Master Biomedizin 2017

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

Pablo Mier
munoz@uni-mainz.de

2) Homology

Pablo Mier
munoz@uni-mainz.de

*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*)

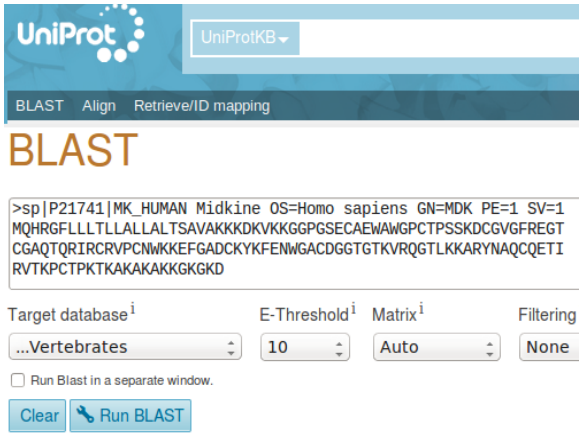
Frog
(*Xenopus laevis*)

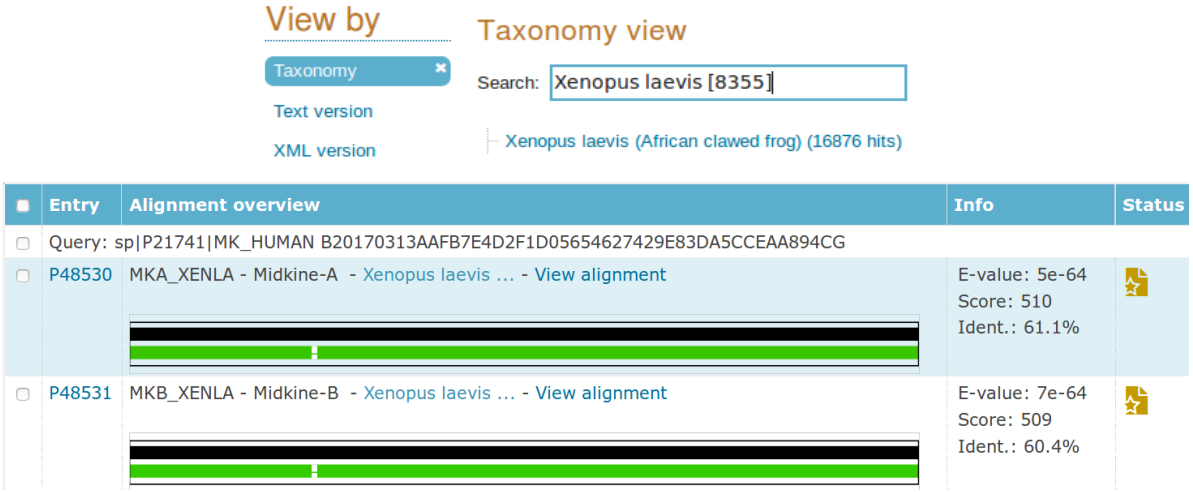
Pablo Mier
munoz@uni-mainz.de

Homology

*Images from: UniProt

3

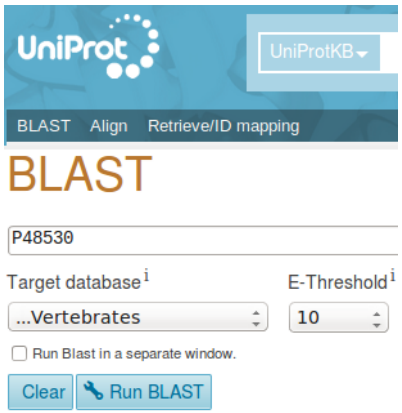


