

Extinct Organisms Project

Part A is out of 10, and Part B is out of 10, with the breakdown noted (and part marks for a given question are possible, such as 0.5 for a 1 mark question).

A. The saber tooth “cat” (saber tooth tiger), *Smilodon populator* lived until 10,000-15,000 years ago. There are currently 11 sequences from this animal that have been submitted to Genbank.

1) What is common to all these sequences, other than that they are sequence fragments? (1 mark)

Ans.

The common part to all these sequences are that they are all derived from the mitochondrion of the *Smilodon populator*. For example, 12S and 16S ribosomal RNA, cytochrome b (cytb), ATPase 8 (ATP8), and NADH dehydrogenase subunit 5 (ND5).

Method

For extraction of the following genes: First go to NCBI website and then select nucleotide Gen Bank on the side panel and in the search bar type “*Smilodon Populator*”. Then sort the results by date released. Pick the first gene of the first sequence and then 5 oldest sequences from the bottom.

Table 1 : extracted genes from GenBank

Sequence Name	Sequence ID
<i>Smilodon populator</i> voucher ZMA20.042 mitochondrion, partial genome	MF871700.1
<i>Smilodon populator</i> isolate ZMA20.041 cytochrome b (cytb) gene, partial cds; mitochondrial	DQ097174.1
<i>Smilodon populator</i> isolate ZMA20.041 12S ribosomal RNA gene, partial sequence; mitochondrial	DQ097171.1
<i>Smilodon populator</i> isolate ZMA20.041 NADH dehydrogenase subunit 5 (ND5) gene, partial cds; mitochondrial	DQ097169.1
<i>Smilodon populator</i> isolate ZMA20.041 ATPase 8 (ATP8) gene, partial cds; mitochondrial	DQ097167.1

<i>Smilodon populator</i> isolate ZMA20.041 16S ribosomal RNA gene, partial sequence; mitochondrial	DQ097165.1
---	------------

2) Obtain the oldest 5 sequences (based on release date), plus the first gene from the largest sequence (most recently released). BLAST each sequence separately against the appropriate NCBI database (which BLAST program would you use?). (1 mark)

Ans: Methodology used to top hits.

Blastn: - was particularly used as we extracted all the nucleotide sequences for all five of the oldest sequences and for the first gene of the first sequence. Also the database used was “Nucleotide collection (nr/nt)”. Results were picked on the basis of high Max Score, Total Score, Query Cover, Identity Value, and low E value. IT was seen that sometimes blastn and megablast both resulted in same results with a little difference in high max. To keep the work consistent, we kept only used on blastn for all sequences to find the closest hits.

4) For each *S. populator* gene, extract the sequences of the closest "hits" (homologous sequences) from at least four different species. Create a gene tree for each *S. populator* gene and its current neighbors.

Ans:

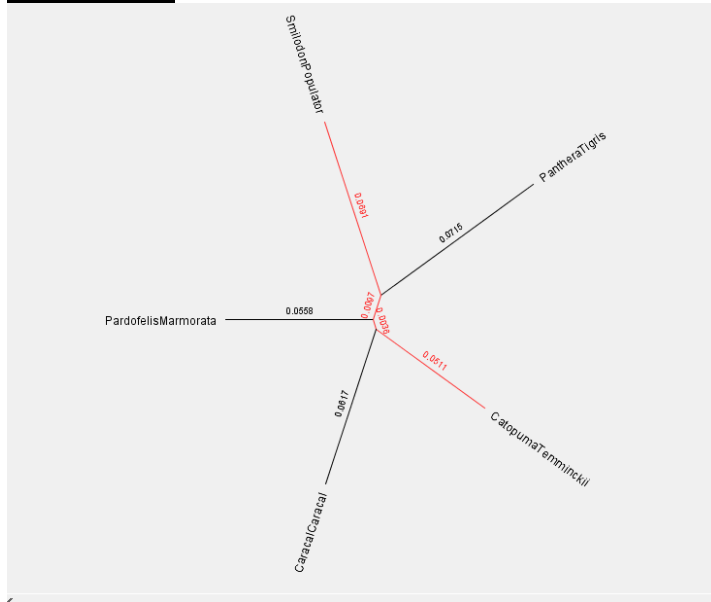
Method:

Blastn : same methodology used for each sequence from table 1. Please refer question 2 for detailed methodology.

