que están más cercanamente relacionadas entre sí que con “test3.”

