Master Biomedizin 2022

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?

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

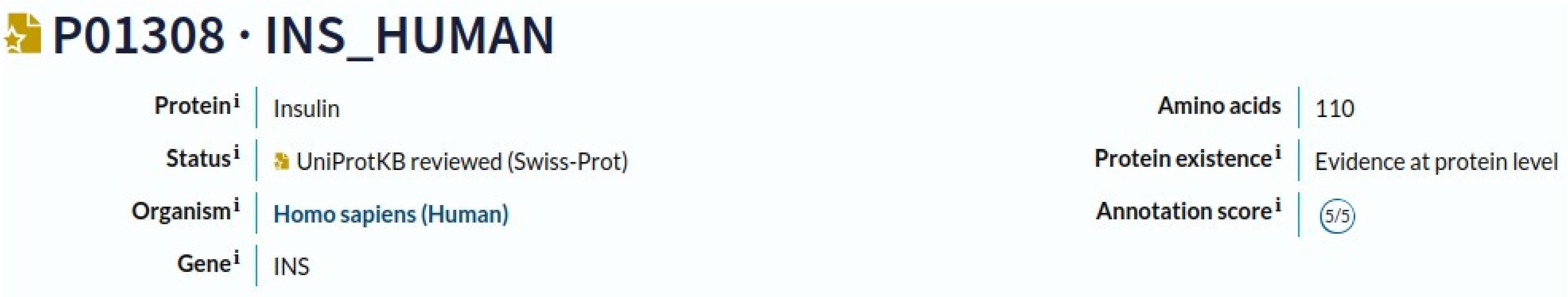
c. How many times has this entry been modified?
… and the protein sequence?

d. With how many proteins does the human insulin interact?

UniProt database

1

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



P01308 · INS_HUMAN

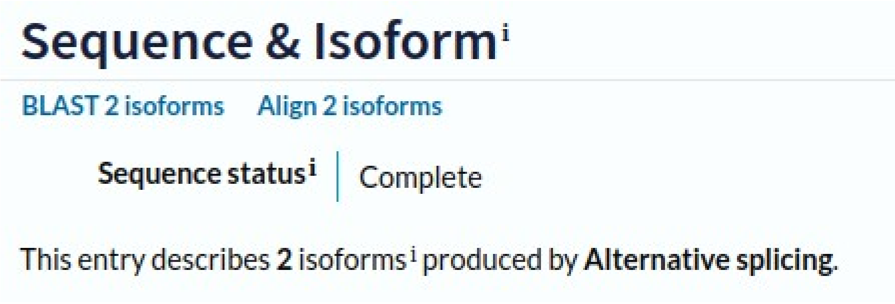
Protein	Insulin	Amino acids	110
Status	UniProtKB reviewed (Swiss-Prot)	Protein existence	Evidence at protein level
Organism	Homo sapiens (Human)	Annotation score	5/5
Gene	INS		

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



Sequence & Isoform i

Sequence status i | Complete

This entry describes **2** isoforms i produced by **Alternative splicing**.

UniProt database

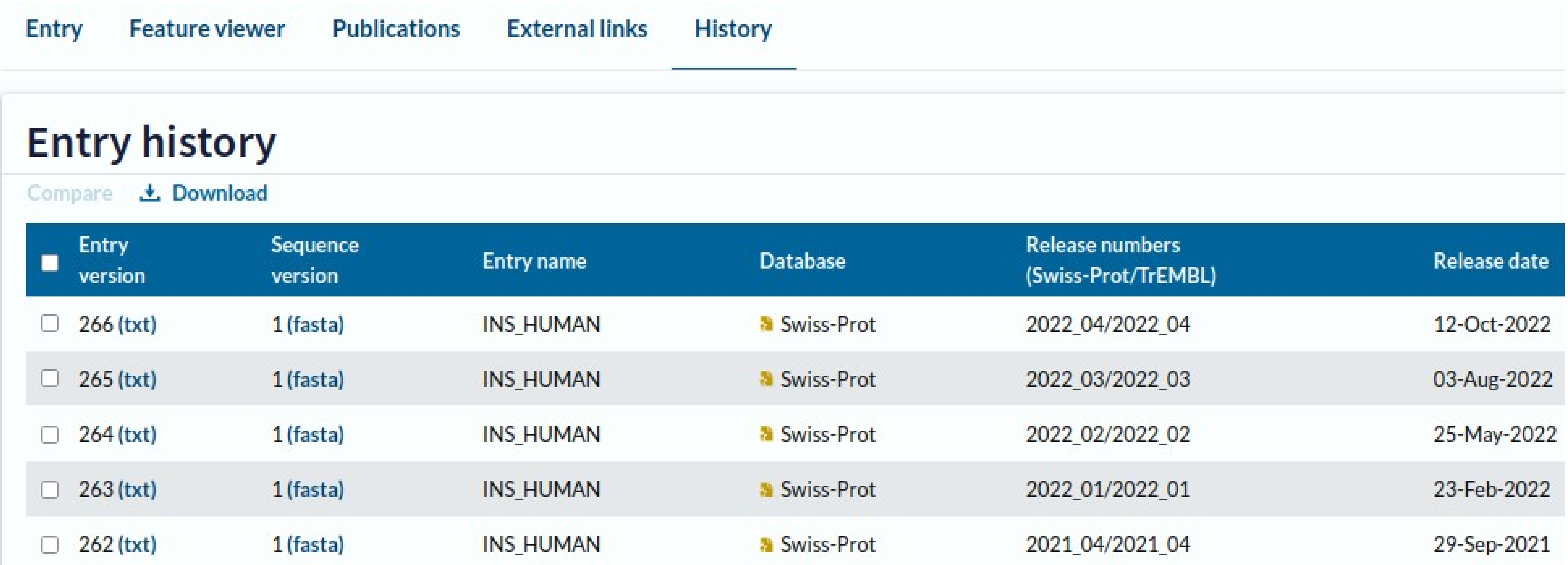
1

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? 265 times; currently in version 266

… and the protein sequence? None; currently in version 1



Entry history

□	Entry version	Sequence version	Entry name	Database	Release numbers (Swiss-Prot/TrEMBL)	Release date
□	266 (txt)	1 (fasta)	INS_HUMAN	Swiss-Prot	2022_04/2022_04	12-Oct-2022
□	265 (txt)	1 (fasta)	INS_HUMAN	Swiss-Prot	2022_03/2022_03	03-Aug-2022
□	264 (txt)	1 (fasta)	INS_HUMAN	Swiss-Prot	2022_02/2022_02	25-May-2022
□	263 (txt)	1 (fasta)	INS_HUMAN	Swiss-Prot	2022_01/2022_01	23-Feb-2022
□	262 (txt)	1 (fasta)	INS_HUMAN	Swiss-Prot	2021_04/2021_04	29-Sep-2021

UniProt database

1

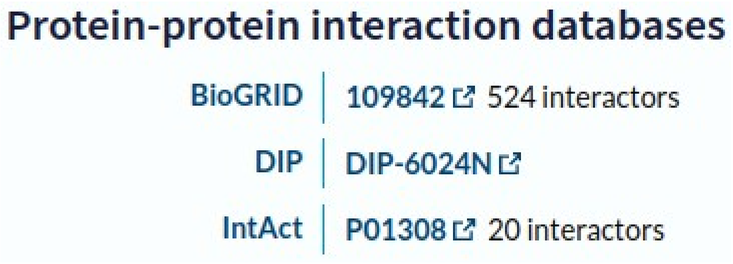
a. What is the AC of the UniProt entry for the human insulin? P01308

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

c. How many times has this entry been modified? 265 times; currently in version 266
… and the protein sequence? None; currently in version 1

d. With how many proteins does the human insulin interact? 524 interactors (BioGRID), 20 interactors (IntAct); databases do not always agree

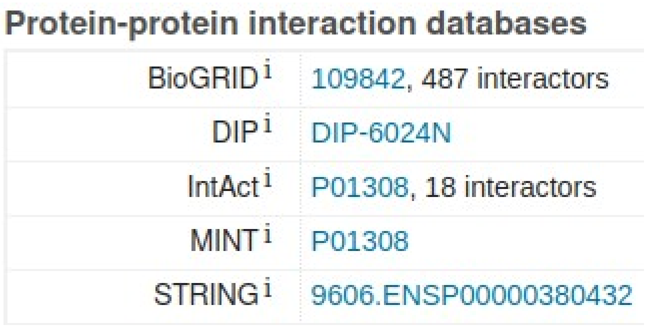
Nov 2022



Protein-protein interaction databases

BioGRID	109842 524 interactors
DIP	DIP-6024N
IntAct	P01308 20 interactors

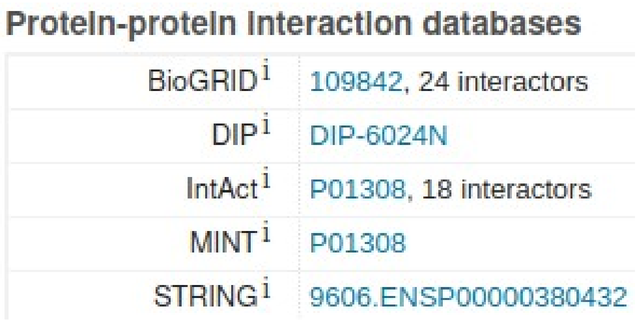
Feb 2022



Protein-protein interaction databases

BioGRID i	109842, 487 interactors
DIP i	DIP-6024N
IntAct i	P01308, 18 interactors
MINT i	P01308
STRING i	9606.ENSP00000380432

Feb 2021



Protein-protein interaction databases

BioGRID i	109842, 24 interactors
DIP i	DIP-6024N
IntAct i	P01308, 18 interactors
MINT i	P01308
STRING i	9606.ENSP00000380432

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

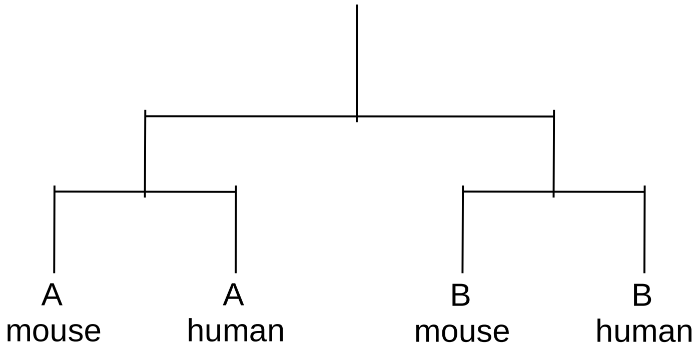
b. Protein A mouse / Protein B mouse

c. Protein A mouse / Protein B human

d. Protein A human / Protein B mouse

e. Protein A human / Protein B human

f. Protein B mouse / Protein B human



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

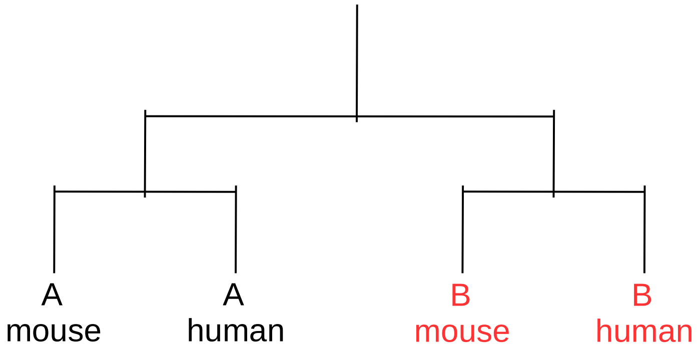
b. Protein A mouse / Protein B mouse → Paralogs

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



Homology

*Images from: Wikipedia

3

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

b. Check the identity between the orthologs (human – frog proteins).

c. Check the identity between the paralogs (frog – frog proteins).

