
Master Biomedizin 2016

Phylogeny

11

- a.** Using the proteins “P13056” and “P49116”, find their orthologs in UniProt in the following organisms: mouse (*Mus musculus*), platypus (*Ornithorhynchus anatinus*), frog (*Xenopus laevis*) and chicken (*Gallus gallus*). Choose reviewed entries whenever possible. Save the sequences in a fasta file.
- b.** Which phylogenetic relations can you describe using these sequences? Use Trex.



Mouse
(*Mus musculus*)



Chicken (rooster)
(*Gallus gallus*)



Frog
(*Xenopus laevis*)



Platypus
(*Ornithorhynchus anatinus*)

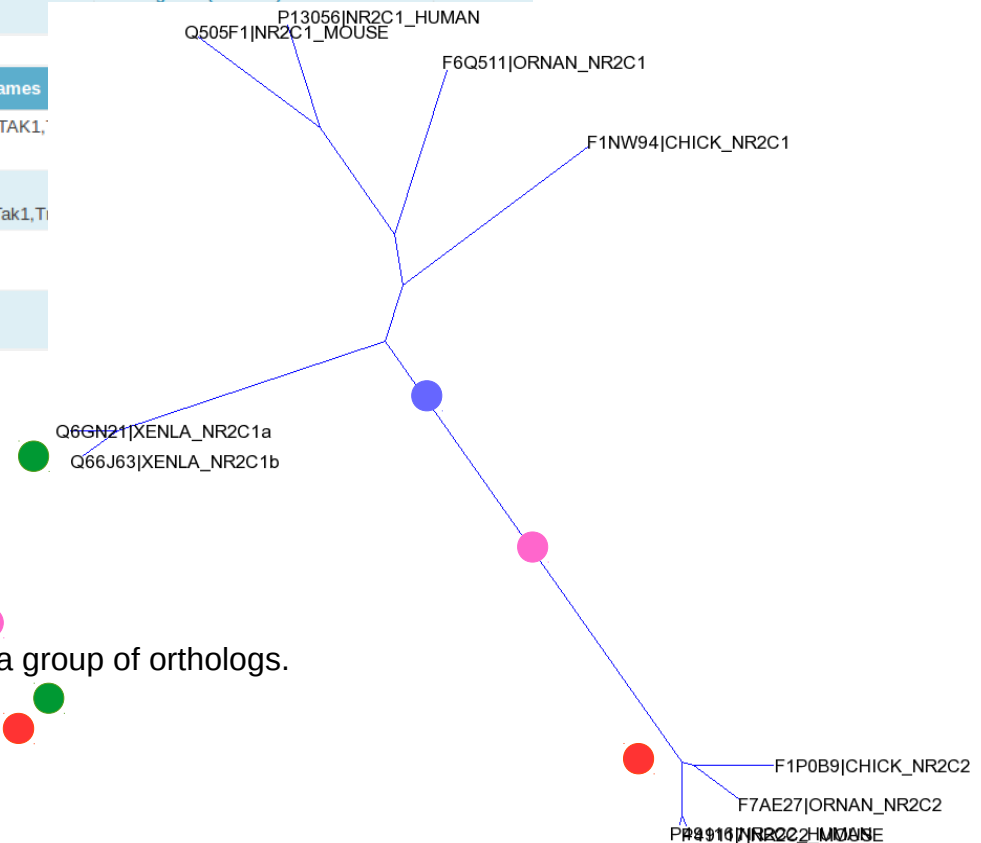
Homology + Phylogeny

*Images from: UniProt, Trex

11

■	Your list:....AXKGJ	Entry	Entry name	Protein names	Gene names	Organism	Length
<input type="checkbox"/>	P13056	P13056	NR2C1_HUMAN	Nuclear receptor subfamily 2 group ...	NR2C1 TR2	Homo sapiens (Human)	603
<input type="checkbox"/>	Q505F1	Q505F1	NR2C1_MOUSE	Nuclear receptor subfamily 2 group ...	Nr2c1 Tr2,Tr2-11	Mus musculus (Mouse)	590
<input type="checkbox"/>	F6Q511	F6Q511	F6Q511_ORNAN	Uncharacterized protein	NR2C1	Ornithorhynchus anatinus (Duckbill platypus)	652
<input type="checkbox"/>	Q6GN21	Q6GN21	N2C1A_XENLA	Nuclear receptor subfamily 2 group ...	nr2c1-a dor2	Xenopus laevis (African clawed frog)	637
<input type="checkbox"/>	Q66J63	Q66J63	N2C1B_XENLA	Nuclear receptor subfamily 2 group ...	nr2c1-b dor2	Xenopus laevis (African clawed frog)	637
<input type="checkbox"/>	F1NW94	F1NW94	F1NW94_CHICK	Uncharacterized protein	NR2C1	Gallus gallus (Chicken)	562

■	Your list:....UZYS3	Entry	Entry name	Protein names	Gene names
<input type="checkbox"/>	P49116	P49116	NR2C2_HUMAN	Nuclear receptor subfamily 2 group ...	NR2C2 TAK1,
<input type="checkbox"/>	P49117	P49117	NR2C2_MOUSE	Nuclear receptor subfamily 2 group ...	Nr2c2 Mtr2r1,Tak1,T
<input type="checkbox"/>	F7AE27	F7AE27	F7AE27_ORNAN	Uncharacterized protein	NR2C2
<input type="checkbox"/>	F1P0B9	F1P0B9	F1P0B9_CHICK	Uncharacterized protein	NR2C2



a. 10 sequences.