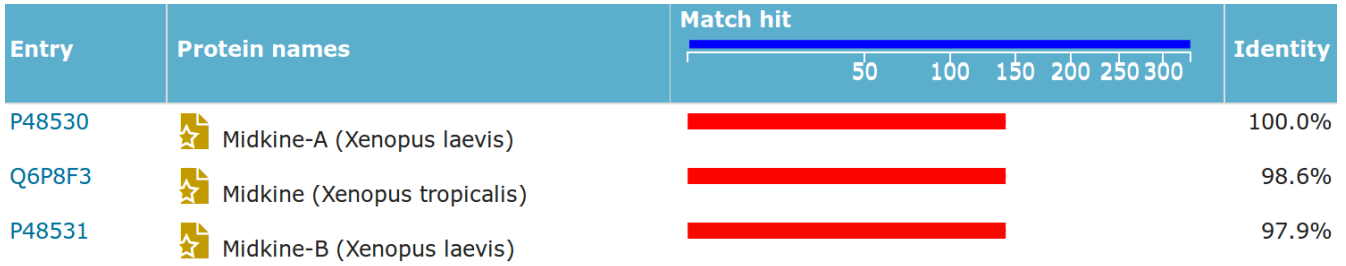


a. Query: P21741.
Ortholog1: P48530.
Ortholog2: P48531.

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

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





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*), 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*)

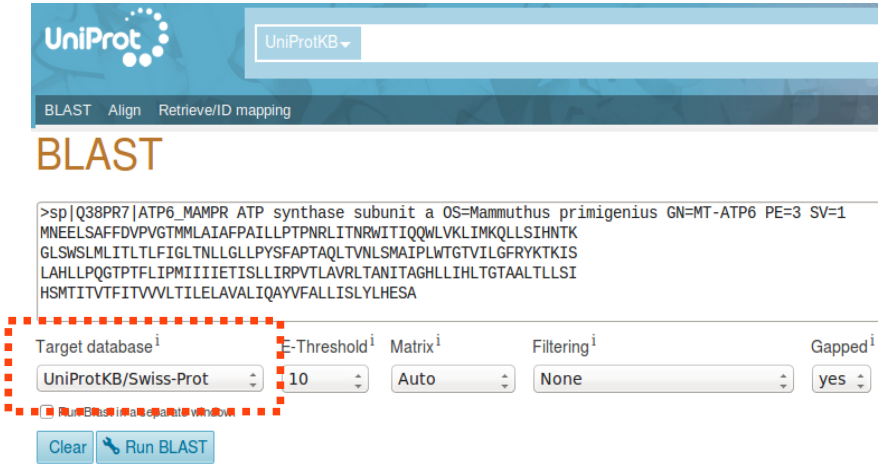
African elephant
(*Loxodonta africana*)