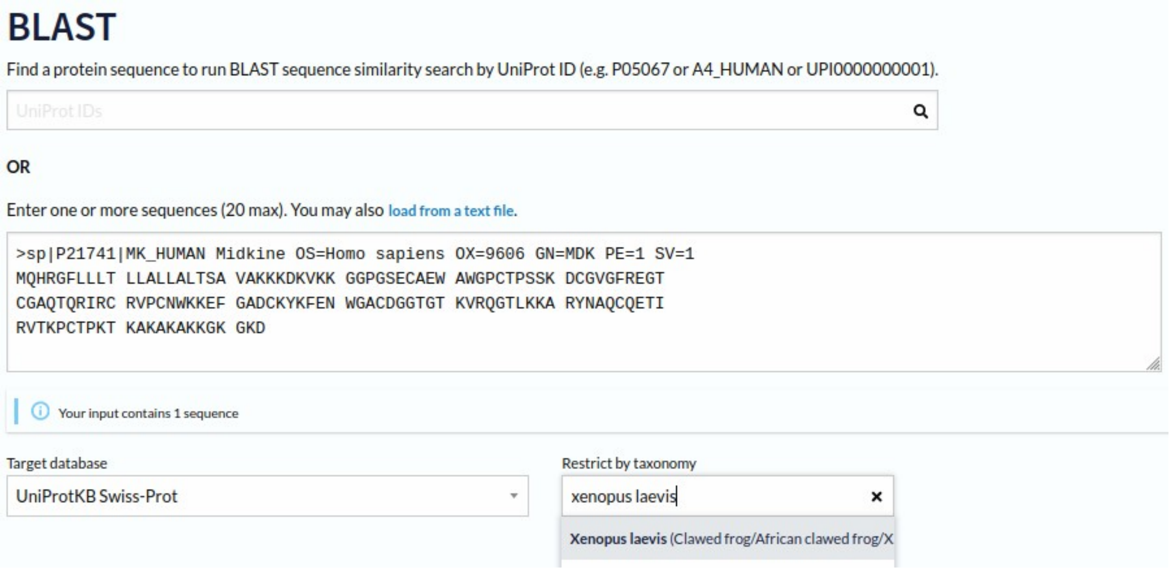
Human
(*Homo sapiens*)

Frog
(*Xenopus laevis*)

Homology

*Images from: UniProt

3



BLAST

Find a protein sequence to run BLAST sequence similarity search by UniProt ID (e.g. P05067 or A4_HUMAN or UPI0000000001).

UniProt IDs

OR

Enter one or more sequences (20 max). You may also load from a text file.

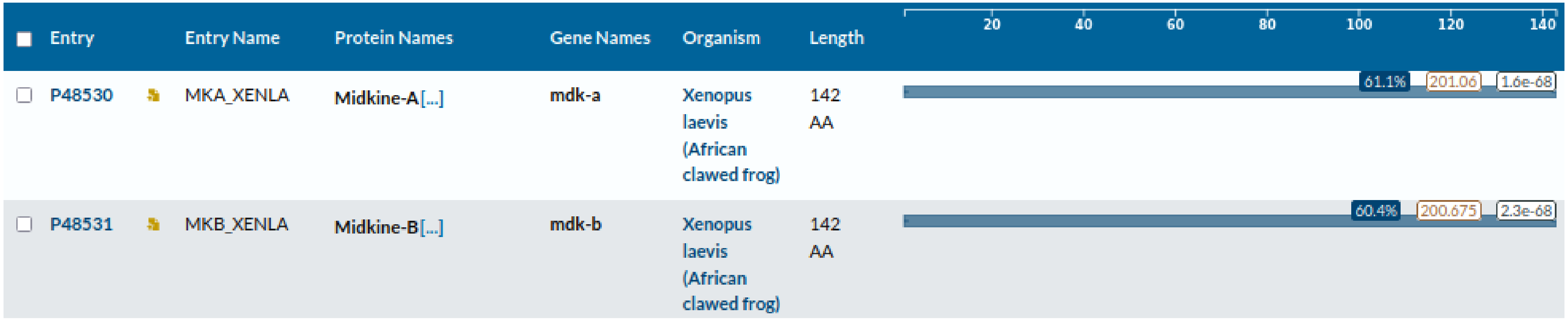
```
>sp|P21741|MK_HUMAN Midkine OS=Homo sapiens OX=9606 GN=MDK PE=1 SV=1
MQHRGFLLLT LLALLALTSA VAKKKDKVKK GGPGSECAEW AWGPCTPSSK DCGVGFREGT
CGAQTQRIRC RVPCNWKKEF GADCKYKFEN WGACDGGTGT KVRQGTLKKA RYNAQCQETI
RVTKPCTPKT KAKAKAKKGK GKD
```

ⓘ Your input contains 1 sequence

Target database: UniProtKB Swiss-Prot

Restrict by taxonomy: xenopus laevis

Xenopus laevis (Clawed frog/African clawed frog/X



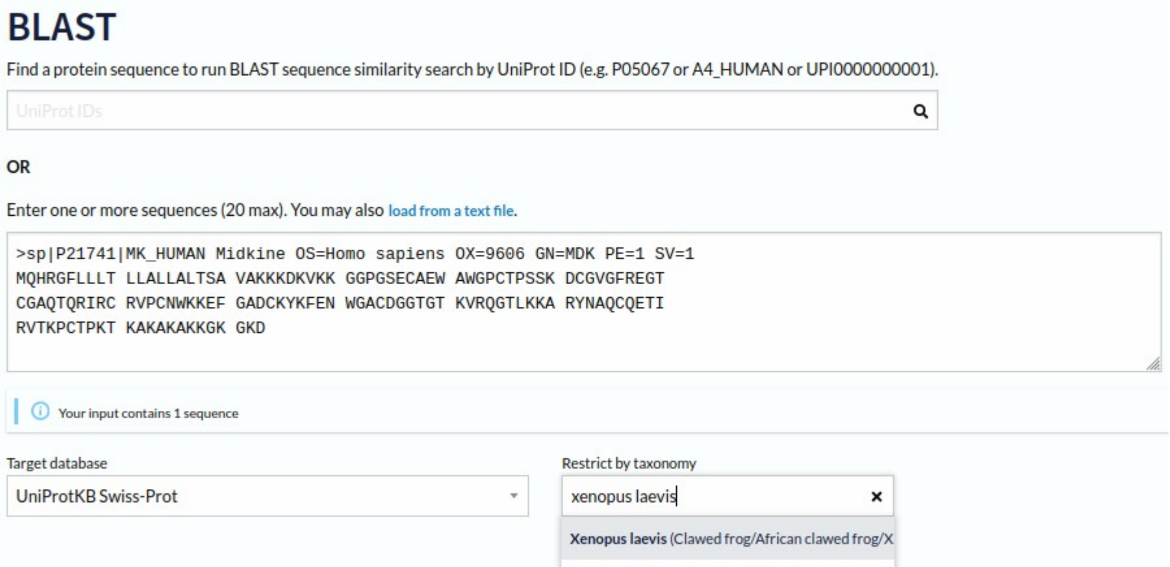
□	Entry	Entry Name	Protein Names	Gene Names	Organism	Length	20 40 60 80 100 120 140
□	P48530	MKA_XENLA	Midkine-A[...]	mdk-a	Xenopus laevis (African clawed frog)	142 AA	61.1% 201.06 1.6e-68
□	P48531	MKB_XENLA	Midkine-B[...]	mdk-b	Xenopus laevis (African clawed frog)	142 AA	60.4% 200.675 2.3e-68

a. Query: P21741.
Ortholog1: P48530.
Ortholog2: P48531.

Homology

*Images from: UniProt

3



BLAST

Find a protein sequence to run BLAST sequence similarity search by UniProt ID (e.g. P05067 or A4_HUMAN or UPI0000000001).

OR

Enter one or more sequences (20 max). You may also load from a text file.

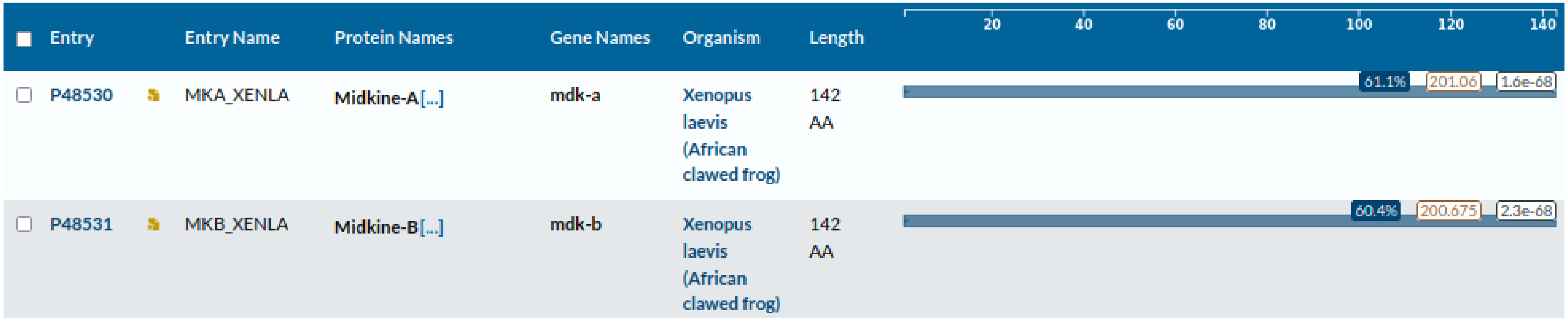
```
>sp|P21741|MK_HUMAN Midkine OS=Homo sapiens OX=9606 GN=MDK PE=1 SV=1
MQHRGFLLLT LLALLALTSA VAKKKDKVKK GGPGSECAEW AWGPCTPSSK DCGVGFREGT
CGAQTQRIRC RVPCNWKKEF GADCKYKFEN WGACDGGTGT KVRQGTLKKA RYNAQCQETI
RVTKPCTPKT KAKAKAKKGK GKD
```

Your input contains 1 sequence

Target database: UniProtKB Swiss-Prot

Restrict by taxonomy: xenopus laevis

Xenopus laevis (Clawed frog/African clawed frog/X



□ Entry		Entry Name	Protein Names	Gene Names	Organism	Length	20 40 60 80 100 120 140
□ P48530		MKA_XENLA	Midkine-A[...]	mdk-a	Xenopus laevis (African clawed frog)	142 AA	61.1% 201.06 1.6e-68
□ P48531		MKB_XENLA	Midkine-B[...]	mdk-b	Xenopus laevis (African clawed frog)	142 AA	60.4% 200.675 2.3e-68

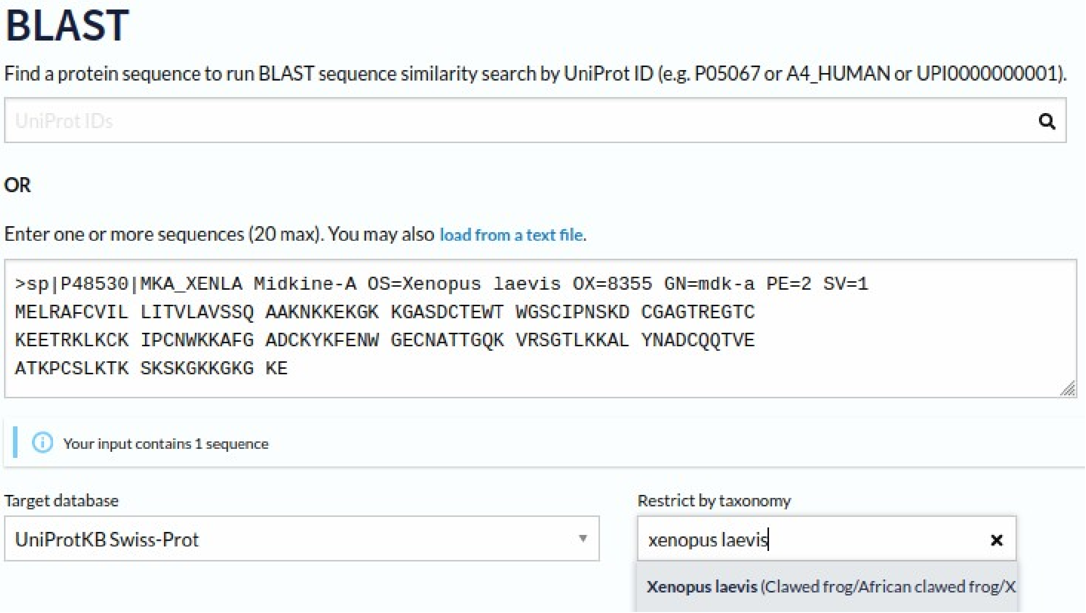
a. Query: P21741.
Ortholog1: P48530.
Ortholog2: P48531.

b. P21741-P48530 = 61.1%
P21741-P48531 = 60.4%

Homology

*Images from: UniProt

3



BLAST

Find a protein sequence to run BLAST sequence similarity search by UniProt ID (e.g. P05067 or A4_HUMAN or UPI0000000001).

