

Alineamientos y Herramientas web

Alineamientos de secuencias

- Forma de representar y comparar **dos o más secuencias** o cadenas de ADN, ARN, o estructuras primarias proteicas para resaltar sus zonas de similitud, que podrían indicar relaciones funcionales o evolutivas entre los genes o proteínas consultados.
- Se trata de la herramienta más básica y potente de la bioinformática.

Alineamientos de secuencias

El alineamiento busca mostrar los residuos derivados de un ancestro común, y los representa gráficamente unos superpuestos a otros.

Seq Ancestral: *MKQPGGLMDK*

- Var_Seq1: MRQPGLMDK
- Var_Seq2: MKQPGGLMDK

Alineamiento...

MRQPGLMDK
MKQPGGLMDK

Alineamiento de 2 secuencias

```
AAB24882      TYHMCQFHCRYVNNHSGEKLYECNERSKAFSCPSHLQCHKRRQIGEKTHEHNQCGKAFPT 60
AAB24881      -----YECNQC GKAF AQHSSLKCHYRTHIGEKPYECNQC GKAFSK 40
                ****: .***: * *:** * :****.:* *****..

AAB24882      PSHLQYHERTHTGKPYECHQCGQAFKKCSLLQRHKRTHHTGKPYE-CNQC GKAF AQ- 116
AAB24881      HSHLQCHKRTHHTGKPYECNQC GKAF SQHGLLQRHKRTHHTGKPYMNVINMVKPLHNS 98
                **** *:*****:***:**.: .*****          : *.: :
```

Alineamientos de secuencias

-Locales y Globales-

Globales: Intentan alinear cada residuo de cada secuencia. Más útiles cuando las secuencias problema iniciales son similares y aproximadamente del mismo tamaño.

Locales: Intentan alinear regiones concretas, de las que conocemos su similitud de antemano. (*Dominios, centros activos...*)

```
Global FTFTALILLAVAV
      F--TAL-LLA-AV
```

```
Local FTFTALILL-AVAV
     --FTAL-LLAAV--
```

Puntuación de los alineamientos

- (*matrices de sustitucion*) -

- Reemplazo de residuos "*similares*" ocurre más a menudo:
 - *Ser --> Thr, Leu --> Ile*

- Similares...¿pero en qué?:

Hidrofobicidad

Tamaño

Carga

Propiedades 3D

Código genético

...

| | C | S | T | P | A | G | N | D | E | Q | H | R | K | M | I | L | V | F | Y | W |
|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|---|---|----|
| C | 9 | | | | | | | | | | | | | | | | | | | |
| S | -1 | 4 | | | | | | | | | | | | | | | | | | |
| T | -1 | 1 | 5 | | | | | | | | | | | | | | | | | |
| P | -3 | -1 | -1 | 7 | | | | | | | | | | | | | | | | |
| A | 0 | 1 | 0 | -1 | 4 | | | | | | | | | | | | | | | |
| G | -3 | 0 | -2 | -2 | 0 | 6 | | | | | | | | | | | | | | |
| N | -3 | 1 | 0 | -2 | -2 | 0 | 6 | | | | | | | | | | | | | |
| D | -3 | 0 | -1 | -1 | -2 | -1 | 1 | 6 | | | | | | | | | | | | |
| E | -4 | 0 | -1 | -1 | -1 | -2 | 0 | 2 | 5 | | | | | | | | | | | |
| Q | -3 | 0 | -1 | -1 | -1 | -2 | 0 | 0 | 2 | 5 | | | | | | | | | | |
| H | -3 | -1 | -2 | -2 | -2 | -2 | 1 | -1 | 0 | 0 | 8 | | | | | | | | | |
| R | -3 | -1 | -1 | -2 | -1 | -2 | 0 | -2 | 0 | 1 | 0 | 5 | | | | | | | | |
| K | -3 | 0 | -1 | -1 | -1 | -2 | 0 | -1 | 1 | 1 | -1 | 2 | 5 | | | | | | | |
| M | -1 | -1 | -1 | -2 | -1 | -3 | -2 | -3 | -2 | 0 | -2 | -1 | -1 | 5 | | | | | | |
| I | -1 | -2 | -1 | -3 | -1 | -4 | -3 | -3 | -3 | -3 | -3 | -3 | -3 | 1 | 4 | | | | | |
| L | -1 | -2 | -1 | -3 | -1 | -4 | -3 | -4 | -3 | -2 | -3 | -2 | -2 | 2 | 2 | 4 | | | | |
| V | -1 | -2 | 0 | -2 | 0 | -3 | -3 | -3 | -2 | -2 | -3 | -3 | -2 | 1 | 3 | 1 | 4 | | | |
| F | -2 | -2 | -2 | -4 | -2 | -3 | -3 | -3 | -3 | -3 | -1 | -3 | -3 | 0 | 0 | 0 | -1 | 6 | | |
| Y | -2 | -2 | -2 | -3 | -2 | -3 | -2 | -3 | -2 | -1 | 2 | -2 | -2 | -1 | -1 | -1 | -1 | 3 | 7 | |
| W | -2 | -3 | -2 | -4 | -3 | -2 | -4 | -4 | -3 | -2 | -2 | -3 | -3 | -1 | -3 | -2 | -3 | 1 | 2 | 11 |