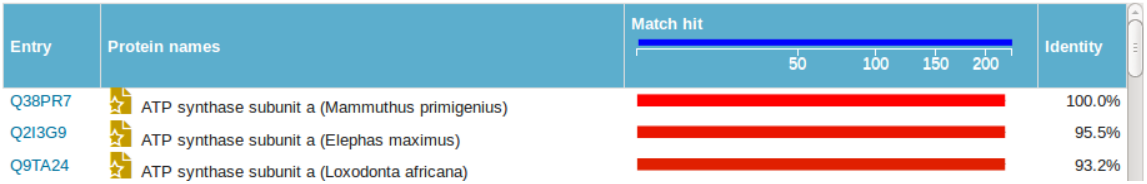
Pablo Mier
munoz@uni-mainz.de

Homology

*Images from: UniProt

4



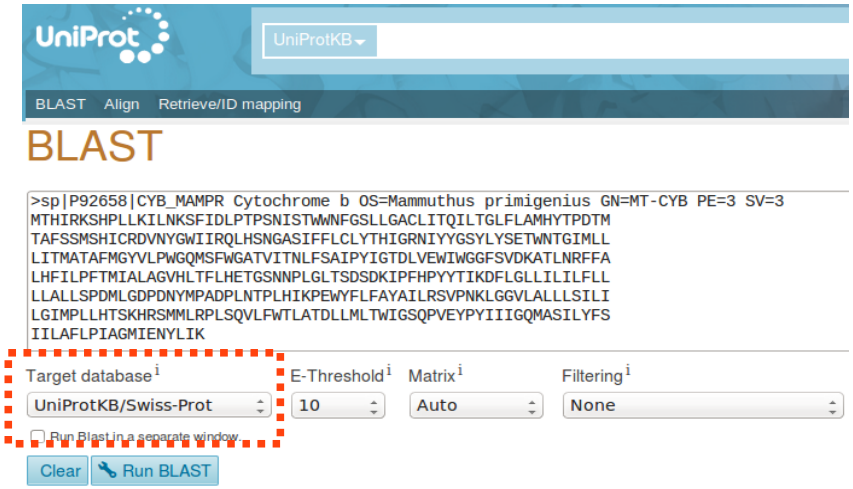


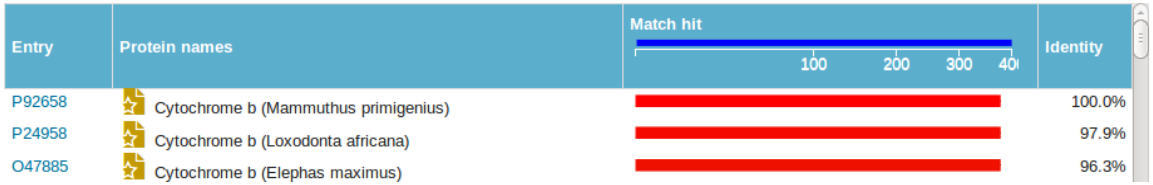
a. *M.primigenius* (Q38PR7) – *E.maximus* (Q2I3G9) = 95.5%
M.primigenius (Q38PR7) – *L.africana* (Q9TA24) = 93.2%

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 works on the woolly mammoth (PMID: 19020620), in 2008, provides evidence enough to determine that it is really closer to the asian elephant; corroborating the similarity shown in exercise 4a.





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

Pablo Mier
munoz@uni-mainz.de

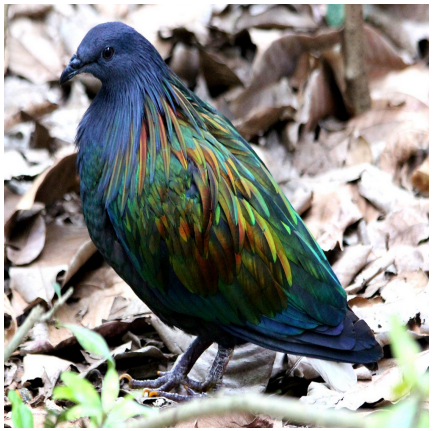
*Images from: Wikipedia

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.

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

Dodo
(*Raphus cucullatus*)



Nicobar pigeon
(*Caloenas nicobarica*)

Chicken (rooster)
(*Gallus gallus*)

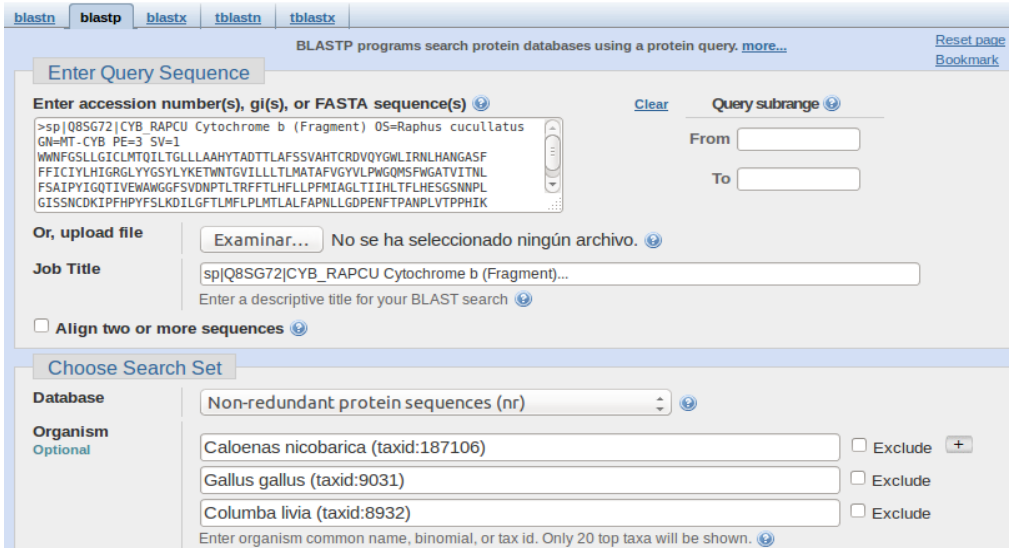
Pigeon
(*Columba livia*)