OR

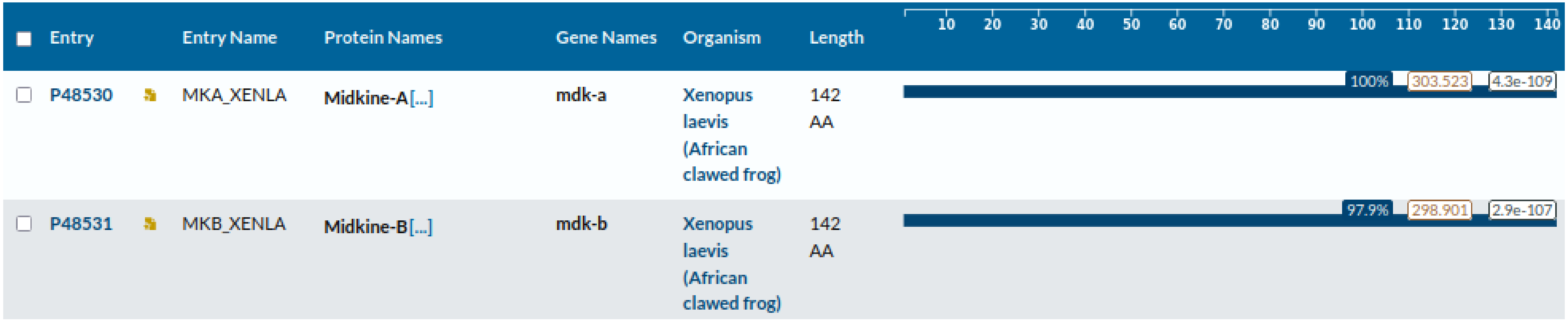
Enter one or more sequences (20 max). You may also load from a text file.

```
>sp|P48530|MKA_XENLA Midkine-A OS=Xenopus laevis OX=8355 GN=mdk-a PE=2 SV=1
MELRAFCVIL LITVLAVSSQ AAKNKKEKGK KGASDCTEWT WGSCIPNSKD CGAGTREGTC
KEETRKLKCK IPCNWKKAFG ADCKYKFENW GECNATTGQK VRSGTLKKAL YNADCQQTVE
ATKPCSLKTK SKSKGKKGKG KE
```

Your input contains 1 sequence

Target database: UniProtKB Swiss-Prot

Restrict by taxonomy: xenopus laevis



□	Entry		Entry Name	Protein Names	Gene Names	Organism	Length	10 20 30 40 50 60 70 80 90 100 110 120 130 140
□	P48530		MKA_XENLA	Midkine-A […]	mdk-a	Xenopus laevis (African clawed frog)	142 AA	100% 303.523 4.3e-109
□	P48531		MKB_XENLA	Midkine-B […]	mdk-b	Xenopus laevis (African clawed frog)	142 AA	97.9% 298.901 2.9e-107

a. Query: P21741.
Ortholog1: P48530.
Ortholog2: P48531.

b. P21741-P48530 = 61.1%
P21741-P48531 = 60.4%

c. P48530-P48531 = 97.9%
Note: may also be done with “alignments”.

Homology

*Images from: Wikipedia

4

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

b. Is there evidence enough to conclude if they are / are not closer?

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

Woolly mammoth
(*Mammuthus primigenius*)

Asian elephant
(*Elephas maximus*)

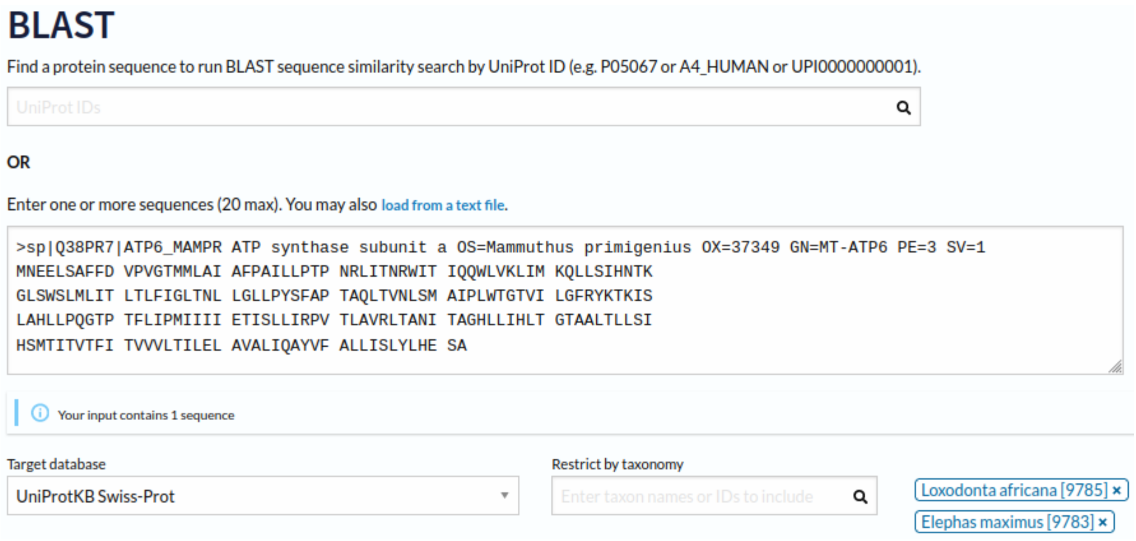
African elephant
(*Loxodonta africana*)

Homology

*Images from: UniProt

4

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



BLAST

Find a protein sequence to run BLAST sequence similarity search by UniProt ID (e.g. P05067 or A4_HUMAN or UPI0000000001).

UniProt IDs

OR

Enter one or more sequences (20 max). You may also load from a text file.

```
>sp|Q38PR7|ATP6_MAMPR ATP synthase subunit a OS=Mammuthus primigenius OX=37349 GN=MT-ATP6 PE=3 SV=1
MNEELSAFFD VPVGTMMLAI AFPAILLPTP NRLITNRWIT IQQWLVKLIM KQLLSIHNTK
GLSWSLMLIT LTLFIGLTNL LGLLPYSFAP TAQLTVNLSM AIPLWTGTVI LGFRYKTKIS
LAHLLPQGTP TFLIPMIIII ETISLLIRPV TLAVRLTANI TAGHLLIHLT GTAALTLLSI
HSMTITVTFI TVVVLTILEL AVALIQAYVF ALLISLYLHE SA
```

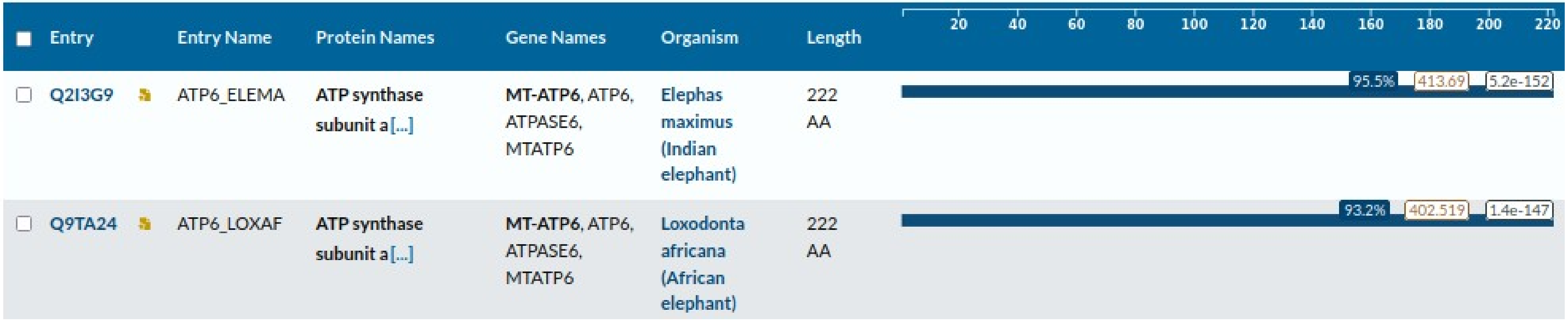
Your input contains 1 sequence

Target database: UniProtKB Swiss-Prot

Restrict by taxonomy: Enter taxon names or IDs to include

Loxodonta africana [9785] ×

Elephas maximus [9783] ×



☐	Entry	Entry Name	Protein Names	Gene Names	Organism	Length	20 40 60 80 100 120 140 160 180 200 220
☐	Q2I3G9	ATP6_ELEMA	ATP synthase subunit a [...]	MT-ATP6, ATP6, ATPASE6, MTATP6	Elephas maximus (Indian elephant)	222 AA	95.5% 413.69 5.2e-152
☐	Q9TA24	ATP6_LOXAF	ATP synthase subunit a [...]	MT-ATP6, ATP6, ATPASE6, MTATP6	Loxodonta africana (African elephant)	222 AA	93.2% 402.519 1.4e-147

Homology

*Images from: UniProt

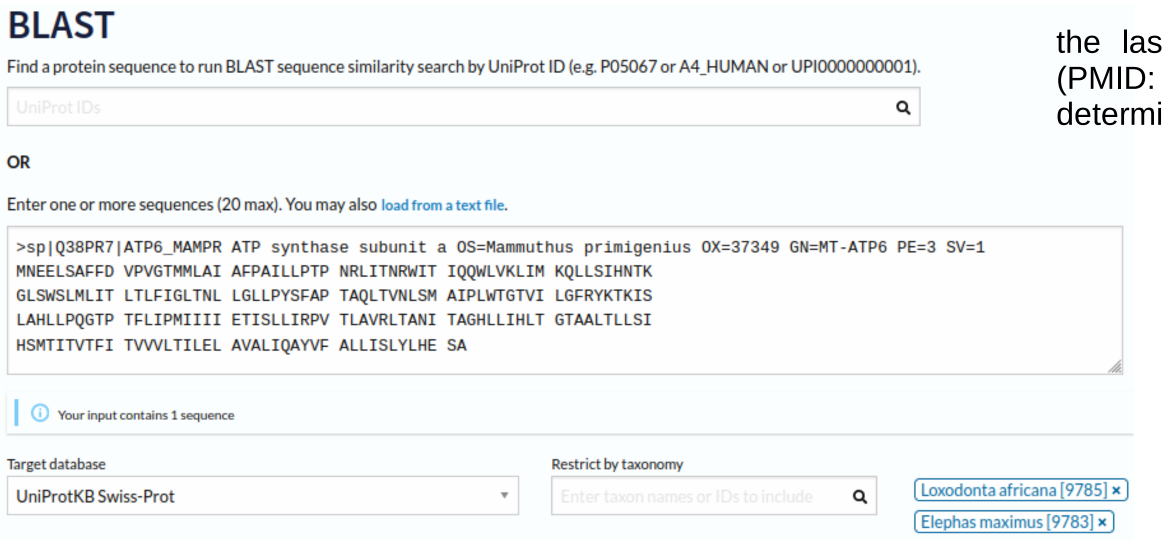
4

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

b. Just this sequence similarity is not evidence enough for claiming the mammoth is closer to the asian elephant than to the african elephant,

BUT

the last genome sequencing work on the woolly mammoth (PMID: 19020620), in 2008, provides evidence enough to determine that it is really closer to the asian elephant.



BLAST

Find a protein sequence to run BLAST sequence similarity search by UniProt ID (e.g. P05067 or A4_HUMAN or UPI0000000001).

