Master Biomedizin 2020

1) UniProt
2) Homology
3) MSA
4) Phylogeny

Pablo Mier
munoz@uni-mainz.de

UniProt database

1

a. What is the AC of the UniProt entry for the human insulin? P01308

b. How many isoforms for this protein are described in that entry? 2 isoforms

c. How many times has this entry been modified? 251 times; currently in version 252
… and the protein sequence? None; currently in version 1

d. With how many proteins does the human insulin interact? 14 interactors (BioGrid), 17 interactors (IntAct); databases do not always agree

Protein-protein interaction databases

BioGrid [i]	109842, 14 interactors
DIP [i]	DIP-6024N
IntAct [i]	P01308, 17 interactors
MINT [i]	P01308
STRING [i]	9606.ENSP00000380432

Pablo Mier
munoz@uni-mainz.de

Homology

2

Classify the following protein pairs based on their evolutionary relationship.
Note: proteins A and B have a common ancestor.

a. Protein A mouse / Protein A human → Orthologs

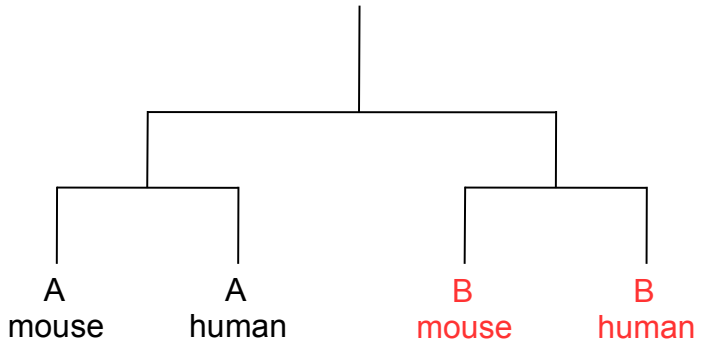
b. Protein A mouse / Protein B mouse → Paralogs

c. Protein A mouse / Protein B human → Homologs

d. Protein A human / Protein B mouse → Homologs

e. Protein A human / Protein B human → Paralogs

f. Protein B mouse / Protein B human → Orthologs



Pablo Mier
munoz@uni-mainz.de

Homology

*Images from: Wikipedia

3

a. Using the human protein “P21741”, find its orthologous proteins in frog (*Xenopus laevis*) and get their UniProt AC. P48530, P48531

b. Check the identity between the orthologs (human – frog proteins).
P21741-P48530 = 61.1%, P21741-P48531 = 60.4%

c. Check the identity between the paralogs (frog – frog proteins).
P48530-P48531 = 97.9%

Human
(*Homo sapiens*)

Frog
(*Xenopus laevis*)

Pablo Mier
munoz@uni-mainz.de

*Images from: Wikipedia

4

a. Based on the sequence of the “ATP synthase subunit a” protein from the extinct mammoth (*Mammuthus primigenius*) [Q38PR7], was the mammoth closer to the asian elephant (*Elephas maximus*) or to the african elefant (*Loxodonta africana*)? Use only SwissProt proteins.

M. primigenius (Q38PR7) – *E. maximus* (Q2I3G9) = 95.5%
M. primigenius (Q38PR7) – *L. africana* (Q9TA24) = 93.2%

b. Is there evidence enough to conclude if they are / are not closer? No.

c. Could you check with the “cytochrome b” protein too? [P92658] Use only SwissProt proteins.

M. primigenius (P92658) – *E. maximus* (O47885) = 96.3%
M. primigenius (P92658) – *L. africana* (P24958) = 97.9%

Woolly mammoth
(*Mammuthus primigenius*)

Asian elephant
(*Elephas maximus*)

African elephant
(*Loxodonta africana*)

Pablo Mier
munoz@uni-mainz.de

*Images from: Wikipedia

5

a. Based solely on the sequence of the "Cytochrome b" protein (Q8SG72) from the extinct dodo (*Raphus cucullatus*), was the dodo closer to the Nicobar pigeon "*Caloenas nicobarica*" or to the chicken (*Gallus gallus*)? Use NCBI Blast.