1. FASTA aligned sequences for the best hits were downloaded and were aligned through ClustalX version 2.1 (Windows).
2. PHB file with 1000 bootstrap trials was created using “Bootstrap N-J tree” with “Exclude Positions with Gaps” and “Correct for Multiple Substitutions” selected, and bootstrap label switched to “Node”.
3. PHB files were viewed in Fig Tree version 1.4.3 (Windows).
4. Branch length was also present on the phylogenetic trees.
5. In each generated we deduced the branch length and the closest relative was chosen on the basis nearest distance from Smilodon Populator.

1. For *Smilodon populator* voucher ZMA20.042 mitochondrion, partial genome - MF871700.1

Gene tree -1

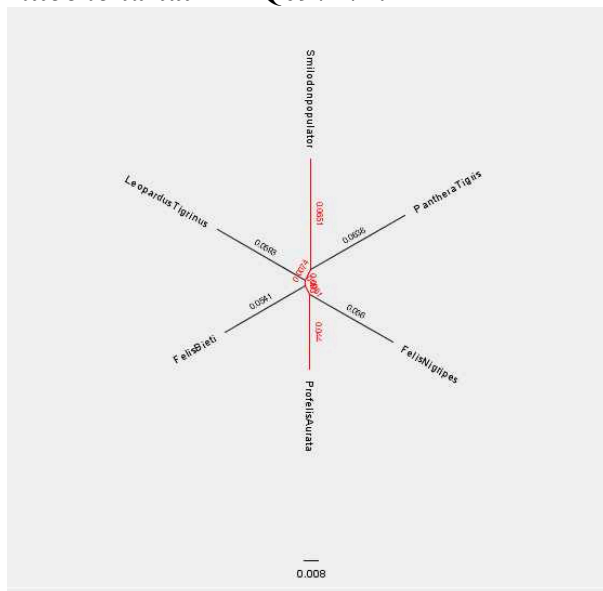


Phylogenetic Tree 1:

Expected related species :
Panthera Tigris, Pardofelis marmorata, Caracal caracal, Catopuma Temminckii.

Closest Relative: Catopuma Temminckii

2. *Smilodon populator* isolate ZMA20.041 cytochrome b (cytb) gene, partial cds; mitochondrial - DQ097174.1

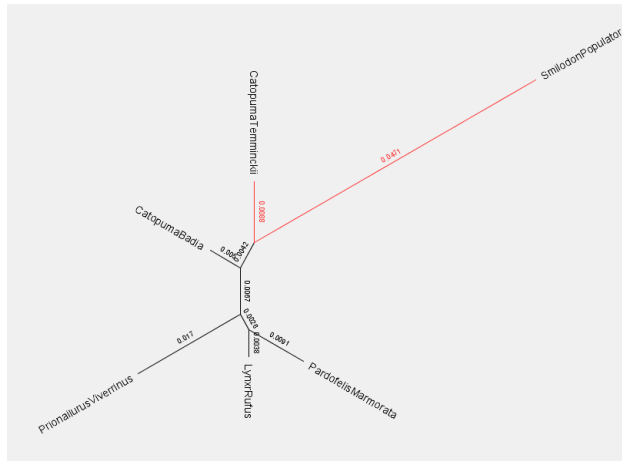


Phylogenetic Tree 2: Expected Closest Species according to our hits: - Panthera Tigris, Leopardus Tigrinus , Felis Beiti, Profelis Aurata, Felis Nigripes.