OR

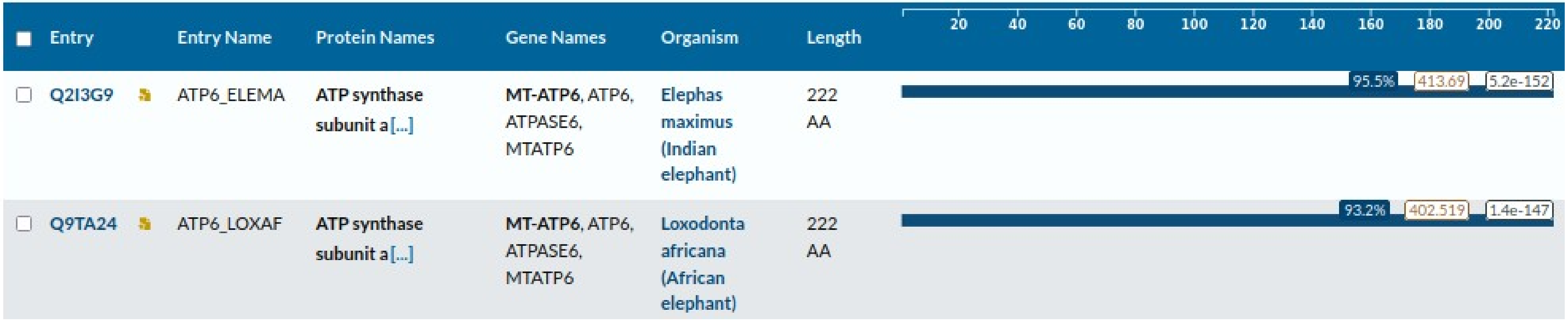
Enter one or more sequences (20 max). You may also load from a text file.

```
>sp|Q38PR7|ATP6_MAMPR ATP synthase subunit a OS=Mammuthus primigenius OX=37349 GN=MT-ATP6 PE=3 SV=1
MNEELSAFFD VPVGTMMLAI AFPAILLPTP NRLITNRWIT IQQWLVKLIM KQLLSIHNTK
GLSWSLMLIT LTLFIGLTNL LGLLPYSFAP TAQLTVNLSM AIPLWTGTVI LGFRYKTKIS
LAHLLPQGTP TFLIPMIIII ETISLLIRPV TLAVRLTANI TAGHLLIHLT GTAALTLLSI
HSMTITVTFI TVVVLTILEL AVALIQAYVF ALLISLYLHE SA
```

ⓘ Your input contains 1 sequence

Target database: UniProtKB Swiss-Prot

Restrict by taxonomy: Loxodonta africana [9785] ×, Elephas maximus [9783] ×



□	Entry	Entry Name	Protein Names	Gene Names	Organism	Length	20 40 60 80 100 120 140 160 180 200 220
□	Q2I3G9	ATP6_ELEMA	ATP synthase subunit a [...]	MT-ATP6, ATP6, ATPASE6, MTATP6	Elephas maximus (Indian elephant)	222 AA	95.5% 413.69 5.2e-152
□	Q9TA24	ATP6_LOXAF	ATP synthase subunit a [...]	MT-ATP6, ATP6, ATPASE6, MTATP6	Loxodonta africana (African elephant)	222 AA	93.2% 402.519 1.4e-147

Homology

*Images from: UniProt

4

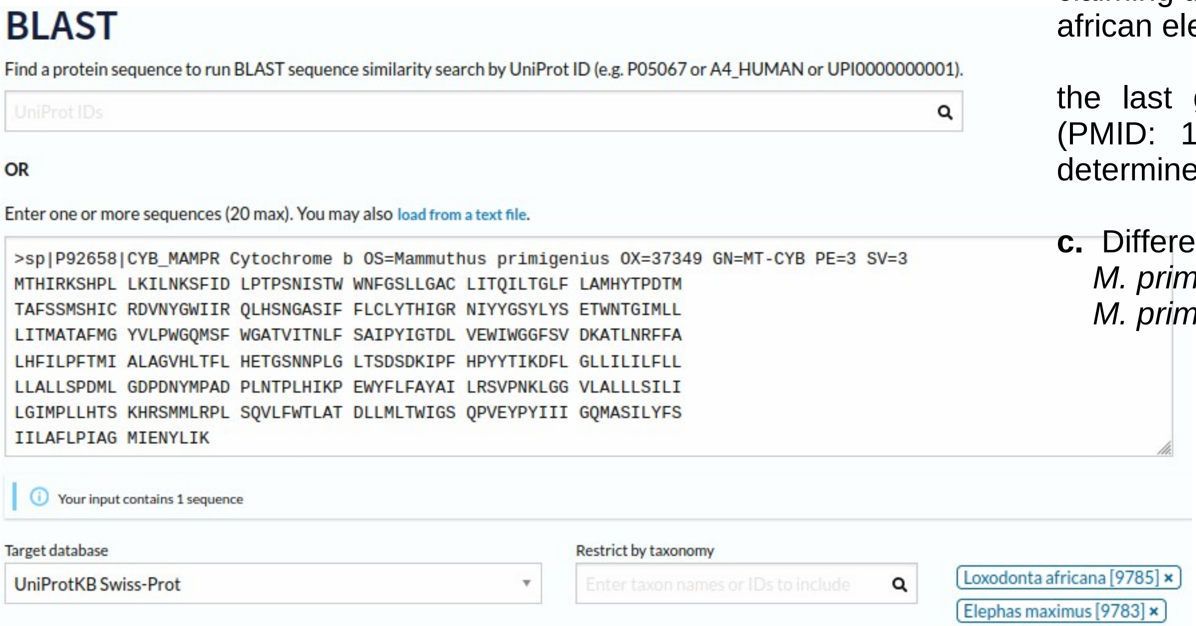
a. *M. primigenius* (Q38PR7) – *E. maximus* (Q2I3G9) = 95.5%
M. primigenius (Q38PR7) – *L. africana* (Q9TA24) = 93.2%

b. Just this sequence similarity is not evidence enough for claiming the mammoth is closer to the asian elephant than to the african elephant,

BUT

the last genome sequencing work on the woolly mammoth (PMID: 19020620), in 2008, provides evidence enough to determine that it is really closer to the asian elephant.

c. Different results! (read “b” again...)
M. primigenius (P92658) – *E. maximus* (O47885) = 96.3%
M. primigenius (P92658) – *L. africana* (P24958) = 97.9%



BLAST

Find a protein sequence to run BLAST sequence similarity search by UniProt ID (e.g. P05067 or A4_HUMAN or UPI0000000001).

UniProt IDs

OR

Enter one or more sequences (20 max). You may also load from a text file.

```
>sp|P92658|CYB_MAMPR Cytochrome b OS=Mammuthus primigenius OX=37349 GN=MT-CYB PE=3 SV=3
MTHIRKSHPL LKILNKSFID LPTPSNISTW WNFGSLLGAC LITQILTGLF LAMHYTPDTM
TAFSSMSHIC RDVNYGWIIR QLHSNGASIF FLCLYTHIGR NIYYGSYLYS ETWNTGIMLL
LITMATAFMG YVLPWGQMSF WGATVITNLF SAIPYIGTDL VEWIWGGFSV DKATLNRFFA
LHFILPFTMI ALAGVHLTFL HETGSNNPLG LTSDSDKIPF HPYYTIKDFL GLLILILFLL
LLALLSPDML GDPDNYMPAD PLNTPLHIKP EWYFLFAYAI LRSVPNKLGG VLALLLSILI
LGIMPLLHTS KHRSMMLRPL SQVLFWTLAT DLLMLTWIGS QPVEYPYIII GQMASILYFS
IILAFLPIAG MIENYLIK
```

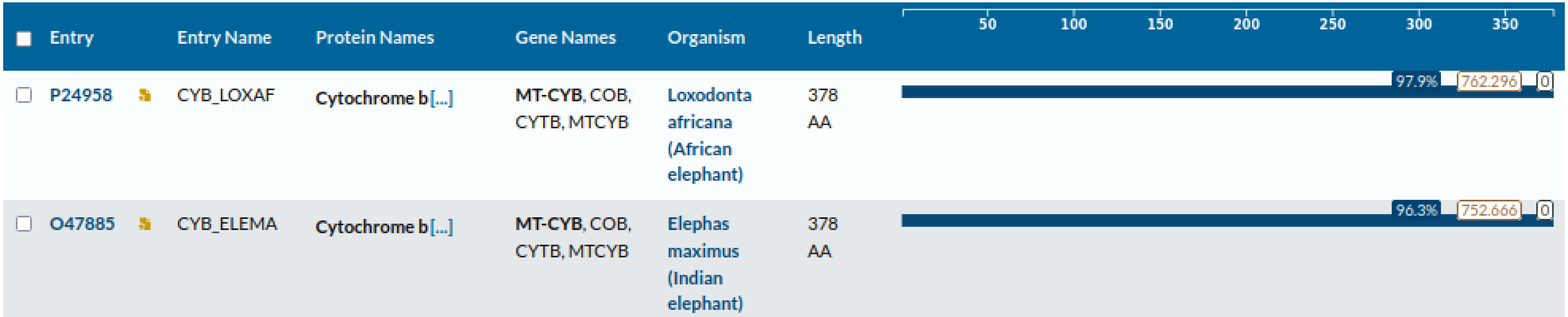
Your input contains 1 sequence

Target database: UniProtKB Swiss-Prot

Restrict by taxonomy: Enter taxon names or IDs to include

Loxodonta africana [9785] ×

Elephas maximus [9783] ×



□	Entry	Entry Name	Protein Names	Gene Names	Organism	Length	
□	P24958	CYB_LOXAF	Cytochrome b [...]	MT-CYB, COB, CYTB, MTCYB	Loxodonta africana (African elephant)	378 AA	97.9% 762.296 0
□	O47885	CYB_ELEMA	Cytochrome b [...]	MT-CYB, COB, CYTB, MTCYB	Elephas maximus (Indian elephant)	378 AA	96.3% 752.666 0

Homology

5

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*)?

b. The Gag-Pol polyprotein is composed of more than one protein. Can you identify them? Use only SwissProt proteins.

5

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*)?

Many retroviral proteins integrated in the human genome.

b. The Gag-Pol polyprotein is composed of more than one protein. Can you identify them? Use only SwissProt proteins.

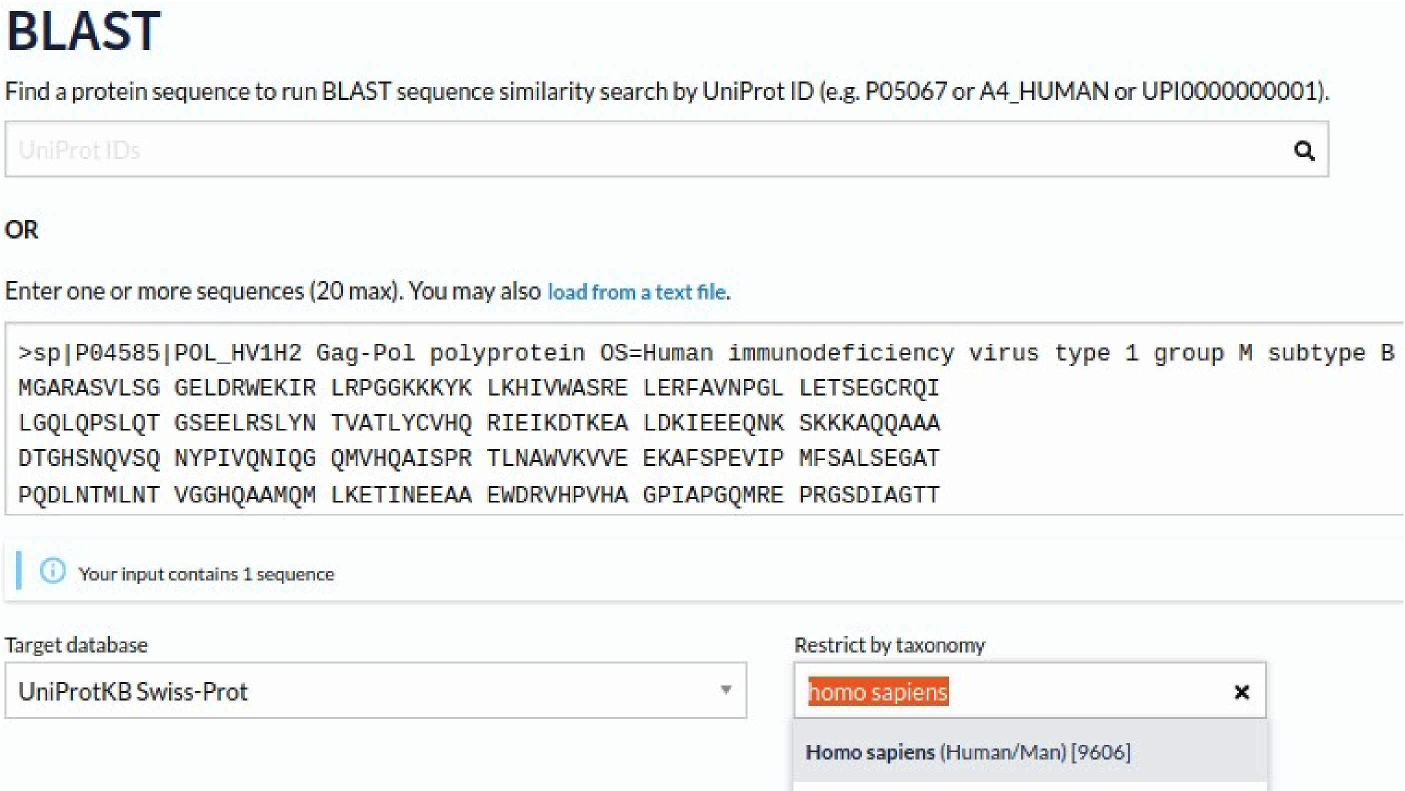
Homology

5

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*)?