R. cucullatus – C. nicobarica = 98.50%
R. cucullatus – G. gallus = 92.51%

b. There are more than 300 species of pigeons. Do the results differ if you consider the street pigeon (*Columba livia*)?

R. cucullatus – C. livia = 95.88%
R. cucullatus – G. gallus = 92.51%

Dodo
(*Raphus cucullatus*)

Nicobar pigeon
(*Caloenas nicobarica*)

Chicken (rooster)
(*Gallus gallus*)

Pigeon
(*Columba livia*)

Pablo Mier
munoz@uni-mainz.de

6

a. The UniProt entry “P04585” contains the Gag-Pol polyprotein from the virus HV1H2. Do you think it would resemble any protein in the human proteome (*Homo sapiens*)? Check it using NCBI Blast.

Many retroviral proteins integrated in the human genome.

b. The Gag-Pol polyprotein is composed of many proteins. Using only protein entries from *Chlamydia trachomatis*, can you identify some of the individual proteins of the Gag-Pol polyprotein?

Database: protein entries from “*Chlamydia trachomatis*”.

- Gag protein p24
- Reverse transcriptase
- Integrase

Pablo Mier
munoz@uni-mainz.de

7

Given the following alignments,

classify them in:

- Pairwise / multiple
- Local / global

calculate their:

- % similarity
- % identity

```
>Protein_A
KKKYYWWKKT
>Protein_B
AKKYYWW
>Protein_C
RKRWWWWRT
```

a)
```
Protein_A     YYWW
Protein_B     YYWW
              ****
```

Pairwise
Local
100% similarity
100% identity

b)
```
Protein_A     KKKYYWWKKT
Protein_B     AKKYYWW---
               ******
```

Pairwise
Global
60% similarity
60% identity

c)
```
Protein_A     KKKYYWWKKT
Protein_B     AKKYYWW---
Protein_C     AKRWWWWR-T
               *:::**
```

Multiple
Global
60% similarity
30% identity

Pablo Mier
munoz@uni-mainz.de

8

a. The proteins “P11582” and “P02226” are paralogs, but they differ in length (152 vs 161 amino acids). Is there an extra region in P02226, or the extra amino acids are dispersed along the protein? Use UniProt.
Extra region in P02226: “IGNESN”.

b. How could that have happened in evolution?
Deletion in P11582, or insertion in P02226.

c. If you blast P02226 and use only SwissProt proteins as database, how many results do you get? Is this protein present in any bacterial proteome?
159 results; 15 in Bacteria and 144 in Eukaryota.

// FYI: Protein entry P02226 is one of the oldest in SwissProt; check the date! July 21, 1986.

Pablo Mier
munoz@uni-mainz.de

*Images from: Wikipedia

9

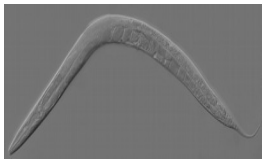
a. Both “P17861” (XBP1_HUMAN) and “Q3SZZ2” (XBP1_BOVIN) are “X-box binding protein 1” proteins. Can you detect which region/s of these proteins is/are important for their function? Why? Use Clustal Omega.
What should you do to detect them? No. They are too similar. We would need a protein from a more distant organism.

b. Add the proteins “G5EE07” (G5EE07_CAEEL) and “Q8UVQ5” (Q8UVQ5_DANRE) to the study. Are you able to identify that region/s now? Why? Use Clustal Omega.
Yes. They are not as similar.

c. Check the positional annotations in the entry of the human protein. Was the region you identified annotated as a domain?
bZIP (basic-leucine zipper) domain in positions:
70-133 (human)
70-133 (cattle)
61-117 (worm)
69-132 (zebrafish)

Human
(*Homo sapiens*)

Cattle
(*Bos taurus*)



Worm
(*Caenorhabditis elegans*)

Zebra fish
(*Danio rerio*)

Pablo Mier
munoz@uni-mainz.de

10

a. Using the following set of orthologous proteins, predict which regions are important for their function. Use T-Coffee.