Closest Relative: Profelis aurata

Expected Closest Species according to our hits: - Panthera Tigris, Leopardus Tigrinus , Felis Beiti, Profelis Aurata, Felis Nigripes.

3. *Smilodon populator* isolate ZMA20.041 12S ribosomal RNA gene, partial sequence;
mitochondrial- DQ097171.1
Gene Tree 3

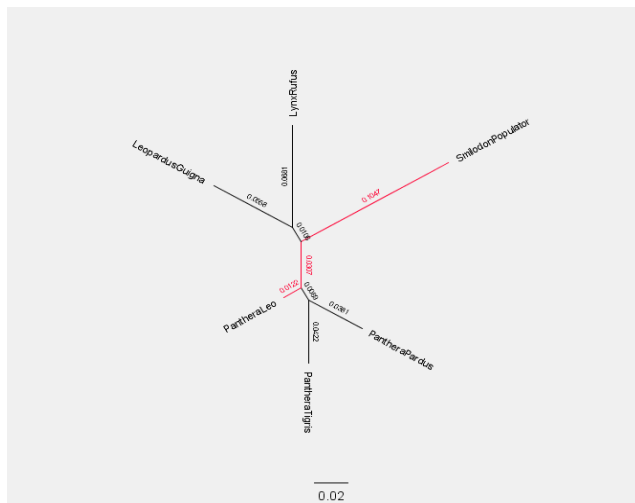


Phylogenetic Tree 3:

Expected Closest Species :
Catopuma Temminckii,
Catopuma Badia, Prionailurus
Viverrinus, LynX Rufus,
Pardofelis Mamorata.

Closest Relative : Catopuma
Temminckii

4. *Smilodon populator* isolate ZMA20.041 NADH dehydrogenase subunit 5 (ND5) gene,
partial cds; mitochondrial - DQ097169.1

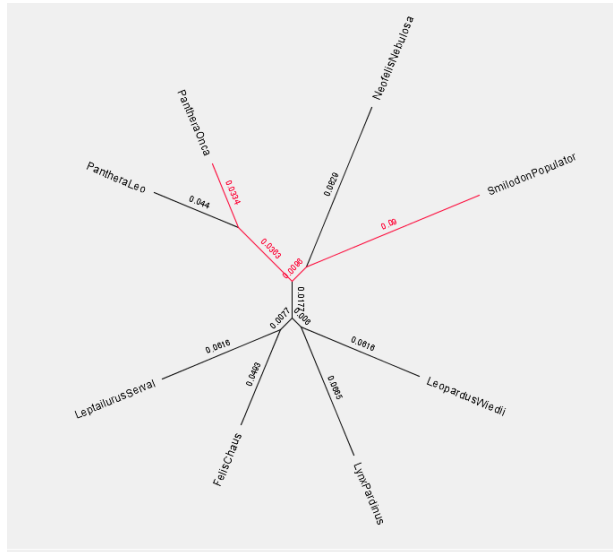


Phylogenetic Tree 4:

Expected Closest Species: Lynx
Rufus, Leopardus Guigna,
Panthera Leo, Panthera Tigris,
Panthera Pardus.

Closest Relative: Panthera Leo.

5. *Smilodon populator* isolate ZMA20.041 ATPase 8 (ATP8) gene, partial cds; mitochondrial-DQ097167.1

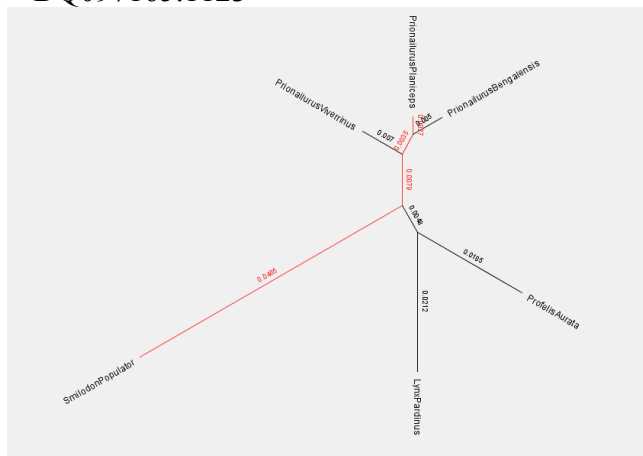


Phylogenetic Tree 5:

Expected Closest Species:
Neofelis Nebulosa, Panthera
Onca, Panthera Leo, Leptailurus
Serval, FelisChaus,
LynxPardinus, Leopardus Weidii.

Closest Relative: Panthera Onca.

6. *Smilodon populator* isolate ZMA20.041 16S ribosomal RNA gene, partial sequence;
mitochondrial
- DQ097165.1123



Phylogenetic Tree 6:

Expected Closest Species:
Prionaliurus Viverrinus,
Prionaliurus planiceps,
Prionaliurus Bengalensis, Profelis
Aurata, LynxPardinus.

Closest Relative : Prionaliurus
Planiceps

5) What can you say about the closest extant relatives of the Saber Tooth Cat? (1 mark) If the gene trees differ, why might that occur? (1 mark)
Is it better to call this organism a saber tooth “cat” or saber tooth “tiger”? (1 mark) Why? (1 mark)

For each of the nucleotide sequence in Table (1), the closest living relatives are

Table 2: The closest Living Relatives for each Sequence from table 1.

Sequence Name	Closest Relative
<i>Smilodon populator</i> voucher ZMA20.042 mitochondrion, partial genome	Catopuma temminckii (Asian golden Cat)
<i>Smilodon populator</i> isolate ZMA20.041 cytochrome b (cytb) gene, partial cds; mitochondrial	Profelis aurata (African golden Cat)
<i>Smilodon populator</i> isolate ZMA20.041 12S ribosomal RNA gene, partial sequence; mitochondrial	Catopuma temminckii (Asian golden Cat)
<i>Smilodon populator</i> isolate ZMA20.041 NADH dehydrogenase subunit 5 (ND5) gene, partial cds; mitochondrial	Panthera leo (Lion)
<i>Smilodon populator</i> isolate ZMA20.041 ATPase 8 (ATP8) gene, partial cds; mitochondrial	Panthera onca (Jaguar-Wild Cat)
<i>Smilodon populator</i> isolate ZMA20.041 16S ribosomal RNA gene, partial sequence; mitochondrial	Prionailurus planiceps (Flat head cat)

Each phylogenetic tree differs from the other because overtime genes undergo mutation and nucleotide base changes at different rates, thus, new species can emerge with genetic variation differs from species to species from their common ancestor. Some species genes may be drastically altered from its ancestor or it may remain conserved.

It is better to call the Saber Tooth Cat as a Cat not a tiger because almost all of the generated fig trees that we got infer towards “cats”. Being more close to cat, it would wrong to say that they were tigers. For prove of this we can refer to tree 4, and see that *Smilodon Popoulator* is really far

away from the *Panthera tigris*. And also in figure 1, we can see that it is closer to *Catopuma temminckii* rather than *panthera tigris* (tiger specie). We can also look at tree 2, and see that the distance from the panther tigris is higher than *Profelis aurata*. Also, rest of the generated phylogenetic trees (3,5,6) confirm that *Smilodon Populator* is closer to a cat specie than a tiger.

B. The following “mystery” sequences were obtained from animal bones in a glacier in the arctic that are estimated to be >10,000 years old.

Methodology for part B.