Many retroviral proteins integrated in the human genome.

b. The Gag-Pol polyprotein is composed of more than one protein. Can you identify them? Use only SwissProt proteins.



BLAST

Find a protein sequence to run BLAST sequence similarity search by UniProt ID (e.g. P05067 or A4_HUMAN or UPI0000000001).

OR

Enter one or more sequences (20 max). You may also load from a text file.

```
>sp|P04585|POL_HV1H2 Gag-Pol polyprotein OS=Human immunodeficiency virus type 1 group M subtype B
MGARASVLSG GELDRWEKIR LRPGGKKKYK LKHIVWASRE LERFAVNPGL LETSEGCRQI
LGQLQPSLQT GSEELRSLYN TVATLYCVHQ RIEIKDTKEA LDKIEEEQNK SKKKAQQAAA
DTGHSNQVSQ NYPIVQNIQG QMVHQAISPR TLNAWVKVVE EKAFSPEVIP MFSALSEGAT
PQDLNTMLNT VGGHQAAMQM LKETINEEAA EWDRVHPVHA GPIAPGQMRE PRGSDIAGTT
```

Your input contains 1 sequence

Target database: UniProtKB Swiss-Prot

Restrict by taxonomy: homo sapiens

Homo sapiens (Human/Man) [9606]

Homology

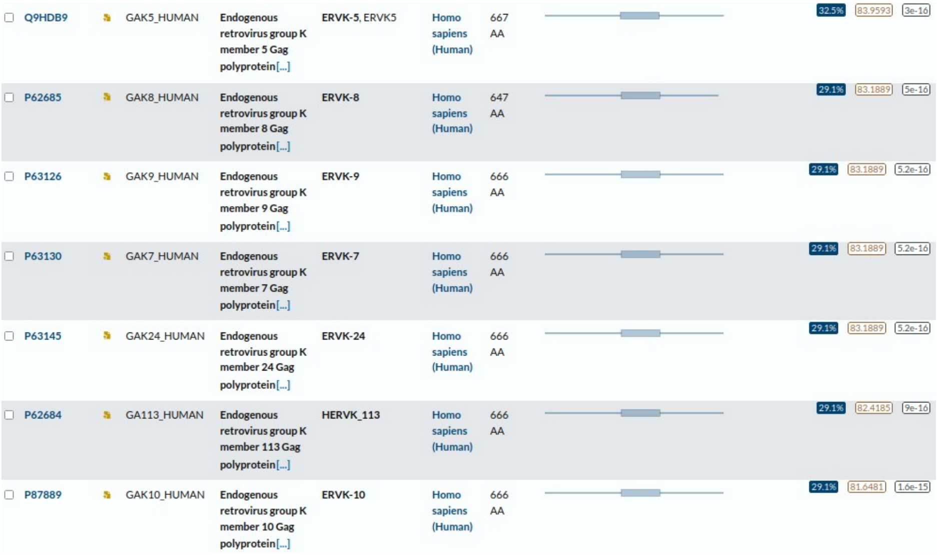
5

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*)?

Many retroviral proteins integrated in the human genome.

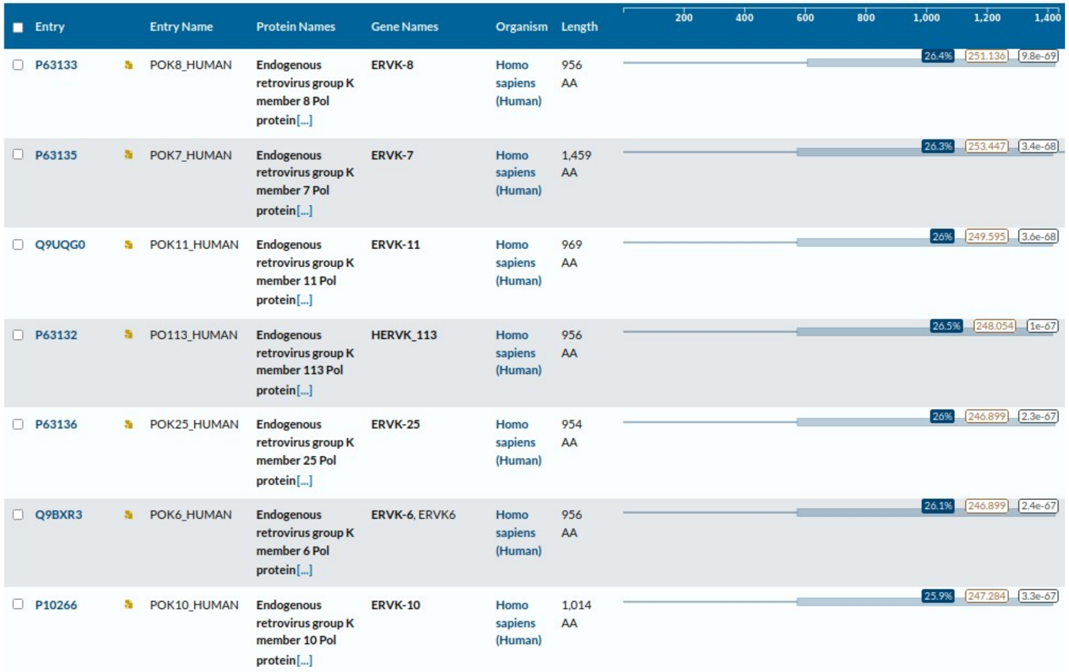
b. The Gag-Pol polyprotein is composed of more than one protein. Can you identify them? Use only SwissProt proteins.

N-terminal → Gag



□	Q9HDB9	GAK5_HUMAN	Endogenous retrovirus group K member 5 Gag polyprotein[...]	ERVK-5, ERVK5	Homo sapiens (Human)	667 AA	32.5%	83.9593	3e-16
□	P62685	GAK8_HUMAN	Endogenous retrovirus group K member 8 Gag polyprotein[...]	ERVK-8	Homo sapiens (Human)	647 AA	29.1%	83.1889	5e-16
□	P63126	GAK9_HUMAN	Endogenous retrovirus group K member 9 Gag polyprotein[...]	ERVK-9	Homo sapiens (Human)	666 AA	29.1%	83.1889	5.2e-16
□	P63130	GAK7_HUMAN	Endogenous retrovirus group K member 7 Gag polyprotein[...]	ERVK-7	Homo sapiens (Human)	666 AA	29.1%	83.1889	5.2e-16
□	P63145	GAK24_HUMAN	Endogenous retrovirus group K member 24 Gag polyprotein[...]	ERVK-24	Homo sapiens (Human)	666 AA	29.1%	83.1889	5.2e-16
□	P62684	GA113_HUMAN	Endogenous retrovirus group K member 113 Gag polyprotein[...]	HERVK_113	Homo sapiens (Human)	666 AA	29.1%	82.4185	9e-16
□	P87889	GAK10_HUMAN	Endogenous retrovirus group K member 10 Gag polyprotein[...]	ERVK-10	Homo sapiens (Human)	666 AA	29.1%	81.6481	1.6e-15

C-terminal → Pol



■	Entry	Entry Name	Protein Names	Gene Names	Organism	Length	200 400 600 800 1,000 1,200 1,400		
□	P63133	POK8_HUMAN	Endogenous retrovirus group K member 8 Pol protein[...]	ERVK-8	Homo sapiens (Human)	956 AA	26.4%	251.136	9.8e-69
□	P63135	POK7_HUMAN	Endogenous retrovirus group K member 7 Pol protein[...]	ERVK-7	Homo sapiens (Human)	1,459 AA	26.3%	253.447	3.4e-68
□	Q9UQG0	POK11_HUMAN	Endogenous retrovirus group K member 11 Pol protein[...]	ERVK-11	Homo sapiens (Human)	969 AA	26%	249.595	3.6e-68
□	P63132	PO113_HUMAN	Endogenous retrovirus group K member 113 Pol protein[...]	HERVK_113	Homo sapiens (Human)	956 AA	26.5%	248.054	1e-67
□	P63136	POK25_HUMAN	Endogenous retrovirus group K member 25 Pol protein[...]	ERVK-25	Homo sapiens (Human)	954 AA	26%	246.899	2.3e-67
□	Q9BXR3	POK6_HUMAN	Endogenous retrovirus group K member 6 Pol protein[...]	ERVK-6, ERVK6	Homo sapiens (Human)	956 AA	26.1%	246.899	2.4e-67
□	P10266	POK10_HUMAN	Endogenous retrovirus group K member 10 Pol protein[...]	ERVK-10	Homo sapiens (Human)	1,014 AA	25.9%	247.284	3.3e-67

MSA

6

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
              ****
```

b)

```
Protein_A     KKKYYWWKKT
Protein_B     AKKYYWW---
               ******
```

c)

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

MSA

6

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

MSA

*Images from: Wikipedia

7

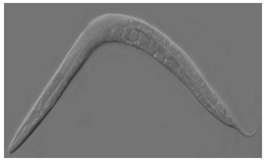
a. Both "P17861" (XBP1_HUMAN) and "Q3SZZ2" (XBP1_BOVIN) are "X-box binding protein 1" proteins. Can you detect which region/s of these proteins is/are important for their function? Why? Use Clustal Omega.
What should you do to detect them?

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.

c. Check the positional annotations in the entry of the human protein. Was the region you identified annotated as a domain?

Human
(*Homo sapiens*)

Cattle
(*Bos taurus*)



Worm
(*Caenorhabditis elegans*)

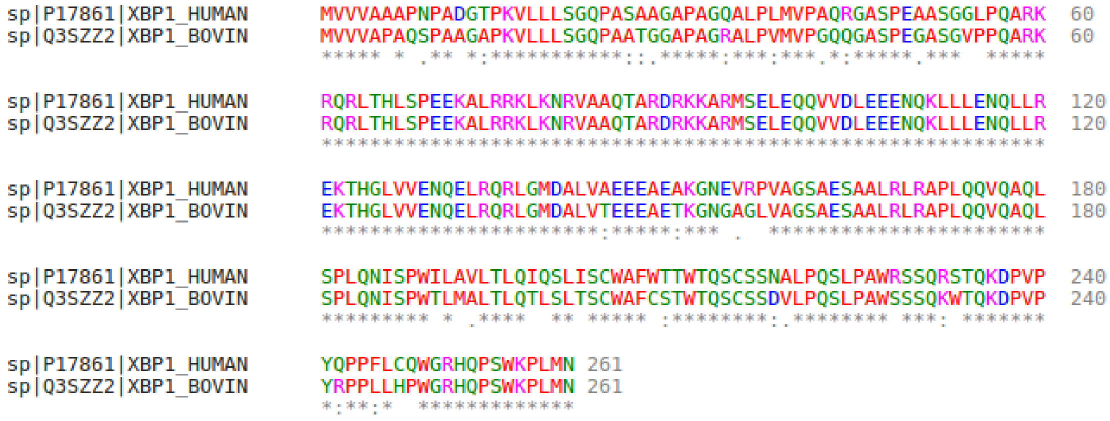
Zebra fish
(*Danio rerio*)

Pablo Mier
munoz@uni-mainz.de

MSA

*Images from: Clustal Omega, UniProt

7



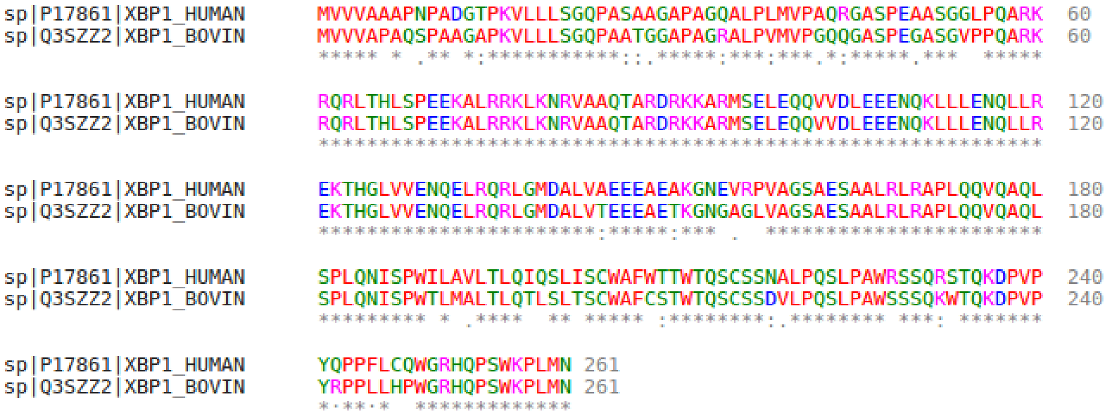
```
sp|P17861|XBP1_HUMAN      MVVVAAAPNPADGTPKVLLLSGQPASAAGAPAGQALPLMVPAQRGASPEAASGGLPQARK  60
sp|Q3SZZ2|XBP1_BOVIN      MVVVAPAQSPAAGAPKVLLLSGQPAATGGAPAGRALPVMVPGQQGASPEGASGVPPQARK  60
                          ***** * .** *:***********::.*****:***:***.*:*****.***  *****