Pablo Mier
munoz@uni-mainz.de

Homology

*Images from: NCBI, Wikipedia

5



a. It seems that the dodo was closer to the pigeon than to the chicken.

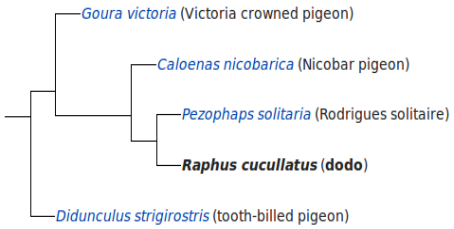
R.cucullatus – C.nicobarica = 99%

R.cucullatus – G.gallus = 93%

b. Same results for different pigeons.

R.cucullatus – C.livia = 96%

R.cucullatus – G.gallus = 93%



Alignments | Download | GenPept | Graphics | Distance tree of results | Multiple alignment

	Description	Max score	Total score	Query cover	E value	Ident	Accession
☑	cytochrome b [Caloenas nicobarica]	535	535	100%	0.0	99%	AAM19503.1
☑	cytochrome b [Columba livia]	526	526	100%	0.0	96%	YP_003540719.1
☐	cytochrome b [Columba livia]	522	522	100%	0.0	95%	AJK30555.1
☐	cytochrome b [Columba livia]	521	521	100%	0.0	95%	AKB93366.1
☑	cytochrome b [Gallus gallus]	509	509	100%	0.0	93%	ADB06697.1

Pablo Mier
munoz@uni-mainz.de

7

Using the protein “P38398”, perform a “tblastn” search in NCBI against human entries.

a. What would this search be used for?

b. Is there any difference between the first and the second result?

Pablo Mier
munoz@uni-mainz.de

Homology

*Images from: NCBI

7

a. Query: protein. Database: nucleotide. To look for the gene encoding the query protein.

	Description	Max score	Total score	Query cover	E value	Ident	Accession
☐	Homo sapiens breast cancer 1 (BRCA1), transcript variant 1, mRNA	3576	3576	100%	0.0	94%	NM_007294.3
☐	Homo sapiens breast and ovarian cancer susceptibility (BRCA1) mRNA, complete cds	3576	3576	100%	0.0	94%	U14680.1

Homo sapiens breast cancer 1 (BRCA1), transcript variant 1, mRNA

Sequence ID: ref|NM_007294.3| Length: 7224 Number of Matches: 1

Range 1: 233 to 5821 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
3576 bits(9273)	0.0	Compositional matrix adjust.	1863/1863(100%)	1863/1863(100%)	0/1863(0%)	+2

```
Query  1    MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQ  60
            MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQ
Sbjct  233  MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQ  412
```

Homo sapiens breast and ovarian cancer susceptibility (BRCA1) mRNA, complete cds

Sequence ID: gb|U14680.1|HSU14680 Length: 5711 Number of Matches: 1

Range 1: 120 to 5708 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
3576 bits(9273)	0.0	Compositional matrix adjust.	1863/1863(100%)	1863/1863(100%)	0/1863(0%)	+3

```
Query  1    MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQ  60
            MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQ
Sbjct  120  MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQ  299
```

b.

First result: NM_007294.3
7224 bp, transcript variant 1, mRNA

Query: 1 ——— 1863

Subject: 233 ——— 5821

Second result: U14680.1
5711bp, complete CDS

Query: 1 ——— 1863

Subject: 120 ——— 5708