- b.
- 1) All NR2C1 proteins together in one branch, as well as the NR2C2 proteins in another branch.
 - 2) All of them are homologs, and each branch contains a group of orthologs.
 - 3) In relation to NR2C1, *X.laevis* suffered a duplication.
 - 4) In relation to NR2C2, *X.laevis* suffered a gene loss.

14

Consider the following sequences: P50225, P50226, P0DMM9, P0DMN0, O43704, O00338, Q6IMI6, O75897. All of them are part of a human paralog family (sulfotransferases).

- a.** Look for the coordinates of the substrate binding region in the UniProt entry “P50225”. Do you expect all the paralogs to share this substrate binding region? Why? Check it.
- b.** Using the previous sequences, could you determine which human paralog is the ortholog of the sequence in “*file4.fasta*” (<https://cbdm.uni-mainz.de/mb16/>)? Do not use BLAST.
- c.** Which protein is the one found in “*file4.fasta*”?

14

(P50225)

Feature key	Position(s)		Length	Description
Region ⁱ	106 – 108		3	Substrate binding
10	20	30	40	50
MELIQDTSRP	PLEYVKGVP	IKYFAEALGP	LQSFQARPDD	LLISTYPKSG
60	70	80	90	100
TTWSQILDM	IYQGGDEK	HRAPIFMRVP	FLEFKAPGIP	SGMETLKDTP
110	120	130	140	150
APRLLKTHLP	LALLPQTLDD	QKVKVYVVAR	NAKDVAVSY	HFYHMAKVHP
160	170	180	190	200
EPGTWDSFLE	KFMVGEVSYG	SWYQHVQEW	ELSRTHPVL	LFYEDMKENP
210	220	230	240	250
KREIQKILEF	VGRSLPEETV	DFVVQHTSFK	EMKKNPMTNY	TTVPQEFMDH
260	270	280	290	
SISPFMRKGM	AGDWKTTFTV	AQNERFDADY	AEKMAGCSLS	FRSEL

Highlight

Annotation

- ☐ Natural variant
- ☒ Region
- ☐ Binding site
- ☐ Nucleotide binding
- ☐ Beta strand
- ☐ Active site
- ☐ Mutagenesis
- ☐ Modified residue
- ☐ Alternative sequence

P50225	ST1A1_HUMAN
P50226	ST1A2_HUMAN
P0DMM9	ST1A3_HUMAN
P0DMN0	ST1A4_HUMAN
O43704	ST1B1_HUMAN
O00338	ST1C2_HUMAN
Q6IMI6	ST1C3_HUMAN
O75897	ST1C4_HUMAN

P50225	ST1A1_HUMAN
P50226	ST1A2_HUMAN
P0DMM9	ST1A3_HUMAN
P0DMN0	ST1A4_HUMAN
O43704	ST1B1_HUMAN
O00338	ST1C2_HUMAN
Q6IMI6	ST1C3_HUMAN
O75897	ST1C4_HUMAN



Align

P50225
P50226
P0DMMS
P0DMNG
043704
000338
Q6IMI6
075897

```

1  -----MELIQDTSRPPLLEYVKGVP LKYFAEALGPLQSFQARPDDL ISTYPKSGTT 52
1  -----MELIQDTSRPPLLEYVKGVP LKYFAEALGPLQSFQARPDDL ISTYPKSGTT 52
1  -----MELIQDTSRPPLLEYVKGVP LKYFAEALGPLQSFQARPDDL INTYPKSGTT 52
1  -----MELIQDTSRPPLLEYVKGVP LKYFAEALGPLQSFQARPDDL INTYPKSGTT 52
1  -----MLSPKDILRKDLKLHVHGYPMTCASFANWKEIQHFSRPDDI VIATYPKSGTT 52
1  -----MALT-SDLGKQIKLKEVEGTLLOPATVDNWSQISQFEAKPDDL ICTYPKAGTT 52
1  MAKIEKNAPTMEKKPELFNIMEVDGVP TLLSKWEKVEKCNFAQKPDLL IATYPKSGTT 60
1  MALHDMEDFT- FDGTKRLSVNYVKGILQPTDTCDIWDKIWNFAQKPDLL ISTYPKAGTT 59


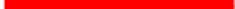