sp|P17861|XBP1_HUMAN      RQRLTHLSPEEKALRRKLKNRVAAQTARDRKKARMSELEQQVVDLEEENQKLLLENQLLR  120
sp|Q3SZZ2|XBP1_BOVIN      RQRLTHLSPEEKALRRKLKNRVAAQTARDRKKARMSELEQQVVDLEEENQKLLLENQLLR  120
                          ************************************************************

sp|P17861|XBP1_HUMAN      EKTHGLVVENQELRQRLGMDALVAEEEAEAKGNEVRPVAGSAESAALRLRAPLQQVQAQL  180
sp|Q3SZZ2|XBP1_BOVIN      EKTHGLVVENQELRQRLGMDALVTEEEAETKGNGAGLVAGSAESAALRLRAPLQQVQAQL  180
                          ***********************:*****:*** .  ***********************

sp|P17861|XBP1_HUMAN      SPLQNISPWILAVLTLQIQSLISCWAFWTTWTQSCSSNALPQSLPAWRSSQRSTQKDPVP  240
sp|Q3SZZ2|XBP1_BOVIN      SPLQNISPWTLMALTLQTLSLTSCWAFCSTWTQSCSSDVLPQSLPAWSSSQKWTQKDPVP  240
                          ********* * .****  ** ***** :********:.******** ***: *******

sp|P17861|XBP1_HUMAN      YQPPFLCQWGRHQPSWKPLMN 261
sp|Q3SZZ2|XBP1_BOVIN      YRPPLLHPWGRHQPSWKPLMN 261
                          *:**:*  *************
```

a. No. They are too similar. We would need a protein from a more distant organism.

MSA

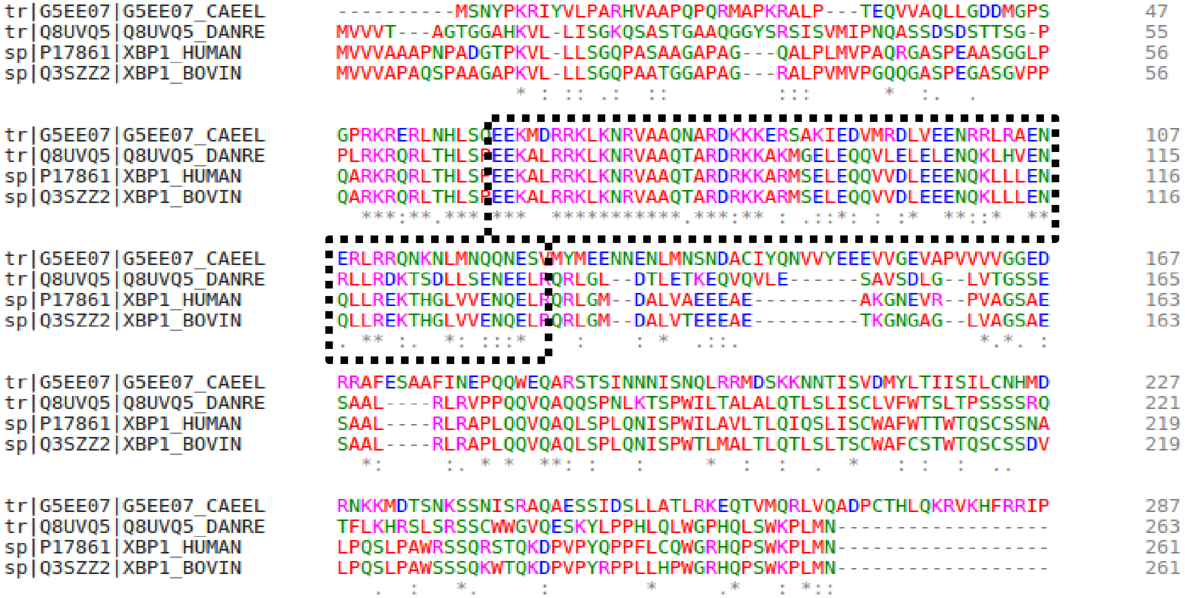
*Images from: Clustal Omega, UniProt

7



```
sp|P17861|XBP1_HUMAN      MVVVAAAPNPADGTPKVLLLSGQPASAAGAPAGQALPLMVPAQRGASPEAASGGLPQARK  60
sp|Q3SZZ2|XBP1_BOVIN      MVVVAPAQSPAAGAPKVLLLSGQPAATGGAPAGRALPVMVPGQQGASPEGASGVPPQARK  60
                          ***** * .** *:***********::.*****:***:***.*:*****.***  *****

sp|P17861|XBP1_HUMAN      RQRLTHLSPEEKALRRKLKNRVAAQTARDRKKARMSELEQQVVDLEEENQKLLLENQLLR  120
sp|Q3SZZ2|XBP1_BOVIN      RQRLTHLSPEEKALRRKLKNRVAAQTARDRKKARMSELEQQVVDLEEENQKLLLENQLLR  120
                          ************************************************************

sp|P17861|XBP1_HUMAN      EKTHGLVVENQELRQRLGMDALVAEEEAEAKGNEVRPVAGSAESAALRLRAPLQQVQAQL  180
sp|Q3SZZ2|XBP1_BOVIN      EKTHGLVVENQELRQRLGMDALVTEEEAETKGNGAGLVAGSAESAALRLRAPLQQVQAQL  180
                          ***********************:*****:*** .  ***********************

sp|P17861|XBP1_HUMAN      SPLQNISPWILAVLTLQIQSLISCWAFWTTWTQSCSSNALPQSLPAWRSSQRSTQKDPVP  240
sp|Q3SZZ2|XBP1_BOVIN      SPLQNISPWTLMALTLQTLSLTSCWAFCSTWTQSCSSDVLPQSLPAWSSSQKWTQKDPVP  240
                          ********* * .****  ** ***** :********:.******** ***: *******

sp|P17861|XBP1_HUMAN      YQPPFLCQWGRHQPSWKPLMN 261
sp|Q3SZZ2|XBP1_BOVIN      YRPPLLHPWGRHQPSWKPLMN 261
                          *:**:*  *************
```

a. No. They are too similar. We would need a protein from a more distant organism.



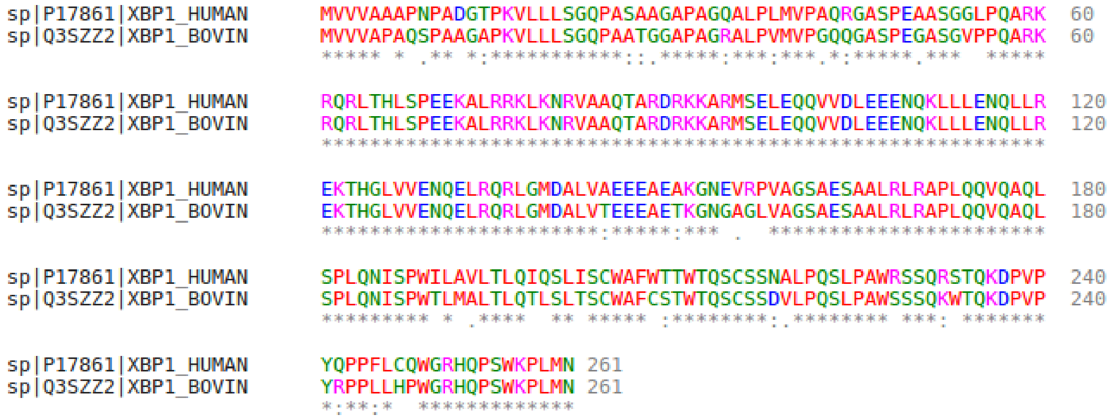
```
tr|G5EE07|G5EE07_CAEEL      ----------MSNYPKRIYVLPARHVAAPQPQRMAPKRALP---TEQVVAQLLGDDMGPS      47
tr|Q8UVQ5|Q8UVQ5_DANRE      MVVVT---AGTGGAHKVL-LISGKQSASTGAAQGGYSRSISVMIPNQASSDSDSTTSG-P      55
sp|P17861|XBP1_HUMAN        MVVVAAAPNPADGTPKVL-LLSGQPASAAGAPAG---QALPLMVPAQRGASPEAASGGLP      56
sp|Q3SZZ2|XBP1_BOVIN        MVVVAPAQSPAAGAPKVL-LLSGQPAATGGAPAG---RALPVMVPGQQGASPEGASGVPP      56
                                           * : :: .:  ::         :::      *  :.  .

tr|G5EE07|G5EE07_CAEEL      GPRKRERLNHLSQEEKMDRRKLKNRVAAQNARDKKKERSAKIEDVMRDLVEENRRLRAEN      107
tr|Q8UVQ5|Q8UVQ5_DANRE      PLRKRQRLTHLSPEEKALRRKLKNRVAAQTARDRKKAKMGELEQQVLELELENQKLHVEN      115
sp|P17861|XBP1_HUMAN        QARKRQRLTHLSPEEKALRRKLKNRVAAQTARDRKKARMSELEQQVVDLEEENQKLLLEN      116
sp|Q3SZZ2|XBP1_BOVIN        QARKRQRLTHLSPEEKALRRKLKNRVAAQTARDRKKARMSELEQQVVDLEEENQKLLLEN      116
                              ***:**.*** ***  ***********.***:** : .::*: : :*  **::*  **

tr|G5EE07|G5EE07_CAEEL      ERLRRQNKNLMNQQNESVMYMEENNENLMNSNDACIYQNVVYEEEVVGEVAPVVVVGGED      167
tr|Q8UVQ5|Q8UVQ5_DANRE      RLLRDKTSDLLSENEELRQRLGL--DTLETKEQVQVLE------SAVSDLG--LVTGSSE      165
sp|P17861|XBP1_HUMAN        QLLREKTHGLVVENQELRQRLGM--DALVAEEEAE---------AKGNEVR--PVAGSAE      163
sp|Q3SZZ2|XBP1_BOVIN        QLLREKTHGLVVENQELRQRLGM--DALVTEEEAE---------TKGNGAG--LVAGSAE      163
                            . ** :.  *: :::*    :    : *  .::.                    *.*. :

tr|G5EE07|G5EE07_CAEEL      RRAFESAAFINEPQQWEQARSTSINNNISNQLRRMDSKKNNTISVDMYLTIISILCNHMD      227
tr|Q8UVQ5|Q8UVQ5_DANRE      SAAL----RLRVPPQQVQAQQSPNLKTSPWILTALALQTLSLISCLVFWTSLTPSSSSRQ      221
sp|P17861|XBP1_HUMAN        SAAL----RLRAPLQQVQAQLSPLQNISPWILAVLTLQIQSLISCWAFWTTWTQSCSSNA      219
sp|Q3SZZ2|XBP1_BOVIN        SAAL----RLRAPLQQVQAQLSPLQNISPWTLMALTLQTLSLTSCWAFCSTWTQSCSSDV      219
                              *:     :. * *  **: :   :     *  :  :  .  *   : :  :  ..

