

Alineamiento

Datos Moleculares en Biodiversidad II

Nelson R. Salinas

Septiembre 12, 2020

Repositorio sesión:

https://github.com/nrsalinas/datos_moleculares_ii/tree/master/Taller_4.

1. Instale un editor de texto aceptable, dado el caso que aún no tenga instalado uno. Un editor de texto eficiente es una herramienta básica de cualquier bioinformático y debe cumplir al menos dos tareas básicas: 1) modificar la codificación de los archivos y 2) cambiar el caracter de final de línea (*EOL character*). Los editores de texto que Windows ofrece por defecto (e.g., Microsoft Word o Notepad) **no son** editores de texto aceptables. Algunas opciones sugeridas son [Visual Studio Code](#) (editor oficial de Microsoft), [Atom](#), [Brackets](#) y [Jedit](#) (necesita máquina virtual de Java).
2. Instale los siguientes programas de alineamiento:
 - (a) [Mafft](#).
 - (b) [Clustal W](#).
3. Instale un programa de visualización de alineamiento. Algunas opciones son [SeaView](#), [Mega](#) o [Ugene](#).
4. Descargue los archivos `vaccinieae_its.fasta` y `vaccinieae_matk.fasta` del repositorio de la clase. Abra el archivo con el editor de texto instalado en el punto 1. ¿A qué formato corresponden los archivos? ¿Son alineamientos? ¿Por qué?
5. Si concluye que los archivos no están alineados, intente alinearlos con los programas Mafft y Clustal W. La versión en Linux de ambos programas se ejecuta a través de la terminal del sistema operativa (a veces también llamada línea de comando).
 - (a) Si es la primera vez que ejecuta un programa a través de la terminal, le será de utilidad desplegar la ayuda del programa, lo cual se realiza ejecutando el nombre del programa y luego la opción `help`:

```
mafft --help
```

```
clustalw -help
```
 - (b) A través de la ayuda desplegada, identifique las opciones que le permiten direccionar los datos de entrada y salida al programa: el archivo que va a ser procesado se señala con la opción de *input file*, mientras que el nombre para el nuevo archivo alineado se configura con la opción de *output file*.

- (c) Realice los alineamientos utilizando ambos programas bajo los parámetros por defecto; luego visualice los resultados. ¿Observa diferencias en los alineamientos producidos por los dos programas? Si su respuesta es afirmativa, ¿puede decidir cuál alineamiento es mejor?
 - (d) A través de la información de ayuda de los programas, identifique las opciones que cambian los valores de la penalidad de abertura y extensión de gaps (*gap opening* y *gap extension*).
 - (e) Alinee nuevamente los archivos empleando varias combinaciones en los valores de creación (0, 1, 3 y 6 para ambos programas) y extensión de gaps (5, 20, 40 para ClustalW; 0, 5 y 10 para Mafft).
 - (f) Compare los alineamientos producidos por el paso anterior. Tabule los resultados obtenidos digitando varias de las propiedades básicas de un alineamiento: número total de columnas, número de columnas con gaps, número de sitios variables, número de sitios informativos. Todas estas propiedades usualmente la calcula el programa de visualización de alineamientos (e.g., en Seaview se pueden ver en la opción “Statistics” del menú “Props”).
 - (g) Algunas veces los alineamientos comienzan y/o terminan con sitios dominados por gaps. ¿Es el caso de las matrices recién alineadas? ¿Cuál cree que es la causa? ¿Dicho fenómeno puede tener impacto en los análisis posteriores?
 - (h) Después de haber visto todos los alineamientos del punto anterior y de haber extraído las propiedades cuantitativas de los mismos, ¿cuál cree que es la característica más apropiada para decidir qué alineamiento es mejor? ¿Existen diferentes efectos de los parámetros en los dos conjuntos de datos? ¿Cuál cree que es la razón para todas esas diferencias, si hubieren?
6. Compile las secuencias de su grupo de estudio en un archivo fasta y trate de alinearlo con el programa de su preferencia, empleando varios valores para los parámetros de creación y extensión de gaps. ¿Existen diferencias entre los alineamientos producidos?
 7. Después de ver los alineamientos de su grupo de estudio, ¿observó que en algunos sitios los programas no hicieron un buen trabajo? ¿Cree que puede realizar ajustes manuales (alineación “al ojo”)?
 8. Algunos programas de alineamiento permiten guardar el archivo de salida en formatos distintos a fasta. Pruebe alinear los datos de su grupo de estudio en ClustalW salvando el resultado en formato Phylip. Para este fin debe consultar la ayuda del programa de alineamiento y seleccionar el valor apropiado en la opción `output`.
 9. Algunos formatos de matrices solo son empleados dentro del campo de la filogenética, por lo que los programas de alineamiento usualmente no los incluyen como una opción para el archivo de salida. Un ejemplo es el formato Nexus. Afortunadamente, existe una multitud de programas para hacer conversiones de formato. Desafortunadamente, la mayoría de ellos pueden producir resultados defectuosos, por lo cual lo más recomendable es el conocimiento detallado de los formatos de matrices filogenéticas. Para este taller se realizará un ejercicio simple y se convertirá el mejor alineamiento del punto 6 a formato Nexus usando el servicio web [ALTER](#).