BLASTX

1. BLASTX under default settings was used to determine the possible proteins encoded from unknown sequences. *Sequence 1* and *Sequence 2* were run individually into BLASTX.
2. For each unknown sequence, the protein was selected for its high Max Score, Total Score, Query Cover, Identity Value, and low E-value.
3. First sequence 1, ref-seq database was used and rest of the options were kept default. And top five results were taken into account. And sequence 2, non-redundant protein database was used.
4. Also for each you can go and see related protein by clicking related protein link on the side screen on NCBI Blastx website.
5. For Introns the graphics option was chosen on the top left to see graphical representation of the protein.
6. Translated Proteins of these sequences were picked by clicking on the hit and then clicking on the sequence id. Translated protein was found at the bottom of the page. See figure below.

Protein Data Bank (PDB)

1. Translated protein was picked and used in protein Data bank.
2. We directly paste the translated sequence and then look for nearest similar protein.
For Sequence 1:- PDB ID : 1OCC was used to visualize in Deep view.
For Sequence 2:- PDB ID : 1BE3 was used to visualize in Deep view.

ClustalX - 2.1 Windows

1. All BLASTX FASTA sequences alongside their respective unknown FASTA sequence were loaded into ClustalX.

2. PHB file was created using 1000 bootstrap trials with “Bootstrap N-J tree”, “Exclude Positions with Gaps” and “Correct for Multiple Substitutions” selected, and bootstrap label switched to “Node”.

FigTree 1.4.3 (Windows).

1. CluslX resulted PHB file, was used with FigTree to create visual representation of sequenced based relatedness in a phylogenetic tree.
2. “Node Labels” to present 1000 bootstrap trials and “Radial Tree Layout” was used.
3. Branch length was also present.
4. Branch distances were deduced and closest relative was decided on the basis of the nearest distance Smilodon Populator.

Deep View 4.10 (Windows)

1. Downloaded nearest similar sequences from PDB and loaded the file of each sequence in Deep View.
 2. Removed unnecessary parts of the domain and only translated sequences that aligned with the related protein was viewed in Deep view.
 3. Another view was saved where gaps and mismatches were incorporated into the actual protein sequence and its effects on the alpha helices and beta sheets were also viewed. Screenshot of these are shown in this report.
 4. Alpha helices are colored yellow and beta sheets are colored red.
1. Can you identify any genes in these sequences? (1 mark) Are there any introns? (1 mark) Why? (1 mark)

Sequence 1: gene Identified was COX 2 – cytochrome c oxidase subunit II

```

GBY/..
Method: conceptual translation.
FEATURES             Location/Qualifiers
     source            1..227
                        /organism="Mammuthus primigenius"
                        /organelle="mitochondrion"
                        /db_xref="taxon:37349"
     Protein           1..227
                        /product="cytochrome c oxidase subunit II"
                        /name="COX2"
                        /calculated_mol_wt=25877
     Region            1..226
                        /region_name="COX2"
                        /note="Cytochrome c oxidase subunit II; Validated;
                        MTH0098"
                        /db_xref="CDD:177160"
     CDS               1..227
                        /gene="COX2"
                        /coded_by="NC_007596.2:7021..7704"
                        /transl_table=2
                        /db_xref="GeneID:373144"
ORIGIN
1  mayplqlgfg datspvmeel lhfdhtlmi ifliisslvly iimlmlttkl vhtnmmnvqe
61  memiwtlpa iililialps lhtlymdei nplltiktm ghqfwvsyey tdyedlafds
121 ymittsalkf gelrllevdn rmvlptdlpv rvlssedvl hswavpslgl ktdaipgrln
181 qvtltmrpg lfyggcseic gahsfmpiv lelvpkyfe swsasla
//

```


For sequence 2 gene identified was: cytb – cytochrome B – c1 subunit.

