

---

# Master Biomedizin 2017

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

---

## 2) Homology

## 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*)

3



UniProtKB

BLAST Align Retrieve/ID mapping

## BLAST

>sp|P21741|MK\_HUMAN Midkine OS=Homo sapiens GN=MDK PE=1 SV=1  
MQHRGFLLLTLLALLLTSAAVAKKDKVKKGGPGSECAEWAGPCTPSSKDCGVGFREGT  
CGAQTQIRICRVPCNWKKEFGADCKYKFENWGACDGGTGTQVRQGLKKARYNAQCQETI  
RVTKPCTPKTKAKAKAKGKGKD

Target database<sup>i</sup> E-Threshold<sup>i</sup> Matrix<sup>i</sup> Filtering

...Vertebrates 10 Auto None

☐ Run Blast in a separate window.

Clear Run BLAST

View by

Taxonomy

Text version

XML version

Taxonomy view

Search: Xenopus laevis [8355]

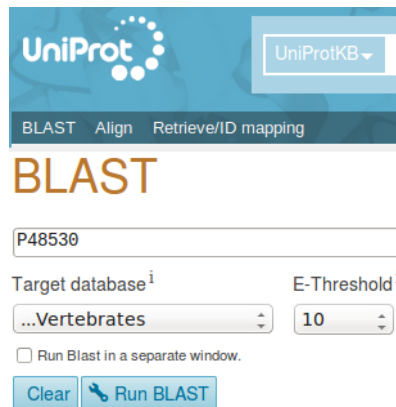
Xenopus laevis (African clawed frog) (16876 hits)

Entry	Alignment overview	Info	Status
Query: sp P21741 MK_HUMAN B20170313AAF7E4D2F1D05654627429E83DA5CCEAA894CG			
P48530	MKA_XENLA - Midkine-A - <a href="#">Xenopus laevis ... - View alignment</a>	E-value: 5e-64 Score: 510 Ident.: 61.1%	
P48531	MKB_XENLA - Midkine-B - <a href="#">Xenopus laevis ... - View alignment</a>	E-value: 7e-64 Score: 509 Ident.: 60.4%	

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”.



UniProtKB

BLAST Align Retrieve/ID mapping

## BLAST

P48530

Target database<sup>i</sup> E-Threshold<sup>i</sup>

...Vertebrates 10

☐ Run Blast in a separate window.

Clear Run BLAST

Entry	Protein names	Match hit	Identity
P48530	Midkine-A (Xenopus laevis)		100.0%
Q6P8F3	Midkine (Xenopus tropicalis)		98.6%
P48531	Midkine-B (Xenopus laevis)		97.9%



## 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 elephant (*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*)

4

UniProtKB

BLAST Align Retrieve/ID mapping

BLAST

>sp|Q38PR7|ATP6\_MAMPR ATP synthase subunit a OS=Mammuthus primigenius GN=MT-ATP6 PE=3 SV=1  
MNEELSAFFDVPVGTMLAIAFPAILLPNRLITNRWITIQQWLVKIMKQLLSIHNTK  
GLSWSLMLITLTLFIGLTNLLGLLPYSFAPTAQLTVNLSMAIPLWTGTVLGFRYKTKIS  
LAHLLPQGTPTFLIPMIIIIETISLLIRPVTAVRLTANITAGHLLIHLTGTAALTLISI  
HSMITITVTFITVVVLTILELAVALIQAYVFALLISLYLHESA

Target database<sup>i</sup> E-Threshold<sup>i</sup> Matrix<sup>i</sup> Filtering<sup>i</sup> Gapped<sup>i</sup>

UniProtKB/Swiss-Prot 10 Auto None yes

Clear Run BLAST

UniProtKB

BLAST Align Retrieve/ID mapping

BLAST

>sp|P92658|CYB\_MAMPR Cytochrome b OS=Mammuthus primigenius GN=MT-CYB PE=3 SV=3  
MTHIRKSHPLKILNKSFIDLPSTNISTWVNFSGLLGACLTITQILTGLFLAMHYTPDTM  
TAFSSMSHICRDVNYGWIIRQLHNSGASIFFLCLYTHIGRNIYGSYLYSETWNTGIMLL  
LITMATAFMGYVLPWGQMSFWGATVITNLSAIPYIGTDLVEWIGGFSVDKATLNRFFA  
LHFILPFTMIALAGVHLTFHETGSNNPLGLTSDSDKIPFHPYTIKDFLGLLILILFLL  
LLALLSPDMLGDPDNYMPADPLNTPHFKPEWYFLFAYAILRSVPNKLGGVLALLSILI  
LGIMPLHTSKHRSMLRPLSQVLFWTLATDMLLTWIGSQPVEYPYIIIGQMASILYFS  
IILAFPLIAGMIENYLIK

Target database<sup>i</sup> E-Threshold<sup>i</sup> Matrix<sup>i</sup> Filtering<sup>i</sup> Gapped<sup>i</sup>

