

Herramientas para alineamientos:

CLUSTAL-OMEGA

MSA-VIEWER

Alejandro Francés - 717361

Víctor Labrador - 701658

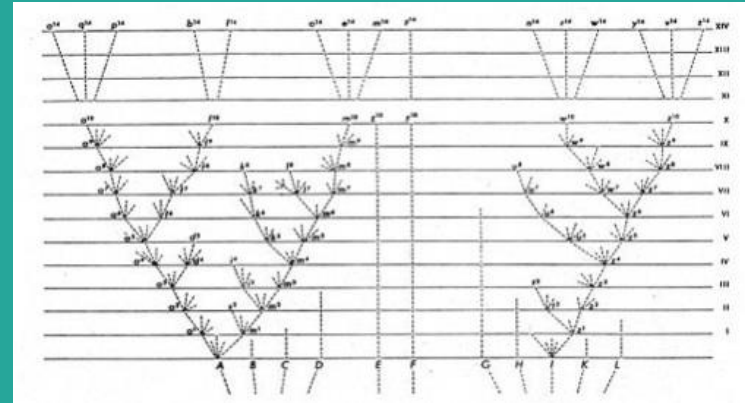
Mengdie Zhou - 700991

CLUSTAL-OMEGA

- ★ Programa de propósito general para el alineamiento secuencial de proteínas y DNA/RNA
- ★ Produce secuencias de alineamiento múltiples y es capaz de manejar paquetes de datos de millones de secuencias en un tiempo razonable
- ★ Se trabaja mediante línea de comando, indicando los parámetros necesarios
- ★ Existe también otra versión con interfaz gráfica

CLUSTAL-OMEGA

- ★ Como resultado, devuelve las sentencias alineadas, según parámetros indicados.
- ★ Tres etapas:
 - Alineamiento por pares
 - Creación de árbol filogenético
 - Alineamiento múltiple



Instalación

Compilador g++ requerida

Descargar los ficheros en clustal-omega requerida

En caso de no tener argtable2, instalarlo

Cuando ya se tenga todos los ficheros, es posible ejecutar los ficheros de auto-configuración que vienen en el paquete

Instalación

Al ejecutar el programa, es posible que haya un error, se trata de que no encuentra el objeto libargtable2.so.0

Para solucionarlo: incluir el path de dicho objeto

Para hacer permanente la variable de entorno, modificarlo en el fichero /etc/environment. Otra opción es indicarle al intérprete shell que existe dicha variable.

Instalación

Para solucionarlo:

```
LD_LIBRARY_PATH=/usr/local/lib
```

```
export LD_LIBRARY_PATH
```

O de forma permanente, modificando el fichero

```
~/.bashrc
```

Configuración

Mediante líneas de comando

clustalo {-opciones}

Las opciones son:

-i, --in, --infile={<file>, -}

Añadir fichero

--hmm-in=<file>
Márkov)

Añadir fichero HHM(modelo oculto de

--dealign

Sirve para asignar secuencias de entrada

Configuración

`--profile1, --p1=<file>`

`--profile2, --p2=<file>`

`--is-profile`

`-t, --seqtype={Protein, RNA, DNA}`

Forzar un tipo de secuencia

`--infmt={a2m=[formato]}` | formato puede ser:
file

Igual que t pero con un

fasta, clustal, msf, phylip, selex, stockholm, vienna

Configuración

- `--distmat-in=<file>` Matriz de distancia por pares (in)
- `--distmat-out=<file>` Para output file
- `--guidetree-in=<file>`
- `--guidetree-out=<file>`
- `--full` Usa la matriz de distancia completa

Configuración

--full-iter

--cluster-size=<n>

--clustering-out=<file>

--use-kimura

--percent-id

Configuración

-o, --out, --outfile=<file>
(fasta)

El fichero de salida

--outfmt={a2m=[formato]}
MSA

Fichero de salida

--residuenumber, --resno

--wrap=<n>

--output-order={input-order, tree-order}

Configuración

--iterations, --iter=<n>
árbol guiado y HMM) iteraciones

Nº de (combinaciones de

--max-guidetree-iterations=<n>

--max-hmm-iterations=<n>

--maxnumseq=<n>

Nº max de secuencias

--maxseqlen=<l>

secuencia

Nº max de longitud de

Configuración

--auto

--threads=<n>

-l, --log=<file>

-h, -v, --version, --long-version

--force

Ejemplo 1

```
./clustalo -i fichero.fa
```

Lee la secuencia en el “fichero.fa” y alinea las secuencias y muestra el resultado por la pantalla en formato FASTA/A2M

Ejemplo 2

```
./clustalo -i ficheroIn.fa -o ficheroOut --outfmt=fasta --force
```

Lee la secuencia del fichero ficheroIn.fa, alinea las secuencias y escribe el resultado en ficheroOut con el formato FASTA. Si el fichero ficheroOut.fasta ya existe, con el “force” lo sobrescribe.