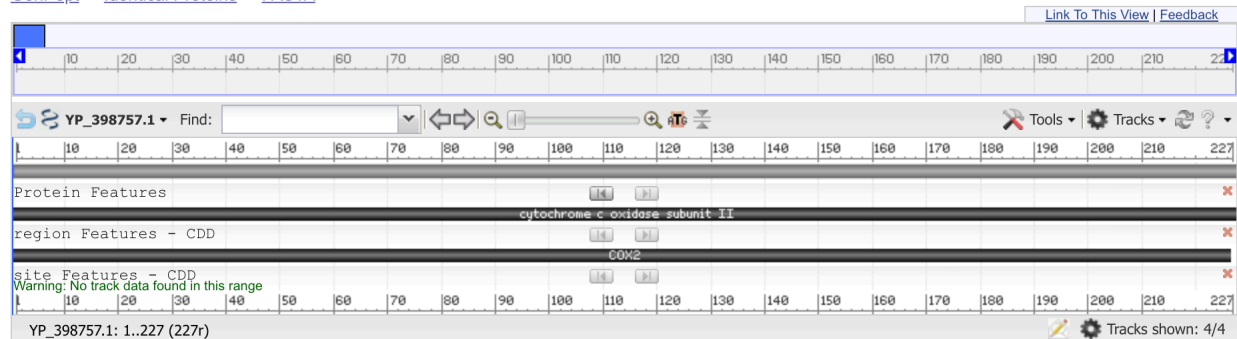
```
Site      /db_xref="CDD:238176"
          order(114,117..118,121..122,135,138..139,143,171,174)
          /site_type="other"
          /note="Qo binding site"
          /db_xref="CDD:238176"
CDS       1..242
          /gene="mtcyb"
          /coded_by="HG962401.1:1..>729"
          /codon_start=3
          /transl_table=2
          /db_xref="GOA:A0A068TLX0"
          /db_xref="InterPro:IPR003797"
          /db_xref="InterPro:IPR003798"
          /db_xref="InterPro:IPR016174"
          /db_xref="InterPro:IPR027387"
          /db_xref="UniProtKB/TrEMBL:A0A068TLX0"
IGIN      1 hpllkinnhs fdlpapsni sawnfgsl glclliqilt glflamhyts ddatafsvt
          61 hircdrvnygw liryahanga slfficflfh vgrglyygy tyfetwnvgv illfavmata
          121 fngpylpwqg mfwgatvit nllaapigv tllvewwgy fsvdkatlr ffafhfilpf
          181 ivaalvmvhl lflhetgenn pglisdsdk ipfhpyytik dvlgvllll lfmlilvfsp
          241 dl
```

No introns were seen when press on the graphics option of the *Mammuthus primigenius* reference sequence.

cytochrome c oxidase subunit II (mitochondrion) [*Mammuthus primigenius*]

NCBI Reference Sequence: YP_398757.1

[GenPept](#) [Identical Proteins](#) [FASTA](#)



No introns were seen when press on the graphics option of the *Sciurus igniventris* reference sequence.

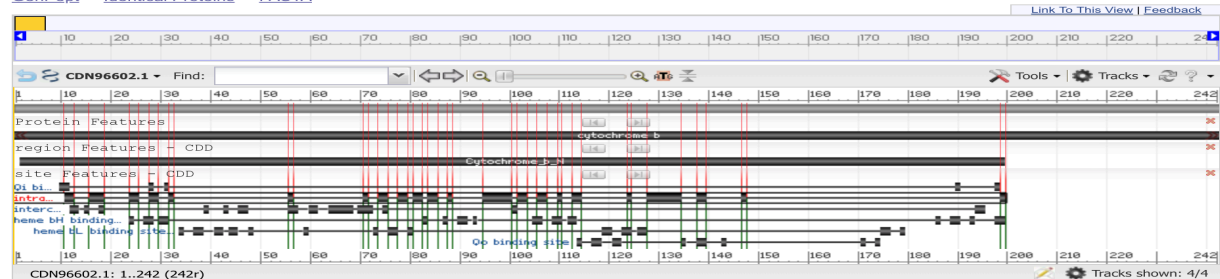
[Graphics](#)

[Send to:](#)

cytochrome b, partial (mitochondrion) [*Sciurus igniventris*]

GenBank: CDN96602.1

[GenPept](#) [Identical Proteins](#) [FASTA](#)



2) What species do they most likely belong too? (1 mark)

Sequence 1: It belongs to Mamuthus premigenius as it was best hit and had 93% identification.