```

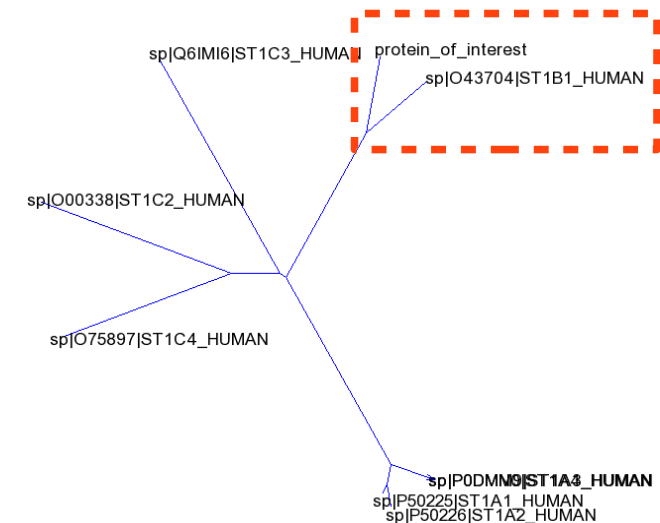
```

53 WVSQLDMIIYGGDLEKCHRAPIFMRVPFLEFKAPGPI-SGMETLKDTPAPRLLT111
53 WVSQLDMIIYGGDLEKCHRAPIFMRVPFLEFKAPGPI-SGMETLKNTAPRLT111
53 WVSQLDMIIYGGDLEKCNRAPIYVRVPFLEVNDPGEP-SGLETLKDTPPPRLT111
53 WVSQLDMIIYGGDLEKCNRAPIYVRVPFLEVNDPGEP-SGLETLKDTPPPRLT111
53 WVEIIDIMLNDGDEIEKCKRGFITEKVPMLTEIPLGLRTSGIEQLEKNPSPRIV112
54 IQTEIVDMIEQNGDVEKQCORAIIOHRHPFIEWARPPQ-SGVEKAKAMPSRLT112
61 WMHETLMDILNDGDEVCKKRAQTLDRHAFLEKPHKEKPDLEFVLEMSSPQLT120
60 WQTEVLEIQNEDGVEKSKRAPHQRFPLEMKIPSLG-SGLEQAHAMSPRLT118

```

- a.** Yes, because they share the motif “K[ST]H”.
- b.** The ortholog of “protein_of_interest” is ST1B1_HUMAN.
- c.** ST1B1_CANLF, from *Canis familiaris* (dog).

Entry	Protein names	Match hit	Identity
Q95JD5	 Sulfotransferase family cytosolic 1B member 1 (Canis lupus familiaris)		100.0%
F6V3V8	 Sulfotransferase (Canis lupus familiaris)		98.3%
D2I2Q5	 Sulfotransferase (Ailuropoda melanoleuca)		94.3%
U6CQE2	 Sulfotransferase (Neovison vison)		93.6%