Ejemplo 3

```
./clustalo -i fileIn.fasta --iter=2
```

Lee el fichero fileIn.fasta y crea un árbol construido a partir de distancias k-tupla y realiza una alineación inicial. Luego se convierte a HMM (modelo de Markov) y se vuelve a construir un árbol guiado. Las secuencias no alineadas se alinean usando información del HMM.

Ejemplo 3 (continuación)

```
./clustalo -i fileIn.fasta --iter=2
```

El segundo alineamiento se vuelve a convertir en HMM y se vuelve a construir un árbol guiado. Entonces las secuencias no alineadas se alinean (por tercera vez) usando información del HMM del paso anterior y el último árbol guiado creado. El resultado del último alineamiento se muestra por pantalla.

Ejemplo 4

```
./clustalo -i fileIn.fasta -o fileOut.fasta -v
```

Si el fichero fileOut no existe (si existe no hace nada porque no está la opción de --force) lee la información de fileIn y muestra el progreso en la pantalla y escribe el resultado en fileOut.fasta.

Ejemplo 5

Para alinear las secuencias, será necesario indicar, por lo tanto, el fichero contenido de las secuencias y el formato del fichero, un ejemplo con las secuencias del crimen:

```
clustalo --infile=crimen.fasta --infmt=a2m=fasta
```

Ejemplo 5 resultado

Devuelve el resultado por la salida, aunque es posible especificar un fichero de salida con la opción previamente nombrada:

```
clustalo --infile=crimen.fasta --infmt=a2m=fasta --  
outfile=crimen_out.fasta
```

Ejemplo 5 resultado

Una de las sentencias, antes de ser alineado:

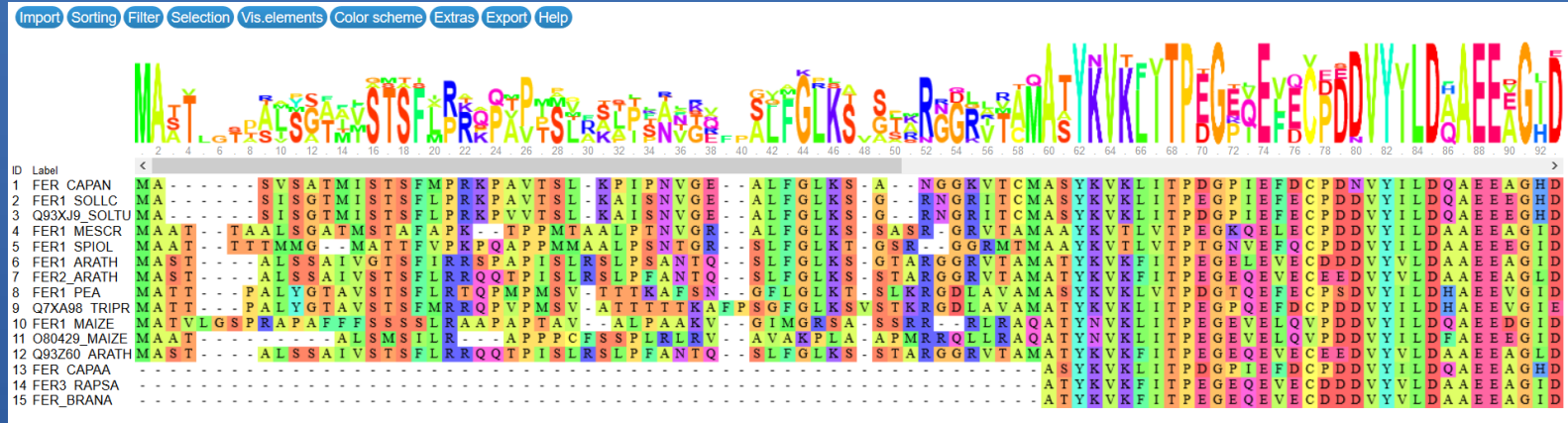
```
>AY156907.1 HIV-1 clone V50 from USA envelope glycoprotein (env) gene, partial cds
AATTAGGCCAACAGTATCAACTCGATTACTGTAAATGGCAGTCTAGCAGAAAAAGAGGTAGTAATTAGA
TCTGAAAATTTACGAGCAATGCTAAAACCATAATAGTACAGCTGACAACACCTGTAGAAATTAATTGTA
TAAGACCCGACAACAATACAAGAAAAAGTATACCTATAGGACCAGGGAGAGCACTTTATACAACAGGAGA
CATAATAGGAAATATAAGAAAAGCATATTGTAACATTAGTAAAGCAAAATGGAATGACACTCTAAGCCAG
GTAGTTAAAAAATTAGGACAACAATTTAATAAAACAACAATAGTCTTTAAGCAATCCTCAGGAGGGGACC
CAGAAATTGCAATGCACAGTTTTAACTGTGGAGGGGAATTTTTCTACTGTAATACAACACAACACTGTTTAA
TAGTACCTGGGATATTACTGAAGGAACAAATATTACTGAAGGAACAAATAACATGGAAGGAAATAATACA
ACAATCATACTCCCATGCAGAATAAAACAAATGATAAACATGTGGCAGGAGGTAGGAAAAGCAATGTATG
CCCCGCCCATCCAAGGAATAATTAGCTGCTCATCAAATATTACAGGGCTACTACTAACAAGAGATGGCGG
TAATGGTAATGACACCAGAGATAATGAGACCTTCAGACCTGGAGGAGGAGATATGAAGGACAATTGGAGA
AGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAA
GGGTGGTGCAGAGAGAAAAAAGA
```