tr|G5EE07|G5EE07_CAEEL      RNKKMDTSNKSSNISRAQAESSIDSLLATLRKEQTVMQRLVQADPCTHLQKRVKHFRRIP      287
tr|Q8UVQ5|Q8UVQ5_DANRE      TFLKHRSLSRSSCWWGVQESKYLPPHLQLWGPHQLSWKPLMN------------------      263
sp|P17861|XBP1_HUMAN        LPQSLPAWRSSQRSTQKDPVPYQPPFLCQWGRHQPSWKPLMN------------------      261
sp|Q3SZZ2|XBP1_BOVIN        LPQSLPAWSSSQKWTQKDPVPYRPPLLHPWGRHQPSWKPLMN------------------      261
                               .  :   *.     :        *     .*   :  *::
```

b. Yes. They are not as similar.

MSA

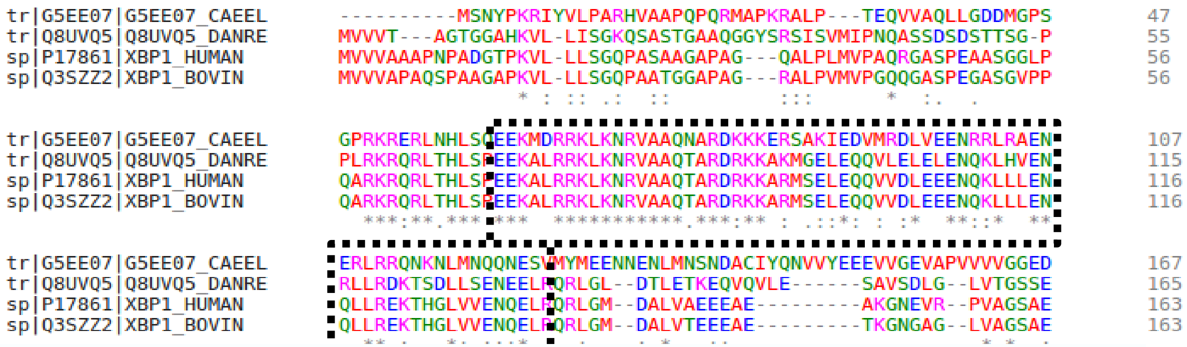
*Images from: Clustal Omega, UniProt

7



```
sp|P17861|XBP1_HUMAN      MVVVAAAPNPADGTPKVLLLSGQPASAAGAPAGQALPLMVPAQRGASPEAASGGLPQARK  60
sp|Q3SZZ2|XBP1_BOVIN      MVVVAPAQSPAAGAPKVLLLSGQPAATGGAPAGRALPVMVPGQQGASPEGASGVPPQARK  60
                          ***** *  .** *:***********:: .*****:***:***.*:*****.***  *****

sp|P17861|XBP1_HUMAN      RQRLTHLSPEEKALRRKLKNRVAAQTARDRKKARMSELEQQVVDLEEENQKLLLENQLLR  120
sp|Q3SZZ2|XBP1_BOVIN      RQRLTHLSPEEKALRRKLKNRVAAQTARDRKKARMSELEQQVVDLEEENQKLLLENQLLR  120
                          ************************************************************

sp|P17861|XBP1_HUMAN      EKTHGLVVENQELRQRLGMDALVAEEEAEAKGNEVRPVAGSAESAALRLRAPLQQVQAQL  180
sp|Q3SZZ2|XBP1_BOVIN      EKTHGLVVENQELRQRLGMDALVTEEEAETKGNGAGLVAGSAESAALRLRAPLQQVQAQL  180
                          ***********************:*****:***  .   ***********************

sp|P17861|XBP1_HUMAN      SPLQNISPWILAVLTLQIQSLISCWAFWTTWTQSCSSNALPQSLPAWRSSQRSTQKDPVP  240
sp|Q3SZZ2|XBP1_BOVIN      SPLQNISPWTLMALTLQTLSLTSCWAFCSTWTQSCSSDVLPQSLPAWSSSQKWTQKDPVP  240
                          ********* *  .****  ** *****  :*********:.******** ***: *******

sp|P17861|XBP1_HUMAN      YQPPFLCQWGRHQPSWKPLMN 261
sp|Q3SZZ2|XBP1_BOVIN      YRPPLLHPWGRHQPSWKPLMN 261
                          *:**:*  *************
```

a. No. They are too similar. We would need a protein from a more distant organism.

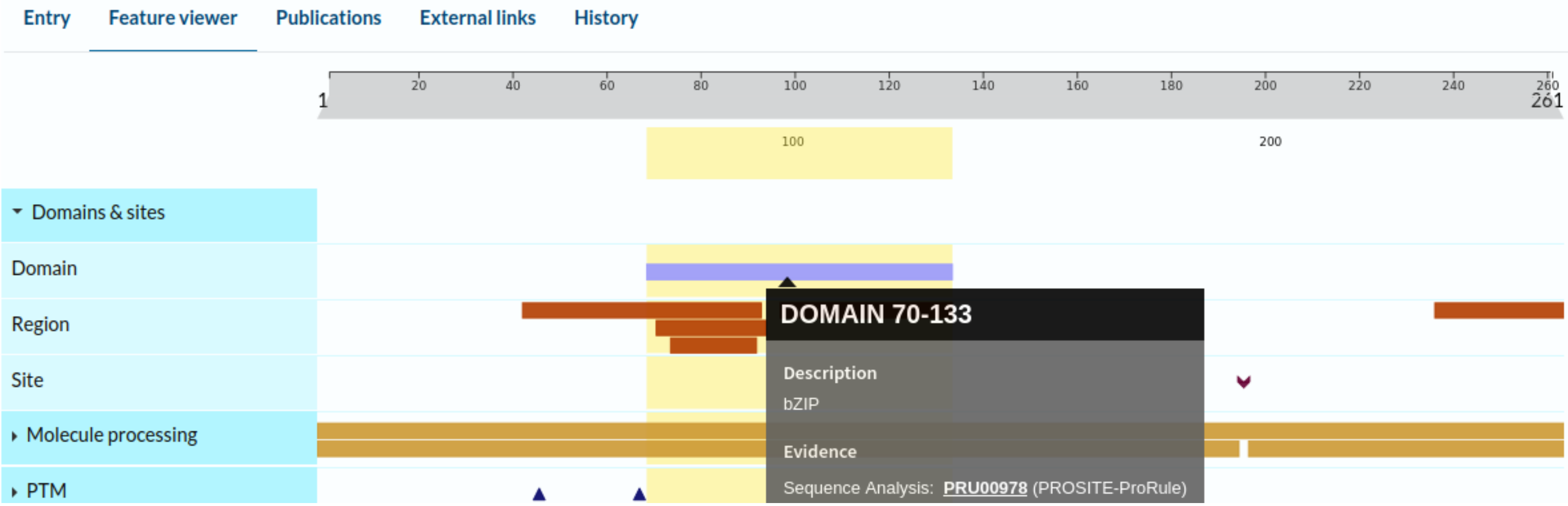


```
tr|G5EE07|G5EE07_CAEEL      ----------MSNYPKRIYVLPARHVAAPQPQRMAPKRALP---TEQVVAQLLGDDMGPS      47
tr|Q8UVQ5|Q8UVQ5_DANRE      MVVVT---AGTGGAHKVL-LISGKQSASTGAAQGGYSRSISVMIPNQASSDSDSTTSG-P      55
sp|P17861|XBP1_HUMAN        MVVVAAAPNPADGTPKVL-LLSGQPASAAGAPAG---QALPLMVPAQRGASPEAASGGLP      56
sp|Q3SZZ2|XBP1_BOVIN        MVVVAPAQSPAAGAPKVL-LLSGQPAATGGAPAG---RALPVMVPGQQGASPEGASGVPP      56
                                           * : :: .:  ::          :::      *  :.  .

tr|G5EE07|G5EE07_CAEEL      GPRKRERLNHLSQEEKMDRRKLKNRVAAQNARDKKKERSAKIEDVMRDLVEENRRLRAEN     107
tr|Q8UVQ5|Q8UVQ5_DANRE      PLRKRQRLTHLSPEEKALRRKLKNRVAAQTARDRKKAKMGELEQQVLELELENQKLHVEN     115
sp|P17861|XBP1_HUMAN        QARKRQRLTHLSPEEKALRRKLKNRVAAQTARDRKKARMSELEQQVVDLEEENQKLLLEN     116
sp|Q3SZZ2|XBP1_BOVIN        QARKRQRLTHLSPEEKALRRKLKNRVAAQTARDRKKARMSELEQQVVDLEEENQKLLLEN     116
                              ***:**.*** ***  ***********.***:** :  .::*: : :*   **::*   **

tr|G5EE07|G5EE07_CAEEL      ERLRRQNKNLMNQQNESVMYMEENNENLMNSNDACIYQNVVYEEEVVGEVAPVVVVGGED     167
tr|Q8UVQ5|Q8UVQ5_DANRE      RLLRDKTSDLLSENEELRQRLGL--DTLETKEQVQVLE------SAVSDLG--LVTGSSE     165
sp|P17861|XBP1_HUMAN        QLLREKTHGLVVENQELRQRLGM--DALVAEEEAE---------AKGNEVR--PVAGSAE     163
sp|Q3SZZ2|XBP1_BOVIN        QLLREKTHGLVVENQELRQRLGM--DALVTEEEAE---------TKGNGAG--LVAGSAE     163

                                                                                         227
                                                                                         221
                                                                                         219
                                                                                         219