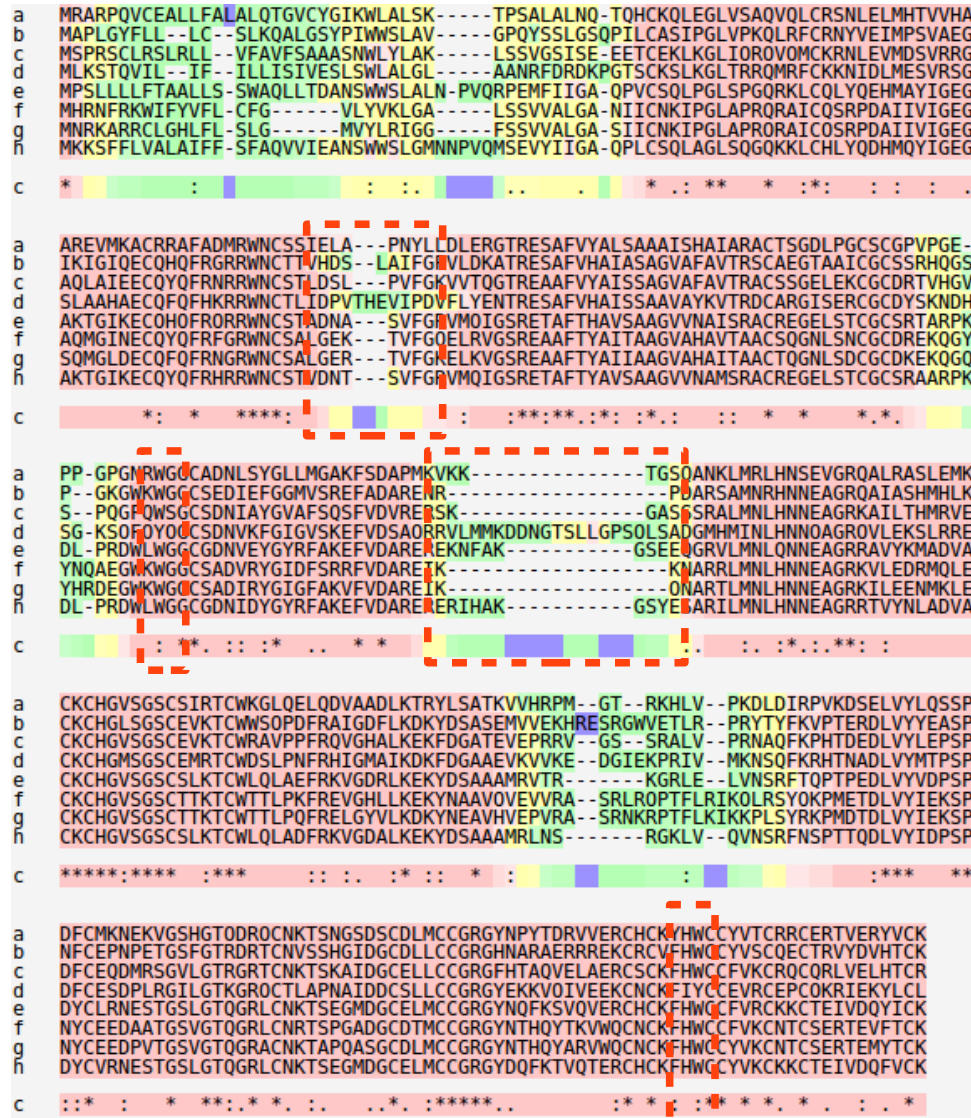


15

Using T-Coffee, align the protein sequences found in “*file3.fasta*” (<https://cbdm.uni-mainz.de/mb16/>). All of them are paralogs, but one ancestor protein to be used as outgroup.

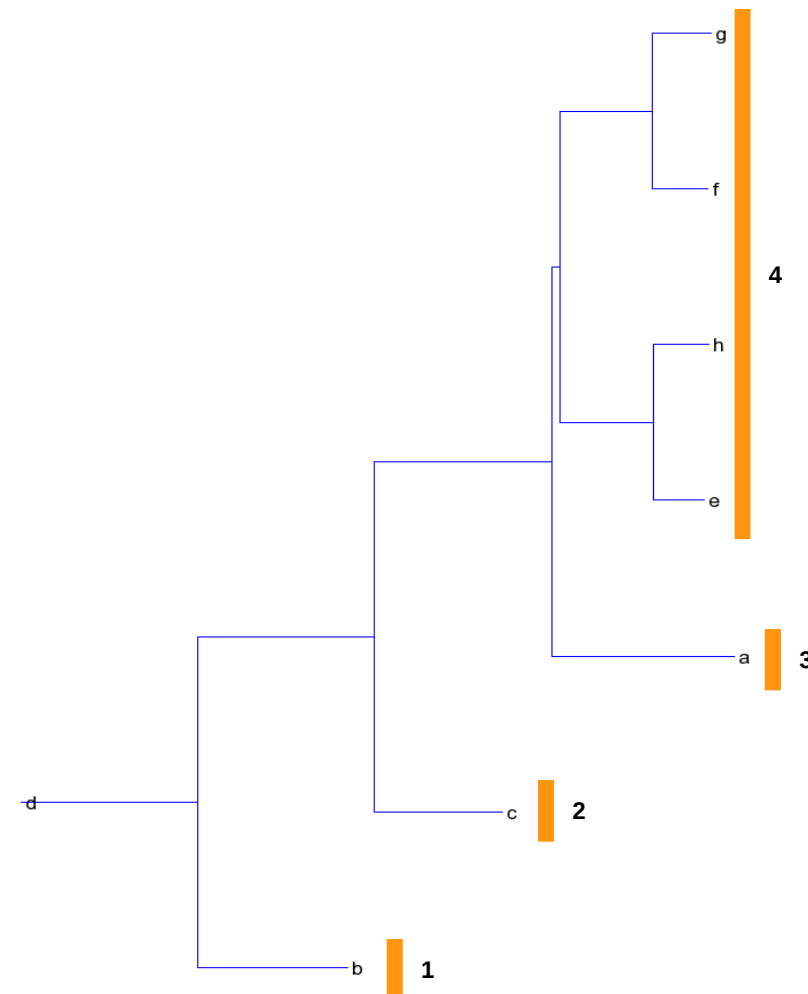
- a. Could you infer from the alignment which sequence should be used as outgroup?
- b. In what order were the paralogs originated after diverging from the ancestor?

15



a. Outgroup: "d". Hints in red boxes.

