

Alineamiento múltiple de secuencias

El alineamiento múltiple de secuencias es el paso previo para:

- Inferir filogenias moleculares
- Relacionar la secuencia primaria de las proteínas con su estructura 3D

Alineamiento múltiple

>Man

```
MENFPIVDMG KLNTEERKAT LDKMKDACEN WGGFELVNHG ISIELMDTVE KLTKEHYKCC MEQRFKEMVE SKGLEYVQSE
INDLDWESTF FLRHLPVSN SEIPDLQDY RKVMKEFAVK LEKLAEEELLD YLCENLGLK GYLKRVFYGS KGNPFGTKVS
NYPPCPKPEL IKGLRAHTDA GGNLLLFQDD KVSGLRLLKD DKWIDVPPMR HSIVINLGDQ LEVITSGKYK SVMHRVIAQT
DGARMSLASF YNPGDDALIS PAPTIVKENE TSEIYPKFVF DDYMKLYVGL KFQAKEPRFE AMKALSSVDV GPVVTA
```

>Rhesus

```
MEMDFPVINM NNLNGESRVS VLNQINDACE NWGGFELVNH GISHELMKDV EKLTKEHYRK CMEQRFKEMV ASKGLDSVET
EINDTDWEST FFLRHLPVSN MSEIGDLDEE YKKVMKEFAD ELEKLAEEVL DLLCENLGLE KGYLKKVFGY SKGNPFGTKV
SNYPPCPKPE LIKGLRAHTD AGGLILLFQD DKVSGLHVLK DGKWVDVPPM HHSIVINLGD QLEVITNGKY KSVHRVIAQ
EDGNRMSIAS FYNPGNDAVI YPAPALVEGE QEKTLYPKF VFDDYMKLYV GLKFQAKEPR FEAMKAMEST NLNMGPIATV
```

>Chimp

```
MENFPIVDMG KLNTEDRKST MELIKDACEN WGGFECVNHG ISIEMMDTVE KLTKEHYKCC MEQRFKEMVA TKGLECVQSE
IDDLWESTF FLRHLPVSSI SEIPDLDDDY RKVMKEFALK LEELAEELLD LLCENLGLK GYLKKAIFYGS KGNPFGTKVS
NYPPCPKPEL IKGLRAHTDA GGIILLFQDD KVSGLQLLKD DQWIDVPPMR HSIVINLGDQ LEVITNGKYK SVMHRVIAQT
DGARMSIASF YNPGDDAVIS PASTLLKENE TSEVYPKFVF DDYMKLYMGL KFQAKEPRFE AMKAMSSVK VGPVVSI
```

>Gorilla

```
MANFPVDMG KLNTEERGTA MEMIKDACEN WGGFELVNHG ISIELMDTVE KLTKEHYKKT MEQRFKEMVA NKGLESVQSE
INDLDWESTF FLRHLPVSNV SENTDLQDY RKIMKQFAEE LEKLAEEHLLD LLCENLGLK GYLKRVFYGS KGNPFGTKVS
NYPPCPTPDL IKGLRAHTDA GGIILLFQDD KVSGLQLLKD DQWIDVPPMR HSIVINLGDQ LEVITNGKYK SVMHRVIAQT
DGTRMSLASF YNPGDDAVIS PAPALVKESD ETSQVYPKFV FNDYMKLYAG LKFQAKEPRF EAMKAVSSVD VGAIATV
```