Ejemplo 5 resultado

Una de las sentencias, después de ser alineado:

```
>AY156907.1 HIV-1 clone V50 from USA envelope glycoprotein (env) gene, partial cds
-----AATTAGGCCAACAGTATCAAC
TCGATTACTGTAAATGGCAGTCTAGCAGAAAAAGAGGTAGTAATTAGATCTGAAAATTT
CACGAGCAATG---CTAAAACCATAATAGTACAGCTGACAACACCTGTAG-----AAATT
AATTGTATAAGACCCGACAACA-----ATACAAGAAAAAGTATACCTATAGGACC-
--AGGGAGAG-----CACTTTATACAACAGGAGACATAATAGGAAATATAAGAAAAGCA
TATTGTAACATTAGTAAAGCAAAATGGAATGACACTCTAA----GCCAGGTAGTTAAAAA
ATTAGGACAACAATTTAA-----TAAACAACAATAGTCTTTAAGCAATCCTCAGGAG
GGGACCCAGAAATTGCAATGCACAGTTTTAACTGTGGAGGGGAA---TTTTTCTACTGTA
ATACAACACAACACTGTTTAATAGTACCTGGGATATTACTGAAGGAACAAATATTACTGAAG
GAACAAATAACATGGAAGGAAATAATACAACAATCATACTCCCATGCAGAATAAAACAAA
TGATAAACATGTGGCA-----GGAGGTAGGAAAAGCAATGTATGCCCCGCCCATCCA
AGGAATAATTAGCTGCTCATCAAATATTACAGGGCTACTACTAACAAGAGATGGCGGTAA
TGGTAATGA-----CACCAGAGATAATGAGACCTTCAGACCTGGAGGAGGAGATAT
GAAGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGG
AGTAGCACCCACCAAGGCAAAGAGAAGGGTGGTGCAGAGAGAAAAAAGA-----
-----
-----
```

MSAViewer



<http://msa.biojs.net/app/>

Herramienta web que permite visualizar de forma fácil las alineaciones de secuencias múltiples

Características



RÁPIDA

Uso de memoria bajo y alto rendimiento, lo que permite mostrar cualquier secuencia de forma rápida

Características



FÁCIL

Permite cargar archivos de formato FASTA o CLUSTAL

Permite personalizar el esquema de color con el que se muestran las secuencias

Características



JAVASCRIPT PURO

Esta implementado usando javascript puro por lo que puede utilizarse en cualquier navegador sin ser necesario su instalación

Características



CÓDIGO LIBRE

Herramienta de código libre, lo que permite incorporarlo a cualquier pagina web

Importar

Permite cargar archivos de entrada desde una URL, una unidad de disco o arrastrando y soltando

Soporta los siguientes formatos:

- MSA en formatos FASTA o CLUSTAL.
- Secuencia de anotación en formato GFFv3 o JalView.
- Árbol filogenético en formato newick.

Ordenación

Permite ordenar las secuencias según los siguientes criterios:

- Identificador
- Etiqueta
- Secuencia
- Identidad
- Porcentaje de huecos

Filtrado y selección

Permite ocultar secuencias y / o columnas por un cierto porcentaje de identidad de secuencia, por cierto porcentaje de huecos o por selección.

Permite buscar una alineación para un motivo utilizando una expresión regular (por ejemplo, K (K | R) RK para una señal de localización nuclear) o invertir la selección de columnas y / o filas.

Otras opciones

Permite adaptar la visualización mostrando u ocultando elementos informativos.

Permite cambiar la paleta de colores.

Permite exportar la visualización de la secuencia en estos formatos:

- URL
- Jalview
- FASTA
- GFF
- PNG

Permite saltar a una columna en concreto.