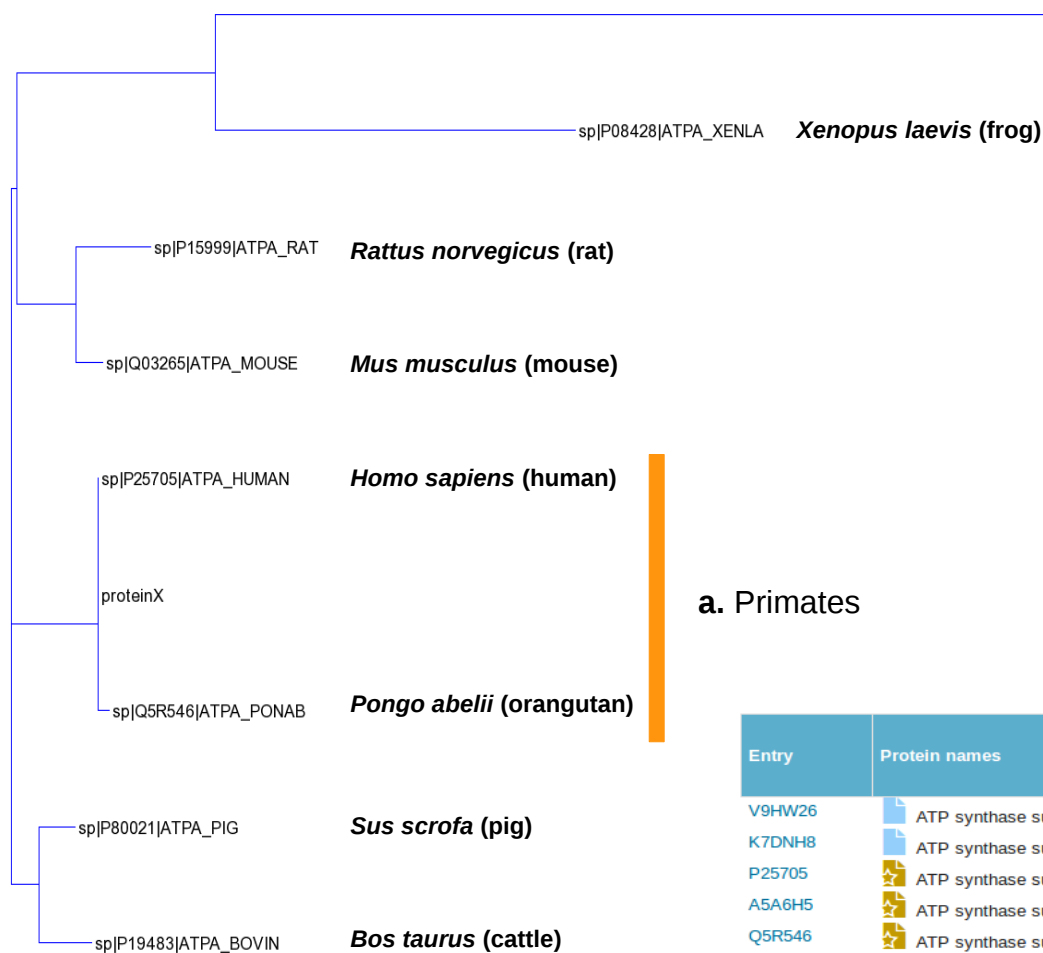
b. Order: b, c, a, ancestor of [e+h] & [f+g].



16

- a. Using “*file2.fasta*” (<https://cbdm.uni-mainz.de/mb16/>) and Trex, can you approximate to which taxonomic division belongs “proteinX”?
- b. From which organism could it be? After guessing, check it.

16



a. Primates

UniProt

BLAST Align Retrieve/ID mapping

UniProtKB

BLAST

How to use this tool

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity evolutionary relationships between sequences as well as help identify members

>proteinX
MLSVRVAAVVRALPRRAGLVSRNALGSSFIARNFHASNTHLQKTGAEMSSILEERIL
GADTSVDLEETGRVLISGDGIARVHGLRNVAEEMVEFSSGLKGMNLNLEPDNVGVVFG
NDKLIKEGDIVKRTGAIQVDPVGEELLGRVVDALGNAIDGKPIGSKTRRRVGLKAPGII
PRISVREPMQTGIKAVDSLVPVIGRQRELIIGDRQTGKTSIAIDTIINQKRFNDGSEKK
KLYCIYVAIGQKRSTVAQLVKRLTDADAMKYTIIVSATASDAAPLQYLAPYSGCSMGVEYF
RDNGKHALIIYDDLKQAVAYRQMSLLRRPPGREAYPGDVLYLHSLLEAAKMDAFG
GGSLTALPVIETQAGDVSAIPTNVISITDGGQIFLETIFYKGIKIRPAINVGLSVSRVGS

Target databaseⁱ E-Thresholdⁱ Matrixⁱ Filterⁱ

...Mammals 10 Auto None

☐ Run Blast in a separate window.

Clear Run BLAST

Entry	Protein names	Match hit	Identity
V9HW26	ATP synthase subunit alpha (Homo sapiens)	<div><div></div></div>	100.0%
K7DNIH8	ATP synthase subunit alpha (Pan troglodytes)	<div><div></div></div>	100.0%
P25705	ATP synthase subunit alpha, mitochondrial (Homo sapiens)	<div><div></div></div>	100.0%
A5A6H5	ATP synthase subunit alpha, mitochondrial (Pan troglodytes)	<div><div></div></div>	100.0%
Q5R546	ATP synthase subunit alpha, mitochondrial (Pongo abelii)	<div><div></div></div>	99.8%

b. *Homo sapiens* (human) or *Pan troglodytes* (chimpanzee), they are 100% identical.

17

Human hemoglobin consists of four protein subunits: two from the alpha globin gene cluster (located on chromosome 16) and two more from the beta globin gene cluster (located on chromosome 11). But there are at least nine different globin genes in these clusters, which are: zeta, alpha, gamma2, theta1, mu, beta, gamma1, epsilon, alpha, delta.