UniProtKB/Swiss-Prot 10 Auto None yes

Clear Run BLAST

Entry	Protein names	Match hit	Identity
Q38PR7	ATP synthase subunit a (Mammuthus primigenius)	<div><div></div></div>	100.0%
Q2I3G9	ATP synthase subunit a (Elephas maximus)	<div><div></div></div>	95.5%
Q9TA24	ATP synthase subunit a (Loxodonta africana)	<div><div></div></div>	93.2%

- 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%

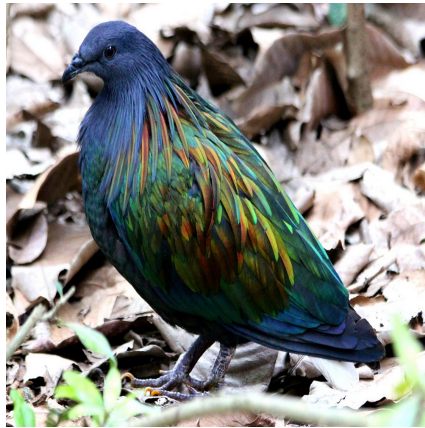
Entry	Protein names	Match hit	Identity
P92658	Cytochrome b (Mammuthus primigenius)	<div><div></div></div>	100.0%
P24958	Cytochrome b (Loxodonta africana)	<div><div></div></div>	97.9%
O47885	Cytochrome b (Elephas maximus)	<div><div></div></div>	96.3%

## 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*)

5

BLASTP programs search protein databases using a protein query. [more...](#)

**Enter Query Sequence**

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) [Query subrange](#)

>sp|Q8SG72|CYB\_RAPCU Cytochrome b (Fragment) 05=Raphus cucullatus  
 GN=MT-CYB PE=3 SV=1  
 MNWFGSLGICLMTQILTGLLLAAHYTADTTAFSSVAHTCRDQVQWLRNLHANGASF  
 FFICIVLHIGRGLYYSYLYKETWNTGVILLTLMATAFVGYLPGQMSFWGATVITNL  
 FSAIPYIGQTIWEAWGGFSVDNPTLTRFFTLHFLFPFIAGLTIIHLTFHESGSNNPL  
 GISSNCDKIPFHPYFSLKIDILGFTLMFLPLMTLALFAPNLLGOPENFTANPLVTPPHIK

Or, upload file [Examinar...](#) No se ha seleccionado ningún archivo. [?](#)

Job Title   
 Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

**Choose Search Set**

Database  [?](#)

Organism [Optional](#)

☐ Exclude [+](#)

☐ Exclude

☐ Exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. [?](#)

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

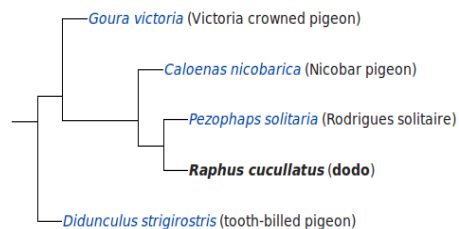
*R.cucullatus* – *C.nicobarica* = 99%

*R.cucullatus* – *G.gallus* = 93%

b. Same results for different pigeons.

*R.cucullatus* – *C.livia* = 96%

*R.cucullatus* – *G.gallus* = 93%



AT	Alignments	Download	GenPept	Graphics	Distance tree of results	Multiple alignment	
	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input checked="" type="checkbox"/>	<a href="#">cytochrome b [Caloenas nicobarica]</a>	535	535	100%	0.0	99%	<a href="#">AAM19503.1</a>
<input checked="" type="checkbox"/>	<a href="#">cytochrome b [Columba livia]</a>	526	526	100%	0.0	96%	<a href="#">YP_003540719.1</a>
<input type="checkbox"/>	<a href="#">cytochrome b [Columba livia]</a>	522	522	100%	0.0	95%	<a href="#">AJK30555.1</a>
<input type="checkbox"/>	<a href="#">cytochrome b [Columba livia]</a>	521	521	100%	0.0	95%	<a href="#">AKB93366.1</a>
<input checked="" type="checkbox"/>	<a href="#">cytochrome b [Gallus gallus]</a>	509	509	100%	0.0	93%	<a href="#">ADB06697.1</a>

## 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?

## 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
<input type="checkbox"/>	<a href="#">Homo sapiens breast cancer 1 (BRCA1), transcript variant 1, mRNA</a>	3576	3576	100%	0.0	94%	<a href="#">NM_007294.3</a>
<input type="checkbox"/>	<a href="#">Homo sapiens breast and ovarian cancer susceptibility (BRCA1) mRNA, complete cds</a>	3576	3576	100%	0.0	94%	<a href="#">U14680.1</a>

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	MDLSALRVVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQ					60
Sbjct 233	MDLSALRVVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQ					412



b. First result: NM\_007294.3  
7224 bp, transcript variant 1, mRNA

Query: 1 \_\_\_\_\_ 1863  
Subject: 233 \_\_\_\_\_ 5821

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	MDLSALRVVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQ					60
Sbjct 120	MDLSALRVVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQ					299



Second result: U14680.1  
5711bp, complete CDS

Query: 1 \_\_\_\_\_ 1863  
Subject: 120 \_\_\_\_\_ 5708