                                                                                         287
                                                                                         263
                                                                                         261
                                                                                         261
```

b. Yes. They are not as similar.

c. bZIP (basic-leucine zipper) domain in positions:

70-133 (human)
70-133 (cattle)
61-117 (worm)
69-132 (zebrafish)



DOMAIN 70-133

Description

bZIP

Evidence

Sequence Analysis: PRU00978 (PROSITE-ProRule)

SeaView

Go to “http://doua.prabi.fr/software/seaview” (or search in Google “SeaView alignment”), download and start SeaView, with which we will generate the phylogenetic trees. Steps:

1. Go to -> http://doua.prabi.fr/software/seaview
2. Click on “MS Windows” to download the software
3. Click on downloaded file "seaview5.exe"
4. Ausführen
5. Extract
6. Go to folder "seaview"
7. Click on "seaview.exe"

If you have problems, do not worry → Use UniProt

Phylogeny

8

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/mb22b/). 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.

#1

#2

#3

#4

#5

#6

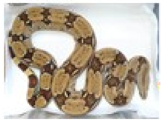
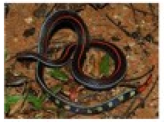
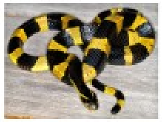
#7

#8

#9

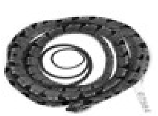
#10

#11

#12

#13

#14

#15

#16

#17

#18

#19

#20

#21

#22

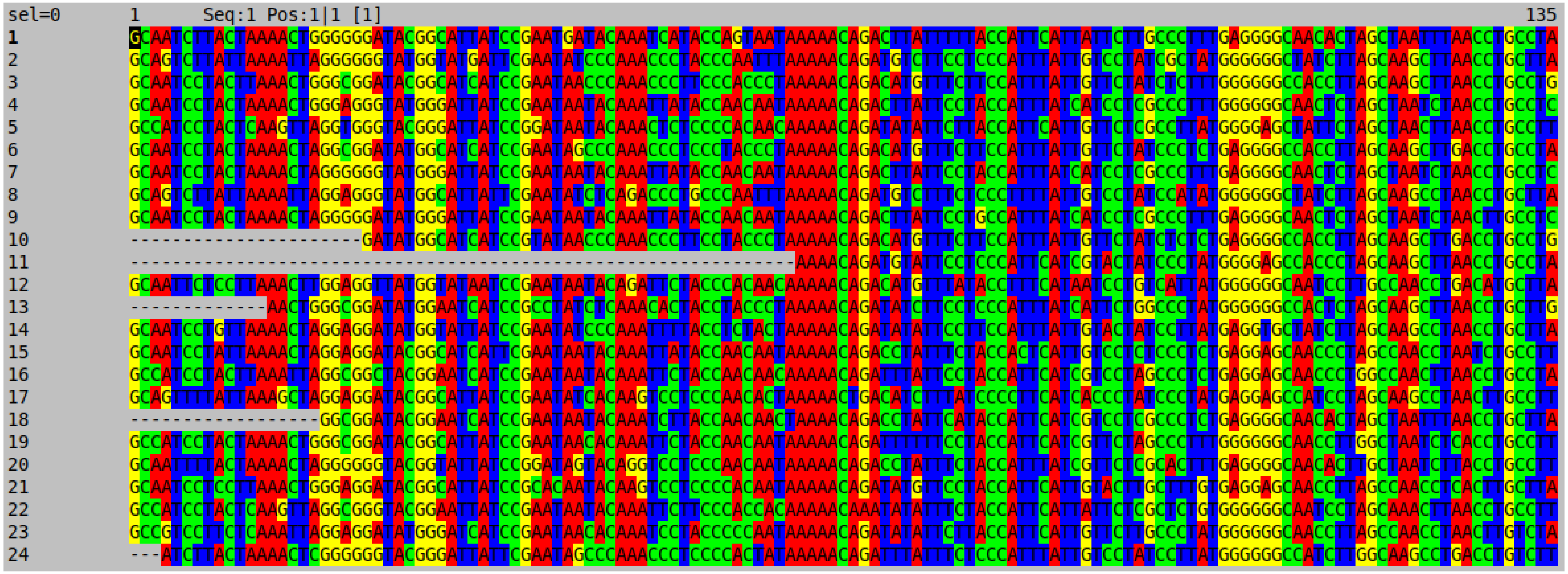
#23

#24

Phylogeny

8

1. Align the sequences → “Align > Align all”

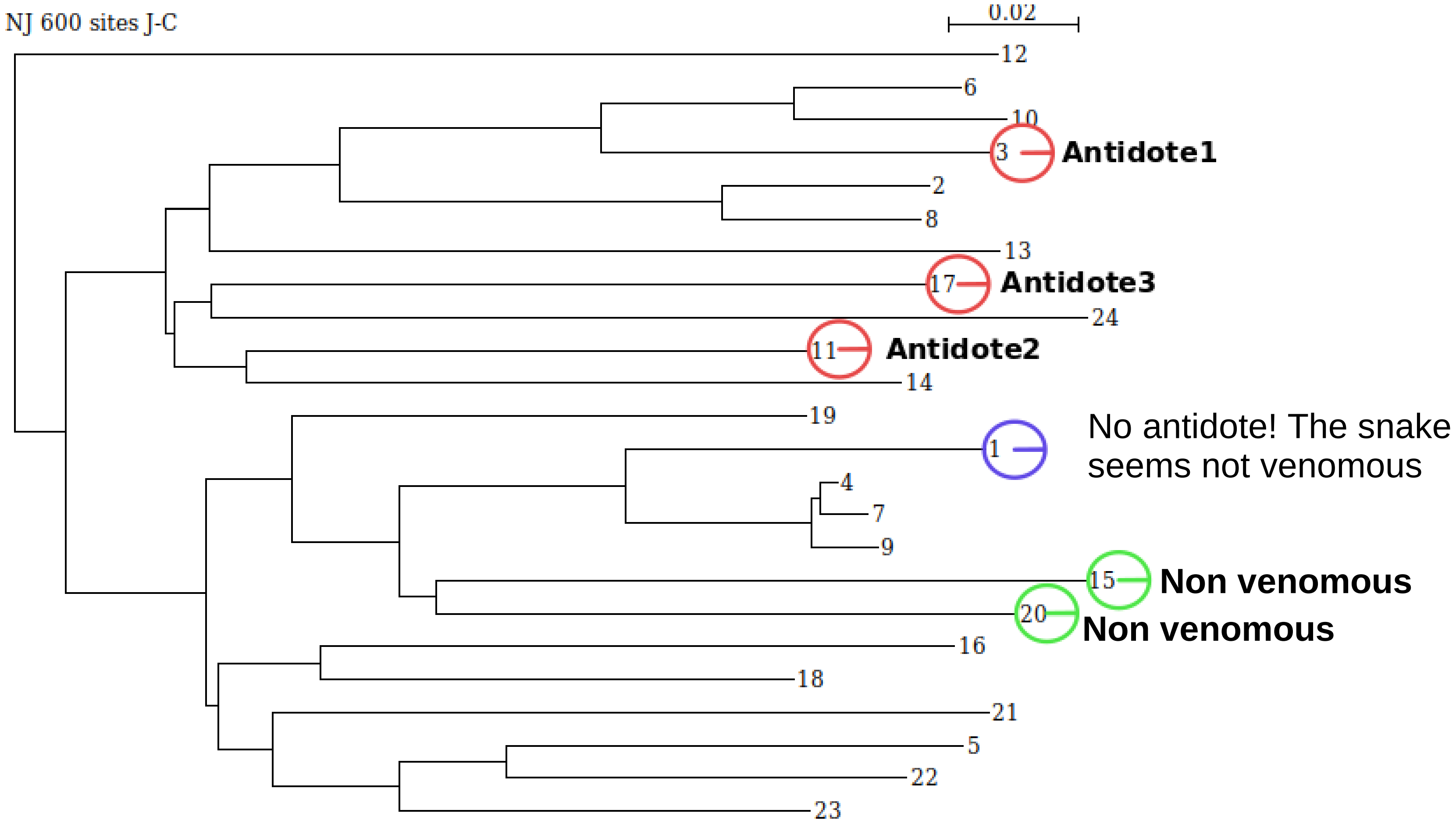


```
sel=0   1      Seq:1 Pos:1|1 [1]                                                                                                                135
1       GCAATCTTACTAAAACTGGGGGGATACGGCATTATCCGAATGATACAAATCATACCAGTAATAAAAACAGACTTATTTTTACCATTCATTATTCTTGCCCTTTGAGGGGCAACACTAGCTAATTTAACCTGCCTA
2       GCAGTCTTATTAAAATTAGGGGGGTATGGTATGATTCGAATATCCCAAACCCTACCCAATTTAAAAACAGATGTCTTCCTCCCATTTATTGTCCTATCGCTATGGGGGGCTATCTTAGCAAGCTTAACCTGCTTA
3       GCAATCCTACTTAAACTGGGCGGATACGGCATCATCCGAATAACCCAAACCCTTCCCACCCTAAAAACAGACATGTTTCTTCCATTTATTGTTCTATCTCTTTGGGGGGCCACCTTAGCAAGCTTAACCTGCCTG
4       GCAATCCTACTAAAACTGGGAGGGTATGGGATTATCCGAATAATACAAATTATACCAACAATAAAAACAGACTTATTCCTACCATTTATCATCCTCGCCCTTTGGGGGGCAACTCTAGCTAATCTAACCTGCCTC
5       GCCATCCTACTCAAGTTAGGTGGGTACGGGATTATCCGGATAATACAAACTCTCCCCACAACAAAAACAGATATATTCTTACCATTCATTGTTCTCGCCTTATGGGGAGCTATTCTAGCTAACTTAACCTGCCTT
6       GCAATCCTACTAAAACTAGGCGGATATGGCATCATCCGAATAGCCCAAACCCTCCCTACCCTAAAAACAGACATGTTTCTTCCATTTATTGTTCTATCCCTCTGAGGGGCCACCTTAGCAAGCTTGACCTGCCTA
7       GCAATCCTACTAAAACTAGGGGGGTATGGGATTATCCGAATAATACAAATTATACCAACAATAAAAACAGACTTATTCCTACCATTTATCATCCTCGCCCTTTGAGGGGCAACTCTAGCTAATCTAACCTGCCTC
8       GCAGTCTTATTAAAATTAGGAGGGTATGGCATTATTCGAATATCTCAGACCCTGCCCAATTTAAAAACAGATGTCTTTCTCCCTTTTATTGTCCTATCCATATGGGGGGCTATCTTAGCAAGCCTAACCTGCTTA
9       GCAATCCTACTAAAACTAGGGGGATATGGGATTATCCGAATAATACAAATTATACCAACAATAAAAACAGACTTATTCCTGCCATTTATCATCCTCGCCCTTTGAGGGGCAACTCTAGCTAATCTAACTTGCCTC
10      ----------------------GATATGGCATCATCCGTATAACCCAAACCCTTCCTACCCTAAAAACAGACATGTTTCTTCCATTTATTGTTCTATCTCTCTGAGGGGCCACCTTAGCAAGCTTGACCTGCCTG
11      ---------------------------------------------------------------AAAACAGATGTATTCCTCCCATTCATCGTACTATCCCTATGGGGAGCCACCCTAGCAAGCTTAACCTGCCTA
12      GCAATTCTCCTTAAACTTGGAGGTTATGGTATAATCCGAATAATACAGATTCTACCCACAACAAAAACAGACATGTTTATACCTTTCATAATCCTGTCATTATGGGGGGCAATCCTTGCCAACCTGACATGCTTA
13      -------------AACTGGGCGGATATGGAATCATCCGCCTATCTCAAACACTACCTACCCTAAAAACAGATATCTTCCTCCCATTTATCATTCTGGCCCTATGGGGGGCCACTCTAGCAAGCTTAACCTGCTTG
14      GCAATCCTGTTAAAACTAGGAGGATATGGTATTATCCGAATATCCCAAATTTTACCTCTACTAAAAACAGATATATTCCTTCCATTTATTGTACTATCCTTATGAGGTGCTATCTTAGCAAGCCTAACCTGCTTA
15      GCAATCCTATTAAAACTAGGAGGATACGGCATCATTCGAATAATACAAATTATACCAACAATAAAAACAGACCTATTTCTACCACTCATTGTCCTCTCCCTCTGAGGAGCAACCCTAGCCAACCTAATCTGCCTT
16      GCCATCCTACTTAAATTAGGCGGCTACGGAATCATCCGAATAATACAAATTCTACCAACAACAAAAACAGATTTATTCCTACCATTCATCGTCCTAGCCCTCTGAGGAGCAACCCTGGCCAACTTAACCTGCCTA
17      GCAGTTTTATTAAAGCTAGGAGGATACGGCATTATCCGAATATCACAAGTCCTCCCAACACTAAAAACTGACATCTTTATCCCCTTCATCACCCTATCCCTATGAGGAGCCATCCTAGCAAGCCTAACTTGCCTT
18      ------------------GGCGGATACGGAATCATCCGAATAATACAAATCTTACCAACAACTAAAACAGACCTATTCATACCATTCATCGTCCTCGCCCTCTGAGGGGCAACACTAGCTAATTTAACCTGCTTA
19      GCCATCCTACTAAAACTGGGCGGATACGGCATTATCCGAATAACACAAATTCTACCAACAATAAAAACAGATTTTTTCCTACCATTCATCGTTCTAGCCCTTTGGGGGGCAACCTTGGCTAATCTCACCTGCCTT
20      GCAATTTTACTAAAACTAGGGGGGTACGGTATTATCCGGATAGTACAGGTCCTCCCAACAATAAAAACAGACCTATTTCTACCATTTATCGTTCTCGCACTTTGAGGGGCAACACTTGCTAATCTTACCTGCCTT
21      GCAATCCTCCTTAAACTGGGAGGATACGGCATTATCCGCACAATACAAGTCCTCCCCACAATAAAAACAGATATGTTCCTACCATTCATTGTACTTGCTTTGTGAGGAGCAACCTTAGCCAACCTCACTTGCTTA
22      GCCATCCTACTCAAGTTAGGCGGGTACGGAATTATCCGAATAATACAAATTCTTCCCACCACAAAAACAAATATATTTCTACCATTCATTATTCTCGCTCTGTGGGGGGCAATCCTAGCAAACTTAACCTGCCTT
23      GCCGTCCTTCTCAAATTAGGAGGATATGGGATCATCCGAATAACACAAATCCTACCCCCAATAAAAACAGATATATTCTTACCATTCATTGTTCTTGCCCTATGGGGGGCAACCTTAGCCAACCTAACTTGTCTA
24      ---ATCTTACTAAAACTCGGGGGGTACGGGATTATTCGAATAGCCCAAACCCTCCCCACTATAAAAACAGATTTATTTCTCCCATTTATTGTCCTATCCTTATGGGGGGCCATCTTGGCAAGCCTGACCTGTCTT
```

Phylogeny

8

1. Align the sequences → “Align > Align all”
2. Build phylogenetic tree → “Trees > Distance methods > NJ
3. Define which antidote to administer



Pablo Mier
munoz@uni-mainz.de