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

[Alignments](#) [Download](#) [GenPept](#) [Graphics](#)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> cytochrome c oxidase subunit II (Mammuthus primigenius)	304	304	97%	3e-102	93%	YP_398757.1
<input type="checkbox"/> cytochrome c oxidase subunit II (Elephas maximus)	304	304	97%	4e-102	93%	YP_626370.1
<input type="checkbox"/> cytochrome c oxidase subunit II (Loxodonta cyclotis)	304	304	97%	4e-102	92%	YP_007626826.1
<input type="checkbox"/> cytochrome c oxidase subunit II (Loxodonta africana)	303	303	97%	1e-101	92%	NP_009282.1
<input type="checkbox"/> cytochrome c oxidase subunit II (Elephas antiquus)	301	301	97%	3e-101	91%	YP_009389989.1
<input type="checkbox"/> cytochrome c oxidase subunit II (Mammot americanum)	295	295	97%	1e-98	89%	YP_009423867.1

Sequence 2: It belongs to Sciurus igniventris as it was the best hit according to our methodology.

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:5

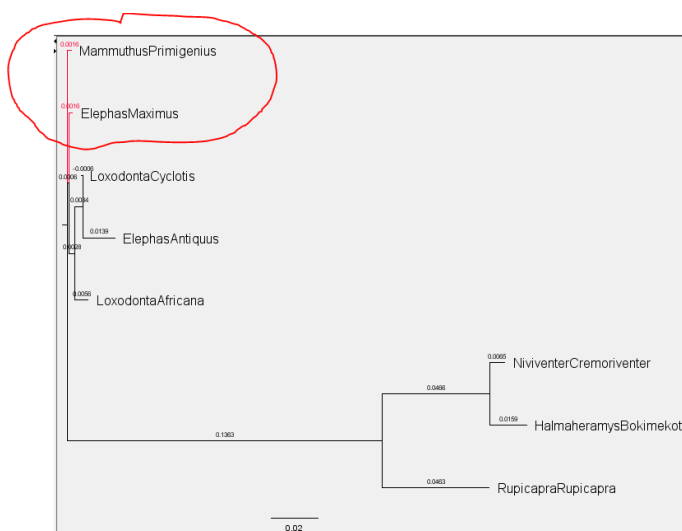
[Alignments](#) [Download](#) [GenPept](#) [Graphics](#)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input checked="" type="checkbox"/> cytochrome b (Sciurus igniventris)	367	367	92%	2e-126	94%	CDN95602.1
<input checked="" type="checkbox"/> cytochrome b (Tamias ruficaudus)	362	362	92%	9e-125	90%	AAL69114.1
<input checked="" type="checkbox"/> cytochrome b (Tamias amoenus)	362	362	92%	1e-124	90%	AAM80693.1
<input checked="" type="checkbox"/> cytochrome b (Tamias minimus)	362	362	92%	1e-124	90%	AXQ12119.1
<input checked="" type="checkbox"/> cytochrome b (Tamias senex)	362	362	92%	1e-124	90%	AAL69283.1
<input type="checkbox"/> cytochrome b (Tamias ruficaudus)	362	362	92%	1e-124	90%	AAL69101.1

3) What is the most likely closest living relative of each species? (1 mark)

For closest living relative we have to look at the phylogenetic tree.