- Se computa siguiendo las indicaciones de las matrices de sustitución

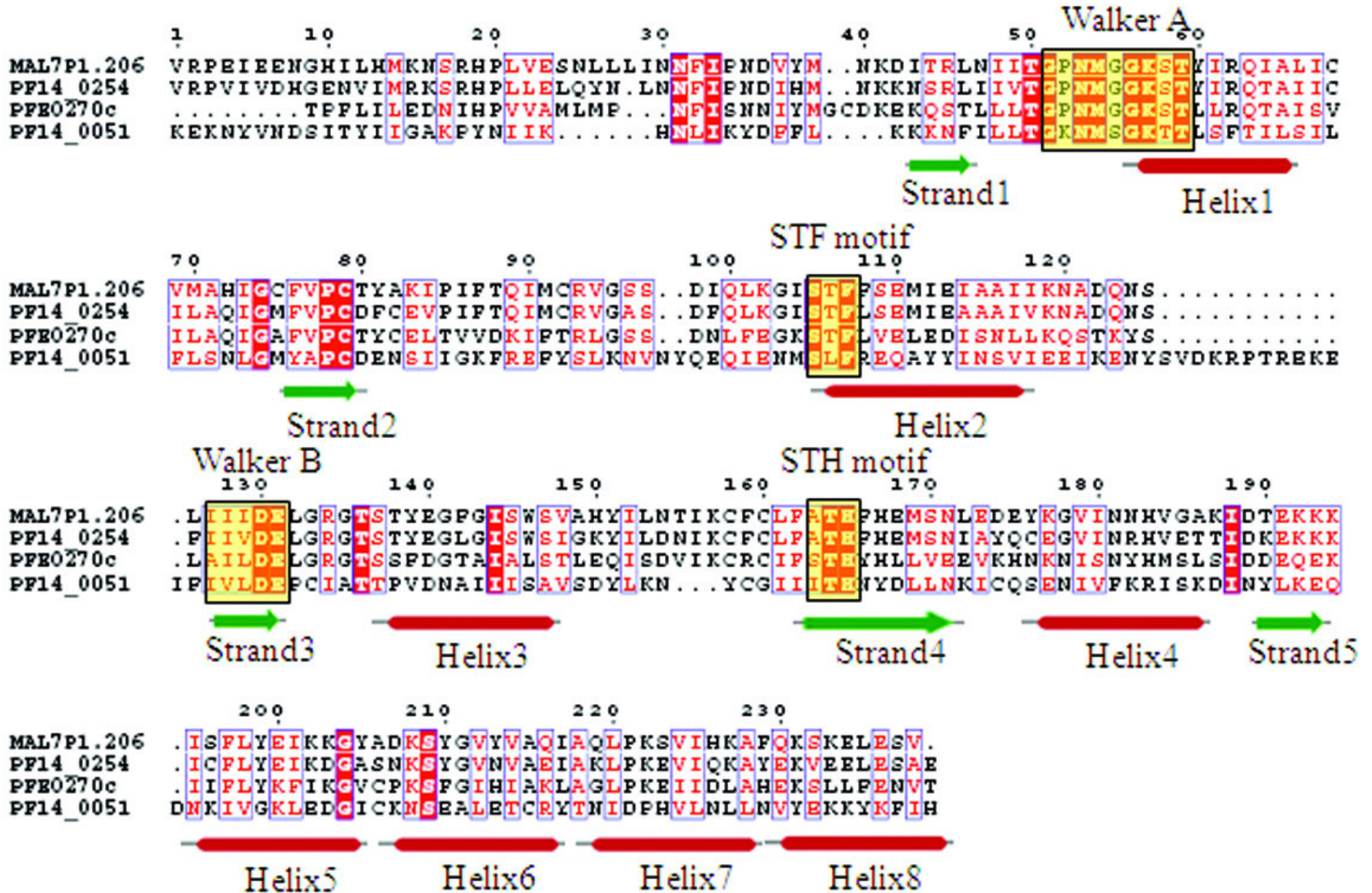
Alineamientos múltiples de secuencias

- Un ***alineamiento múltiple*** es un alineamiento de más de dos secuencias, siguiendo los mismos principios que el alineamiento para dos secuencias.

- Los alineamientos múltiples, son esenciales para la reconstrucción filogenética.