- a.** Find the protein sequence for each one of them.
- b.** Sort them either in cluster alpha or cluster beta.
- c.** Why do you think they are clustered in either cluster alpha or cluster beta?

Phylogeny

*Images from: UniProt, Trex, UCSC

17

<input type="checkbox"/>	Your list...ITXRE	Entry	Entry name	Protein names
<input type="checkbox"/>	HBAZ_HUMAN	P02008	HBAZ_HUMAN	Hemoglobin subunit zeta
<input type="checkbox"/>	HBM_HUMAN	Q6B0K9	HBM_HUMAN	Hemoglobin subunit mu
<input type="checkbox"/>	HBA_HUMAN	P69905	HBA_HUMAN	Hemoglobin subunit alpha
<input type="checkbox"/>	HBAT_HUMAN	P09105	HBAT_HUMAN	Hemoglobin subunit theta-1
<input type="checkbox"/>	HBE_HUMAN	P02100	HBE_HUMAN	Hemoglobin subunit epsilon
<input type="checkbox"/>	HBG1_HUMAN	P69891	HBG1_HUMAN	Hemoglobin subunit gamma-1
<input type="checkbox"/>	HBG2_HUMAN	P69892	HBG2_HUMAN	Hemoglobin subunit gamma-2
<input type="checkbox"/>	HBD_HUMAN	P02042	HBD_HUMAN	Hemoglobin subunit delta
<input type="checkbox"/>	HBB_HUMAN	P68871	HBB_HUMAN	Hemoglobin subunit beta

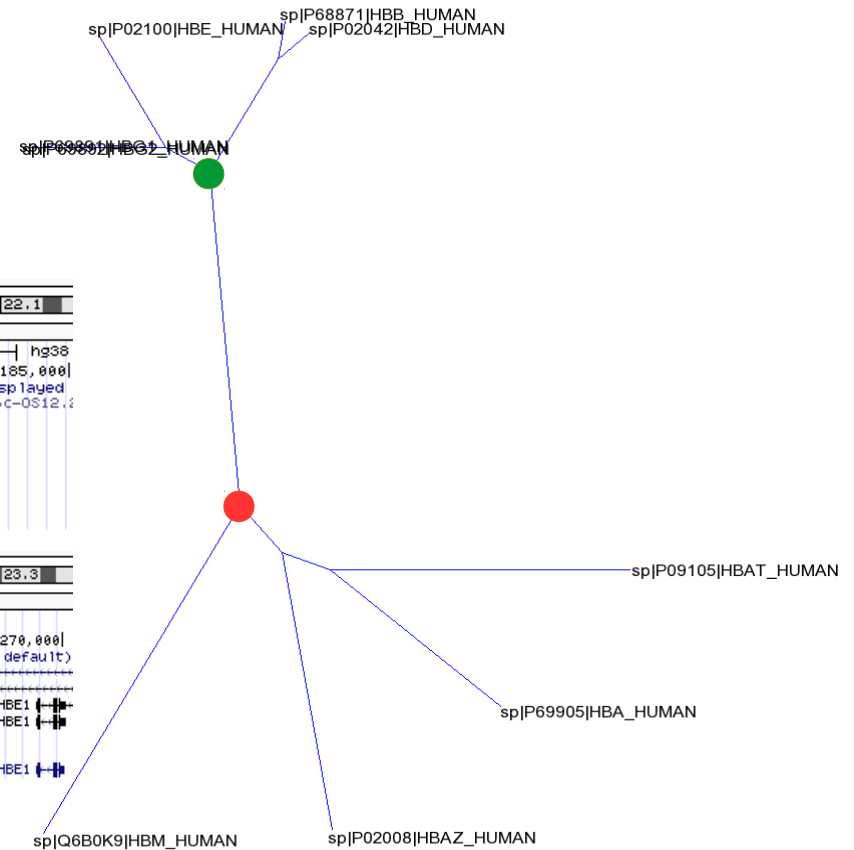
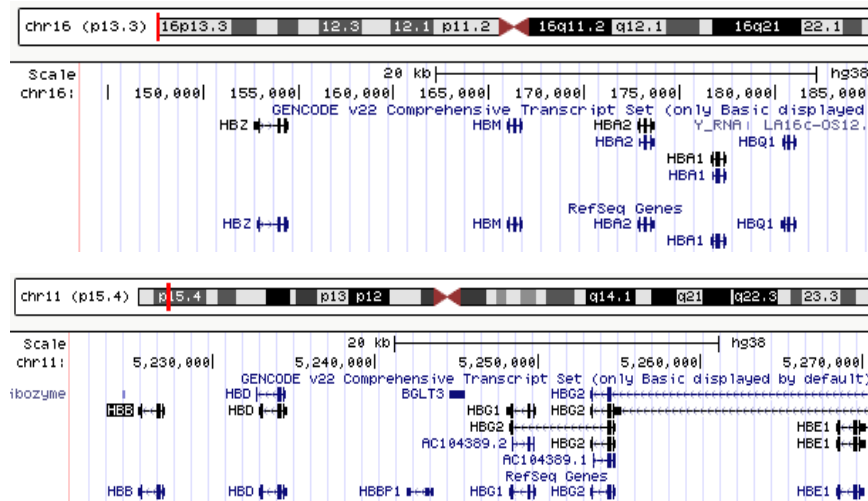
a. In UniProt.