>Orangutan

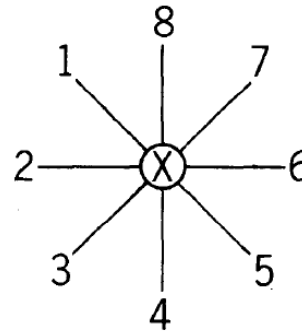
```
MENFPIINLE NLNGDERAKT MEMIKDACEN WGGFELVNHG IPHEVMDTVE KLTKGHYKCC MEQRFKELVA SKGLEAVQAE
VTDLDWESTF FLRHLPTSNI SQVPDLDEEY REVMRDFAKR LEKLAEEELLD LLCENLGLK GYLKNAIFYGS KGNPFGTKVS
NYPPCPKPD L IKGLRAHTDA GGIILLFQDD KVSGLQLLKD EQWIDVPPMR HSIVVNLGDQ LEVITNGKYK SVMHRVIAQT
DGTRMSLASF YNPGNDAVIY PAPSLEEESK QVYPKFVFDD YMKLYAGLKF QPKPRFEAM KAMEANVELV DQIASA
```

The neighbor-joining method

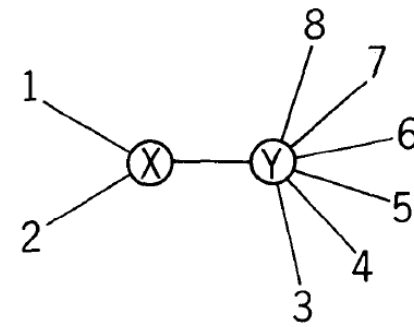
Table 1
Distance Matrix for the Tree in Figure 1

OTU	OTU						
	1	2	3	4	5	6	7
2 ..	7						
3 ..	8	5					
4 ..	11	8	5				
5 ..	13	10	7	8			
6 ..	16	13	10	11	5		
7 ..	13	10	7	8	6	9	
8 ..	17	14	11	12	10	13	8

En cada paso, se calcula la suma de todas las ramas y se elige el par de OTUs que origina la suma mínima.



(a)



(b)

FIG. 2.—(a), A starlike tree with no hierarchical structure; and (b), a tree in which OTUs 1 and 2 are clustered.

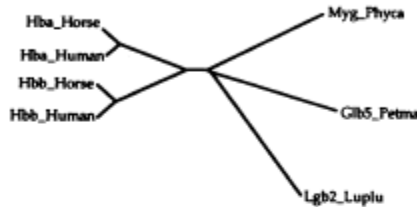
Saitou N, Nei M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 4(4):406-25.

CLUSTAL W

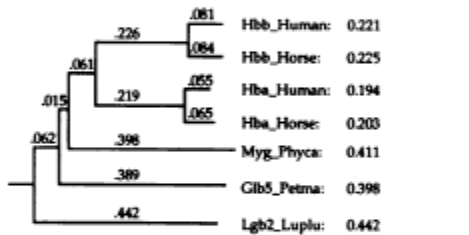
Pairwise alignment:
Calculate distance matrix

Hbb_Human	1	-					
Hbb_Horse	2	.17	-				
Hba_Human	3	.59	.60	-			
Hba_Horse	4	.59	.59	.13	-		
Myg_Phycs	5	.77	.77	.75	.75	-	
Glb5_Petma	6	.81	.82	.73	.74	.80	
Lgb2_Luplu	7	.87	.86	.86	.88	.93	.90
		1	2	3	4	5	6

Unrooted Neighbor-Joining tree



Rooted NJ tree (guide tree)
and sequence weights



Progressive alignment:
Align following the guide tree



1. Se alinean las secuencias separadamente de dos en dos, obteniéndose una matriz de distancias por pares
2. Se deriva un árbol guía a partir de la matriz de distancias
3. Las secuencias se van alineando progresivamente de acuerdo con el árbol guía
4. Los gaps introducidos en las primeras etapas se respetan en etapas posteriores