Se asume que todas las secuencias provienen de un ancestro común, y trataremos de buscar las relaciones entre ellas.

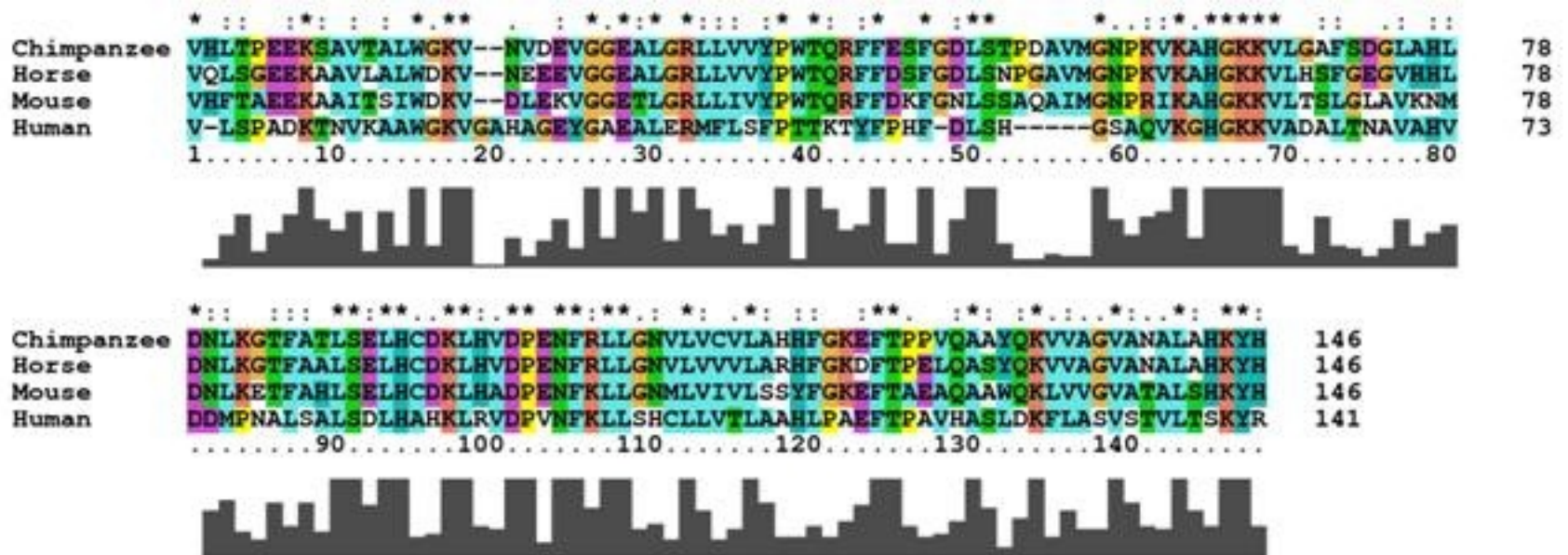
Alineamientos múltiples de secuencias



Herramientas para alineamientos

ClustalW (*ClustalX*)

- Alineamiento progresivo: *Alineando primero secuencias más similares, para ir añadiendo sucesivamente al alineamiento secuencias o grupos menos relacionados*
- Diferentes matrices de puntuación para diferentes alineamientos.
- Penalizaciones por *gap* (*huecos en los alins.*); varían entre secuencias y de manera específica según la posición



ClustalW (*ClustalX*)

El alineamiento *progresivo* es rápido y simple, pero tiene un problema... :

-Cualquier error de alineamiento que ocurra en los primeros alineamientos, no se podrá corregir posteriormente.

-SOLUCIONES-

· **T-Coffee**: Realiza distintos alineamientos apareados (*entre 2 secuencias usando distintos algoritmos, global/local*), detecta los **residuos más consistentes**, y usa esta información para el alineamiento final.

- MUY LENTO
- MUY BUENA CALIDAD

--> **Alternativas**: Alineamientos iterativos: **Muscle**, **MAFFT**

MUSCLE

Multiple sequence alignment

Faster and more accurate than CLUSTALW

- "Más rápido y preciso que CLUSTALW..."

- Construye pares de perfiles que se extraen del alineamiento, se realinean, y se guarda el resultado sólo cuando ha mejorado.