b. Using Trex:

● Cluster alpha

● Cluster beta

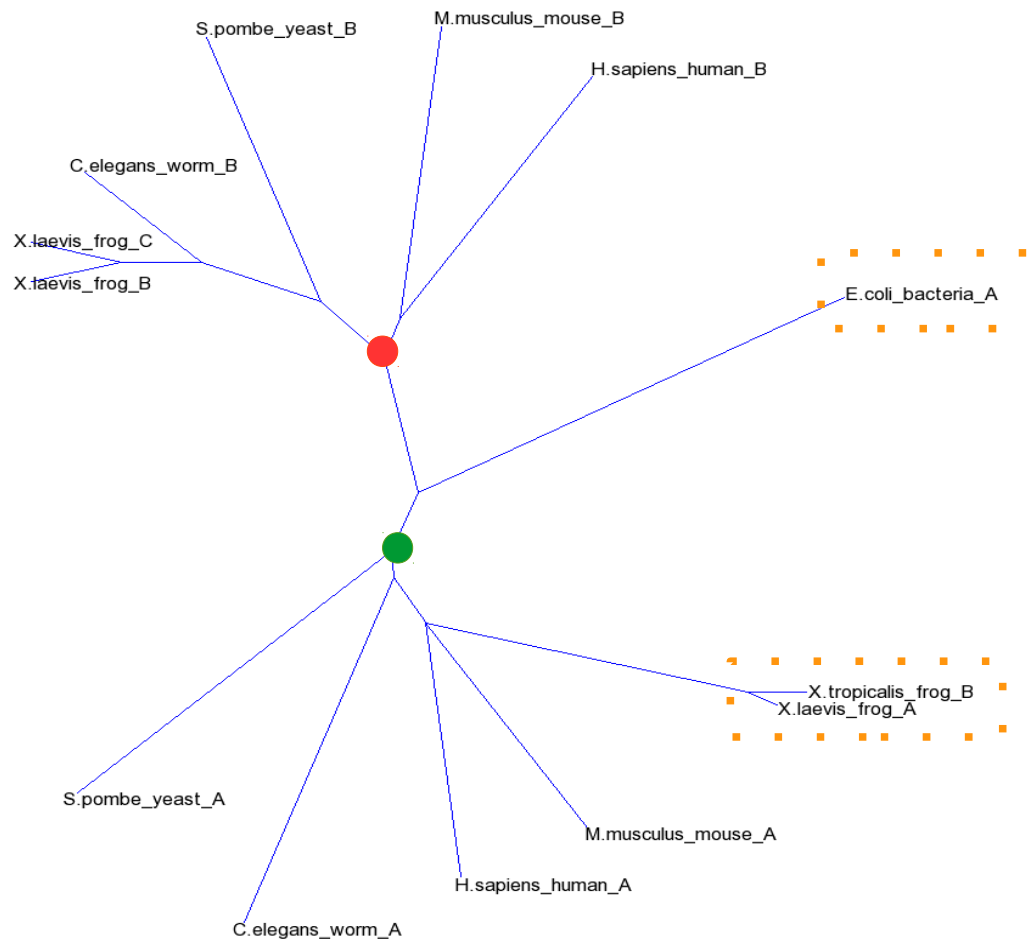
c. Paralogous expansion from one ancestral alpha and one ancestral beta.



18

- a. All of the sequences in “*file1.fasta*” (<https://cbdm.uni-mainz.de/mb16/>) are homologs. How many groups of orthologs would you say there are in this file?
- b. What could you say about the history of this protein family?
- c. Would you say there is any wrongly annotated sequence?

18



a. Two groups of orthologs:

● Proteins "A"

● Proteins "B"

b. E.coli has only one protein, and then it duplicated to form A and B. It is possible that X.laevis_B duplicated later to form B and C.

c. X.tropicalis_B is wrongly annotated. It should be X.tropicalis_A, because they are in the same branch. The actual X.tropicalis_B is not in the dataset, or was lost during evolution.

19

Consider the following sequences: Q9T4B2 (CYB_DELDE, dolphin), P00156 (CYB_HUMAN, human), P00157 (CYB_BOVIN, cattle), Q9MIX8 (CYB_DANRE, zebrafish), P18946 (CYB_CHICK, chicken), Q36461 (CYB_ORNAN, platypus) and P00160 (CYB_XENLA, frog).

- a. Do you think that a phylogenetic tree built using these sequences would be coherent with the evolution?
- b. Check with the taxonomy provided by UniProt (<http://www.uniprot.org/taxonomy/>).