Q90WY9
P97801
Q98SU9
Q9W6S8
W4XFQ6
A0A088A467

Boxed regions.

b. Check your predictions in their UniProt entries.

88-148	Tudor
10-41	P1
92-204	Required for interaction with RPP20/POP7
235-262	P2
274-288	Required for interaction with SYNCRIP

Pablo Mier
munoz@uni-mainz.de

11

a. Using the following set of orthologous proteins, do you think the evolution is pressuring them? Use UniProt.

Q02078
Q5REW7
Q2MJT0
Q60929
A2VDZ3
A2ICN5
Q9W6U8

Yes, they conserve most of the sequence but have some differences.

b. Have they evolved? Did they gain or lose any domain/motif/region?

Differences in:
- Beta domain (LCR Glu, E)
- Glutamines. Maximum Q stretch:
  - Chicken = 3 Q
  - Pig = 4 Q
  - Cattle = 5 Q
  - Mouse = 6 Q
  - Rat = 7 Q
  - Orangutan = 9 Q
  - Human = 11 Q

Pablo Mier
munoz@uni-mainz.de

12

A patient comes to the hospital. He was just bitten by a snake. We have the sequence of the mitochondrial gene ND4 of 24 species of snake ("*snakes.fasta*"; https://cbdm.uni-mainz.de/mb20/). We have three antidotes available. Given the following information, which antidote would you administer the patient?

1) The snake that bit the patient is in terrarium #1.
2) The most distant snake species is in terrarium #12.
3) Antidote1 is indicated against bites from the species in terrarium #3.
4) Antidote2 is indicated against bites from the species in terrarium #11.
5) Antidote3 is indicated against bites from the species in terrarium #17.
6) Snakes in terrariums #15 and #20 are non-venomous.

No antidote! The snake seems not venomous

Pablo Mier
munoz@uni-mainz.de

13

a. All of the sequences in "*file1.fasta*" (https://cbdm.uni-mainz.de/mb20/) are homologs. How many groups of orthologs would you say there are in this file? Use Trex (http://www.trex.uqam.ca/).
Two groups of orthologs: Protein A & protein B.

b. What could you say about the history of this protein family?
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. Would you say there is any wrongly annotated sequence?
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 either not in the dataset or was lost during evolution.

Pablo Mier
munoz@uni-mainz.de

14

a. Using "*file2.fasta*" (https://cbdm.uni-mainz.de/mb20/), can you approximate to which taxonomic division belongs "proteinX"? Primates.

b. From which organism could it be? After guessing, check it.
Homo sapiens (human) or *Pan troglodytes* (chimpanzee); they are 100% identical.

Pablo Mier
munoz@uni-mainz.de

15

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, mu, alpha, theta1, epsilon, gamma1, gamma2, delta and beta. Use the proteins in "*file3.fasta*" (https://cbdm.uni-mainz.de/mb20/), which includes an hemoglobin protein from *Gallus gallus* to be used as outgroup.

a. Sort them either in cluster alpha or cluster beta.

Alpha: zeta, mu, alpha, theta1.
Beta: epsilon, gamma1, gamma2, delta and beta.

b. Why do you think they are clustered in either cluster alpha or cluster beta?

Paralogous expansion from one ancestral alpha and one ancestral beta.

Pablo Mier
munoz@uni-mainz.de

16

The sequences in "*file4.fasta*" (https://cbdm.uni-mainz.de/mb20/) are orthologs from dolphin, human, cattle, zebra fish, chicken, platypus and frog.

a. Build a phylogenetic tree using these sequences.

b. Check in UniProt > Taxonomy (https://www.uniprot.org/taxonomy/) if the phylogenetic tree built only from that one protein family is coherent with evolution.

Dolphin = Delphinus delphis
Human = Homo sapiens
Cattle = Bos taurus
Zebra fish = Danio rerio
Chicken = Gallus gallus
Platypus = Ornithorhynchus anatinus
Frog = Xenopus laevis

No, the tree fails to relate all amniotes in one branch.

Pablo Mier
munoz@uni-mainz.de