MAFFT

- Se basa en el mismo mecanismo de MUSCLE
- Similar en rapidez y precisión a MUSCLE

M-Coffee

- Combina varios alineadores.
- Después usa T-Coffee para combinar estos alineamientos, en lugar de elegir un unico método de alineamiento.
- Puntuación para los alineamientos en función del color



*Paquete de herramientas para
análisis de evolución molecular,
filogenética y filogenómica.*

Phylemon

- Antes de Phylemon:

Uso de un número elevado de programas para inferencia filogenética (PHYLIP, PAUP, MEGA, PhyML, PAML & MrBayes)

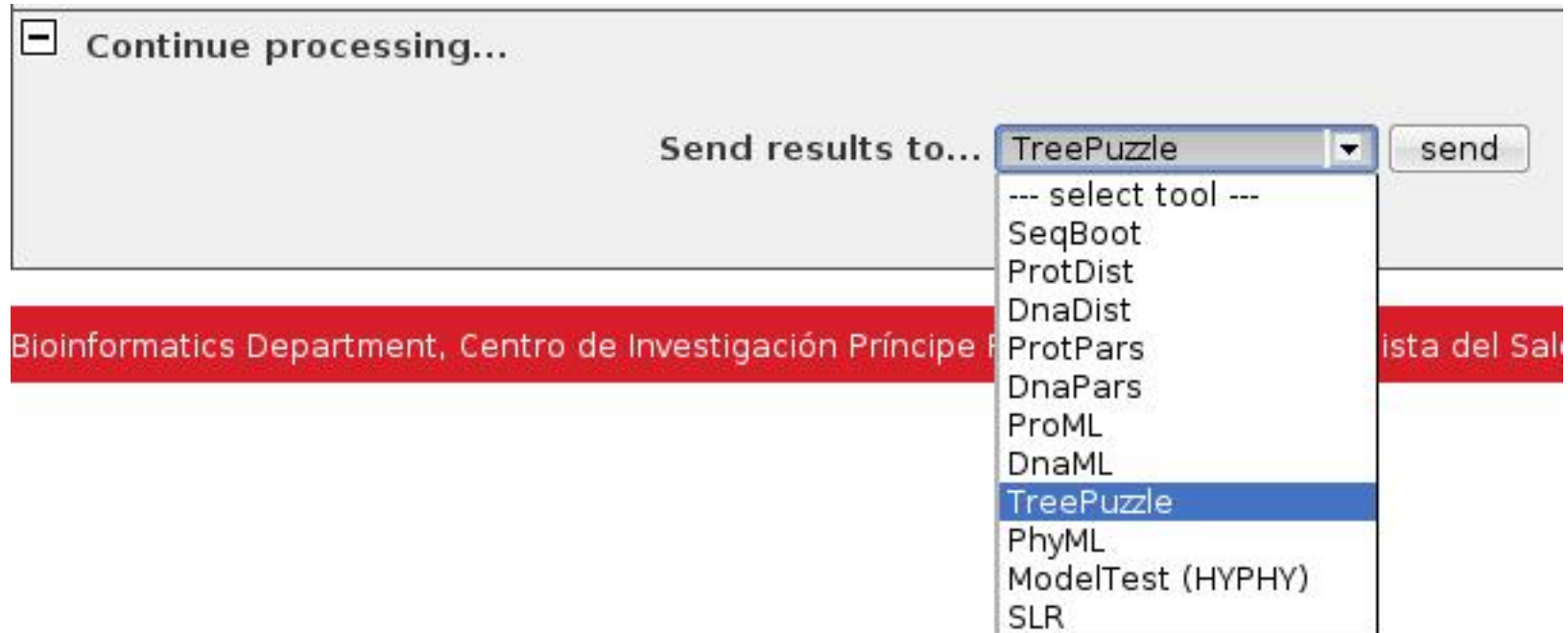
- Phylemon:

Servidor web con todas las herramientas integradas para el análisis de DNA y proteínas, desde un punto de vista filogenético y evolutivo.

Phylemon: Objetivos

Propone a los usuarios no-expertos un servidor web que evite problemas de:

- *Manipulación de datos*
- *Conversion de formatos*
- *Buscar las diferentes herramientas y coordinarlas*
- *Crear un flujo de información en varios pasos para procesar los datos*



[Home](#)
[Job summary page](#)
[User Guide](#)
[Methods](#)
[FAQ](#)
[Credits](#)
[Embedded Tools](#)
[Contact](#)

Results on orthologs tree :

- [O](#) *DCXR (HOMO)*

Explore the results on paralogs trees for the following proteins :

- [P](#) *DCXR (HOMO)*
- *DCXR (PTR)*
- *DCXR (MMU)*
- *Dcxr (MUS)*
- *DCXR (CAF)*
- *DCXR_BOVIN (BTA)*
- *DCXR (ECA)*
- *DCXR (MOD)*
- *DER_CHICK (GAL)*
- *DCXR (TGU)*
- *DCXR (GAC)*
- *dcxr (DAR)*
- *DCXR (ORL)*
- *dcxr (XET)*
- *DCXR (TNI)*

Other results from the same job :

- [O](#) *ARSA (HOMO)*
- [O](#) *HSPA5 (SSC)*

DATA FROM ORTHOLOGS TREE OF DCXR

SPECIES : *Human* - ID : *ENSP00000303356*