Human
(*Homo sapiens*)



Dolphin
(*Delphinus delphis*)



Cattle
(*Bos taurus*)



Chicken (rooster)
(*Gallus gallus*)



Zebrafish
(*Danio rerio*)



Platypus
(*Ornithorhynchus anatinus*)



Frog
(*Xenopus laevis*)

19



UniProt

BLAST Align Retrieve/ID mapping

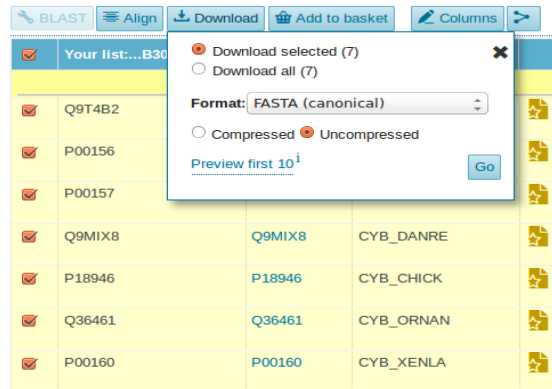
Retrieve/ID mapping

How to use this tool

Enter or upload a list of identifiers to do one of the following:
Retrieve the corresponding UniProt entries to download
Convert identifiers which are of a different type to UniProt

1. Provide your identifiers

Q9T4B2
P00156
P00157
Q9MIX8
P18946
Q36461
P00160



BLAST Align Download Add to basket Columns

Your list: B30

Download selected (7)
Download all (7)

Format: FASTA (canonical)
Compressed Uncompressed

Preview first 10ⁱ

Go

Identifier	Accession	Protein Name
Q9T4B2	Q9T4B2	CYB_DANRE
P00156	P00156	CYB_CHICK
P00157	P00157	CYB_ORNAN
Q9MIX8	Q9MIX8	CYB_XENLA
P18946	P18946	CYB_XENLA
Q36461	Q36461	CYB_XENLA
P00160	P00160	CYB_XENLA



ClustalW is a widely used sequence alignment tool.

Paste your sequence into the window:
(7 input formats are accepted: FASTA, NBRF/PIR, EMBL/Swiss-Prot, GDE, Clustal, GCG/MSF, RSF)

```
>sp|Q9T4B2|CYB_DELDE Cytochrome b 0S=Delphinus delphis GN=MT:CYB PE=3 SV=1
MTNIRKTHPLNKLNDADFIDLPSPNISSWNNF6SLGLCLIMQILTGFLAMHYTPDIS
TAFSSVAHICRDVNYGMFIYRLHANGASMEFICLYAHIGRGLYGSYMFQEDWNIQVLL
LTVMATAFVGTVLPWGQMSFWGATVITNLLSAIPYIGTTLVWINGGFSVDKATLIRFEA
EHFILPFIITAAALVHLLFLHETGSMNPLGITSHSKITFHPYCTIKDALGLLFLLSLM
ALTLFTPOLLGDPNYTPANPLSTPAHIKPEWYELFAYAILBSIPNKLGGVLAALLSILI
LIFIPMLQTSKORSMMFSPFSOLLEWTLADLLTLTWIGGQPVENPYIYGLASILYEL
LTLVLMPTAGLITENKLLKW
>sp|P00156|CYB_HUMAN Cytochrome b 0S=Homo sapiens GN=MT:CYB PE=1 SV=2
MTNIRKTHPLNKLINHSFIDLPSPNISAWNNF6SLGLCLIMQILTGFLAMHYSPDAS
TAFSSIAHITRDVNYGMFIYRLHANGASMEFICLYAHIGRGLYGSFLYSEDNIGILL
LATNATAFNGTVLPWGQMSFWGATVITNLLSAIPYIGTDLVWINGGYSVDSPILIRFEI
EHFILPFIITAAALVHLLFLHETGSMNPLGITSHSKITFHPYCTIKDALGLLFLLSLM
```

File ☐ Pasted ☒ Examinar... No se ha seleccionado ningún archivo.

Align sequences Reset Clear

Dolphin → Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > **Euteleostomi** > **Mammalia** > **Eutheria** > Laurasiatheria > **Cetartiodactyla** > Cetacea > Odontoceti > Delphinidae > Delphinus

Cattle → Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > **Euteleostomi** > **Mammalia** > **Eutheria** > Laurasiatheria > **Cetartiodactyla** > Ruminantia > Pecora > Bovidae > Bovinae > Bos

Human → Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > **Euteleostomi** > **Mammalia** > **Eutheria** > Euarchontoglires > Primates > Haplorrhini > Catarrhini > Hominidae > Homo

Platypus → Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > **Euteleostomi** > **Mammalia** > Monotremata > Ornithorhynchidae > Ornithorhynchus

Frog → Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > **Euteleostomi** > Amphibia > Batrachia > Anura > Pipioidea > Pipidae > Xenopodinae > Xenopus > Xenopus

Chicken → Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > **Euteleostomi** > Archelosauria > Archosauria > Dinosauria > Saurischia > Theropoda > Coelurosauria > Aves > ... > Gallus

Zebrafish → Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > **Euteleostomi** > Actinopterygii > Neopterygii > Teleostei > Ostariophysi > Cypriniformes > Cyprinidae > Danio