CLUSTAL W (1.82) multiple sequence alignment

```

Man      MEN-FPIVDMGKLNTEERKATLDKMKDACENWGGFFELVNHGISIELMDTVEKLTKEHYKK 59
Chimp    MEN-FPIVDMGKLNTEEDRKSTMELIKDACENWGGFFECVNHGISIEMMDTVEKLTKEHYKK 59
Gorilla  MAN-FPVVDMGKLNTEERTAMEMIKDACENWGGFFELVNHGISIELMDTVEKLTKEHYKK 59
Orangutan MEN-FPIINLENLNGDERAKTMEMIKDACENWGGFFELVNHGISIHEVMDTVEKLTKEHYKK 59
Rhesus   MEMDFPVINMNNLNGESRSVSLNQLINDACENWGGFFELVNHGISIELMDKVEKLTKEHYRK 60
*      **::: : ** :.*  .:  ::***** ***** .:*:.***** **:*

Man      CMEQRFKEMVESKGLEYYQSEINDLDWESTFFLRHLPVSNISEIPDLDDQDYRKVMKEFAV 119
Chimp    CMEQRFKEMVATKGLECVQSEIDDLWESTFFLRHLPVSSISEIPDLDDDYRKVMKEFAL 119
Gorilla  TMEQRFKEMVANKGLESVQSEINDLDWESTFFLRHLPVSNVSENTDLDDQDYRKIMQFAE 119
Orangutan CMEQRFKELVASKGLEAVQAEVTDLDWESTFFLRHLPVSNISQVPLDDEEYREVMRDFAK 119
Rhesus   CMEQRFKEMVASKGLDSVETEINDTDWESTFFLRHLPVSNMSEIGDLDEEYKVMKEFAD 120
*****:* .***: *::: * *****.*.:* : **:::***:***

Man      KLEKLAEEELLDYLCENLGLEKGYLKKVIFYGSKGPNFGTKVSNYPPCPKPELIKGLRAHTD 179
Chimp    KLEELAEEELLDLLCENLGLEKGYLKKAFYGSKGPNFGTKVSNYPPCPKPELIKGLRAHTD 179
Gorilla  ELEKLAEHLDDLCCENLGLEKGYLKNAFYGSKGPNFGTKVSNYPPCPTPDLIKGLRAHTD 179
Orangutan RLEKLAEEELLDLLCENLGLEKGYLKNAFYGSKGPNFGTKVSNYPPCPKPDLIKGLRAHTD 179
Rhesus   ELEKLAEEVLDLLCENLGLEKGYLKKVIFYGSKGPNFGTKVSNYPPCPKPELIKGLRAHTD 180
.*:*.*.*.* ** *****.* *****.*:*****

Man      AGGNLLLFQDDKVSGLRLLKDDKWIDVPPMRHSIVINLGDQLEVITSGKYKVMHRVIAQ 239
Chimp    AGGIILLFQDDKVSGLQLLKDDQWIDVPPMRHSIVINLGDQLEVITNGKYKVMHRVIAQ 239
Gorilla  AGGIILLFQDDKVSGLQLLKDDQWIDVPPMRHSIVINLGDQLEVITNGKYKVMHRVIAQ 239
Orangutan AGGIILLFQDDKVSGLQLLKDEQWIDVPPMRHSIVVNLGDQLEVITNGKYKVMHRVIAQ 239
Rhesus   AGGLILLFQDDKVSGLHVLKDGKWDVPPMHHSIVINLGDQLEVITNGKYKVMHRVIAQ 240
*** .*****:*** :*:*****:*****:*****.*****

Man      TDGARMSIASFYNPGD DALISPAPTLVKE-NETSEIYPKFVFDYMKLYVGLKFQAKEPR 298
Chimp    TDGARMSIASFYNPGD DAVIS PASTLLKE-NETSEVYPKFVFDYMKLYMGLKFQAKEPR 298
Gorilla  TDGTRMSIASFYNPGD DAVIS PAPANLVESEDETSQVYPKFVFDYMKLYAGLKFQAKEPR 299
Orangutan TDGTRMSIASFYNPGD DAVIYPAPSLIEE---SKQVYPKFVFDYMKLYAGLKFQAKEPR 296
Rhesus   EDGNRMSIASFYNPGD DAVIYPAPALVEGEQEKTCLYPKFVFDYMKLYVGLKFQAKEPR 300
** ***:*****:***: **.:*:. .:*****:***** *****

Man      FEA-MKALSS--VDVGPVVTA 316
Chimp    FEAMMKAMSS--VKVGPVVISI 317
Gorilla  FEA-MKAVSS--VDVGAIAIV 317
Orangutan FEA-MKAMEANVELVDQIASA 316
Rhesus   FEA-MKAMESTNLMGPIAIV 320
*** **:::  .:  .:

```

➔ Secuencia consenso

Chimp	PNFGTKVSNYPKPELIKGLRAHTDAGGI ILLFQDDKVSGLQLLKDDQW
Gorilla	PNFGTKVSNYPKPTPDLIKGLRAHTDAGGI ILLFQDDKVSGLQLLKDDQW
Man	PNFGTKVSNYPKPELIKGLRAHTDAGGNLLLFQDDKVSGLRLLKDDKW
Rhesus	PNFGTKVSNYPKPELIKGLRAHTDAGGL ILLFQDDKVSGLHVLKDGKW
Orangutan	PNFGTKVSNYPKPDLIKGLRAHTDAGGI ILLFQDDKVSGLQLLKDEQW

cons *****.*:*****:*****::***:*

Chimp	IDVPPMRHSIVINLGDQLEVI TNGKYKSMHRVIAQTDGARMSSIASFYNPG
Gorilla	IDVPPMRHSIVINLGDQLEVI TNGKYKSMHRVIAQTDGTRMSLASFYNPG
Man	IDVPPMRHSIVINLGDQLEVI TSGKYKSMHRVIAQTDGARMSSIASFYNPG
Rhesus	VDVPPMHHSIVINLGDQLEVI TNGKYKSMHRVIAQEDGNRMSIASFYNPG
Orangutan	IDVPPMRHSIVVNLGDQLEVI TNGKYKSMHRVIAQTDGTRMSLASFYNPG

cons :*****:*****:*****.***** ** ***:*****

Chimp	DDAVISPASTLLKE-NETSEVYPKFVFDYMKLYMGLKFQAKEPRFEAMMK
Gorilla	DDAVISPAPALVKESDETSQVYPKFVFNKYMKLYAGLKFQAKEPRFEAM-K
Man	DDALISPAPTLVKE-NETSEIYPKFVFDYMKLYVGLKFQAKEPRFEAM-K
Rhesus	NDAVIYPAPALVEGEQEKTKLYPKFVFDYMKLYVGLKFQAKEPRFEAM-K
Orangutan	NDAVIYPAPSLIEE---SKQVYPKFVFDYMKLYAGLKFQAKEPRFEAM-K

cons :**:* **.:*:. .:*****:***** *****.***** *

Chimp	AMSS--VKVGPVVS I
Gorilla	AVSS--VDVGAIATV
Man	ALSS--VDVGPVVTA
Rhesus	AMESTNLNMGPIATV
Orangutan	AMEANVELVDQIASA

cons *:.: .: .: .:

JalView

Start Jalview

Score

78.0

89.0

86.0

77.0

76.0

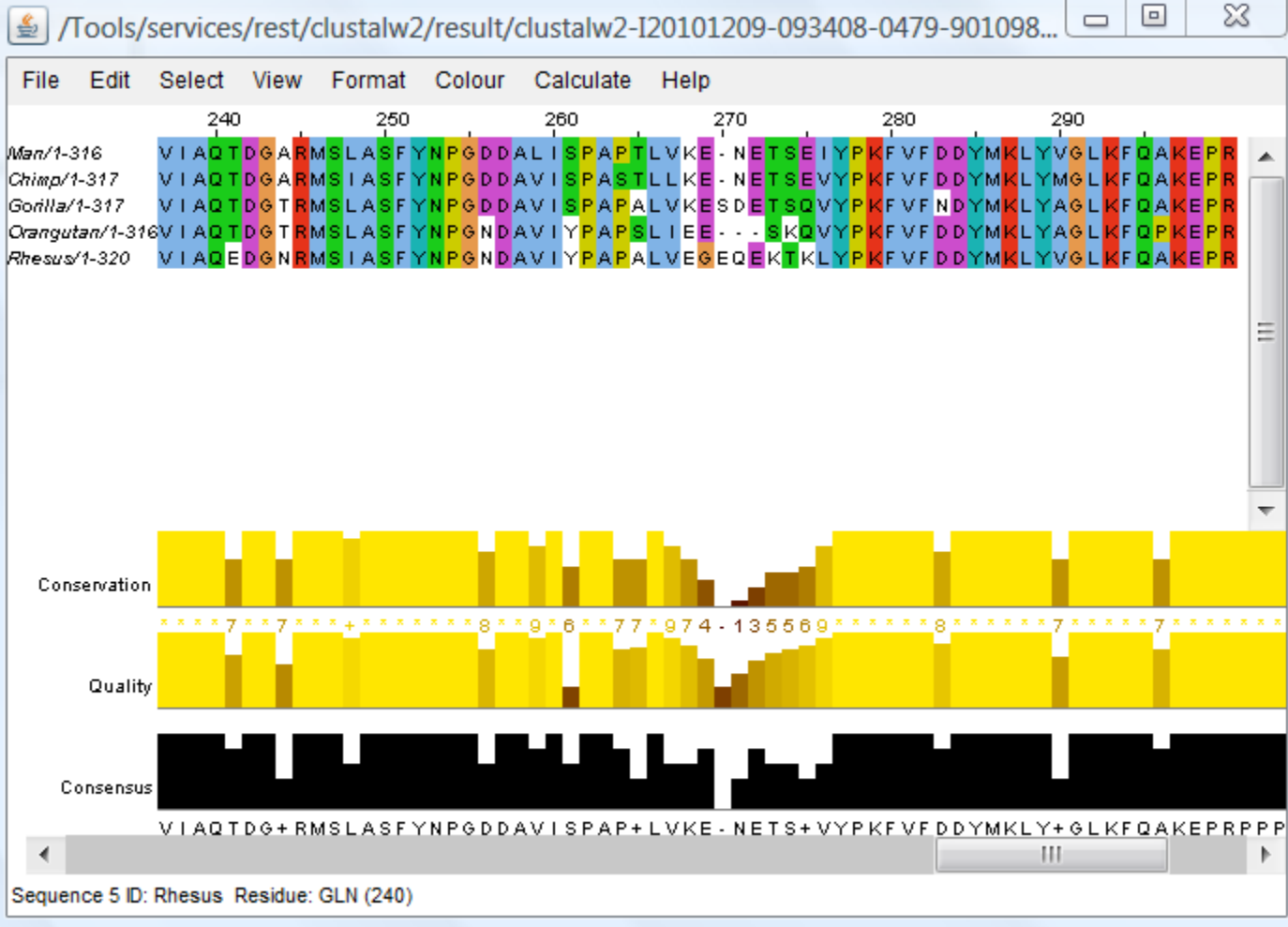
77.0

77.0

86.0

79.0

79.0



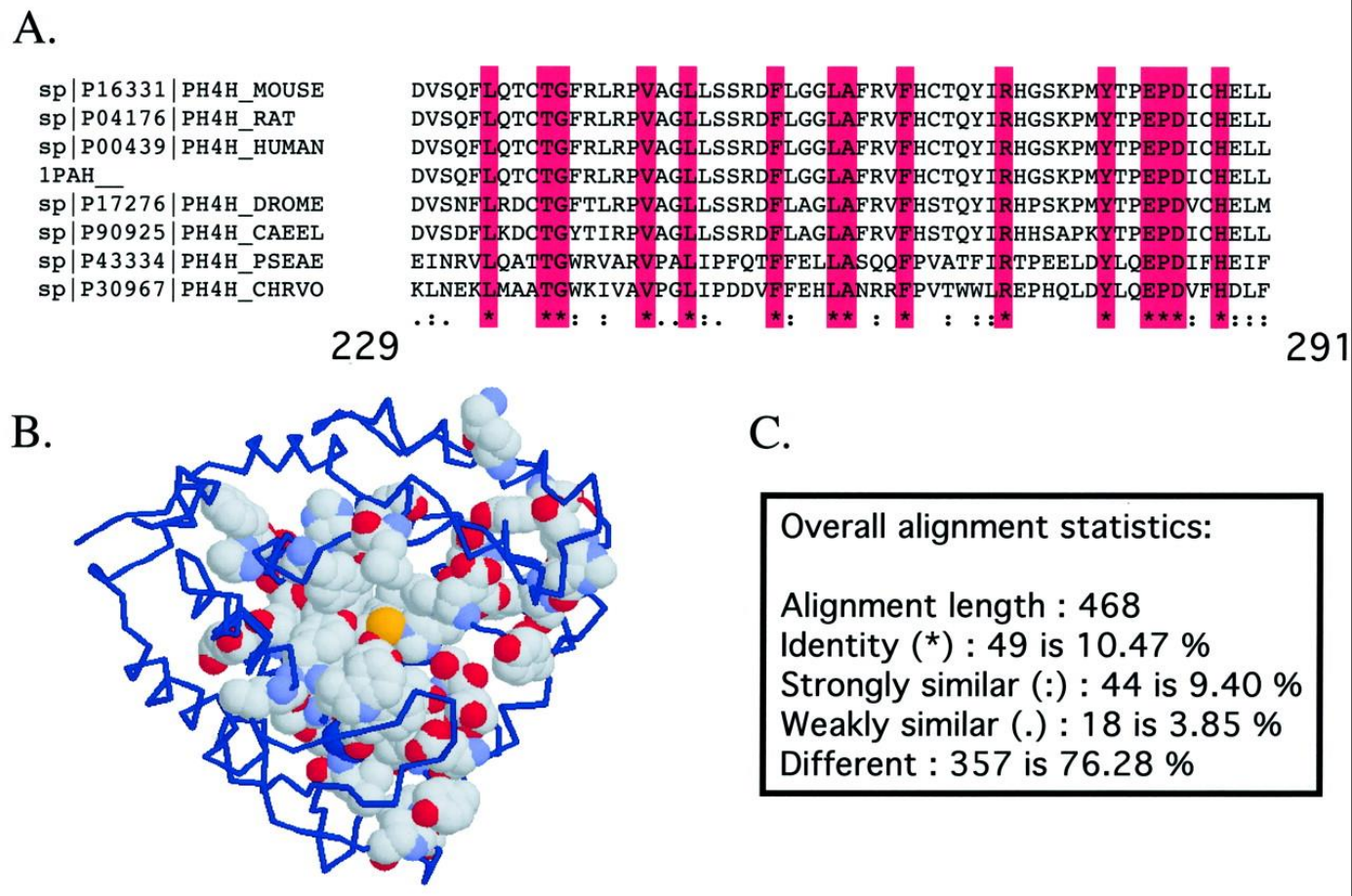


FIG. 3. Sample output from Project 4. *A*, a portion of the multiple sequence alignment used for the analysis of phenylalanine hydroxylase generated by the program ClustalW. Swiss-Prot accession numbers for the sequences used in this exercise were as follows: *P00439*, *Homo sapiens* PAH; *P04176*, *Rattus norvegicus* PAH; *P16331*, *Mus musculus* PAH; *P30967*, *Chromobacterium violaceum* PAH; *P43334*, *Pseudomonas aeruginosa* PAH; *P17276*, *Drosophila melanogaster* PAH; *P90925*, *Caenorhabditis elegans* PAH; and *1PAH*, the sequence from the crystallographic-solved PAH fragment from *Rattus Norvegicus* (13). *B*, structure of PAH in which the 100% conserved residues, identified through the sequence alignment in *A*, are shown in space-filling mode, whereas the remainder of the protein is shown as a backbone trace. *C*, overall statistics of the ClustalW alignment.