Sequence 1

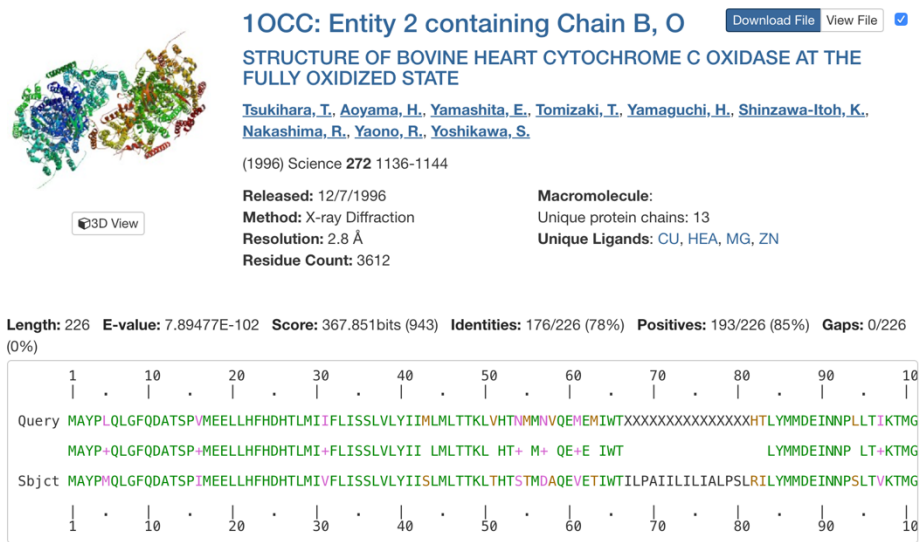


Phylogenetic tree 7:

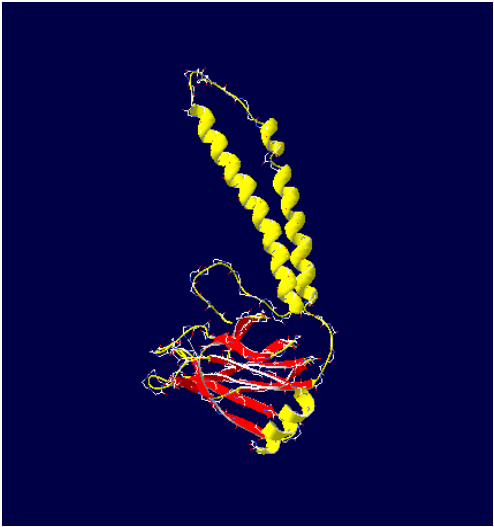
Expected relatives: Elephas Maximus, Loxodonta Cyclotis, Elphas Antiquus, Loxodonta Africana, NiviventerCremoriventer , Bokimekot, Rupicaprarupicapra.

Closest Living relative: Elephas maximus

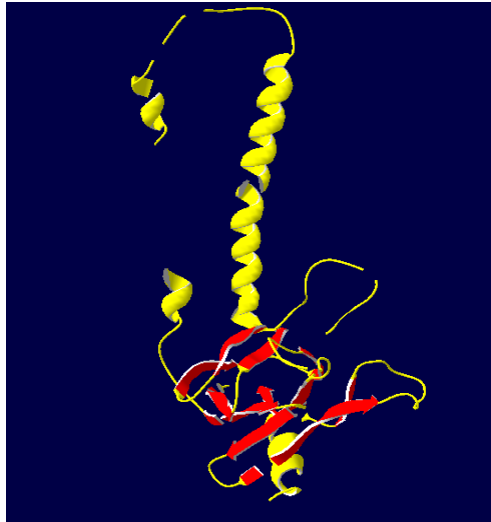
Protein Structures related to these sequences are: 1OCC – PDB Data shown below. Organism : Bos Taurus.



Deep View of the translated sequence 1 with and without gaps.



Deep View Figure 1: represent the part of Cytochrome C oxidase subunit II which is closely related the translated sequence 1. Three Alpha helices and 7 beta sheets are shown here in this figure.



Deep View Figure 2: represent the part of Cytochrome C oxidase subunit II which is closely related the translated sequence 1 but with removed missing links after alignment with 1OCC. As we can see actual translated sequence is incomplete and we see distorted/broken alpha helices and beta sheets in this figure.

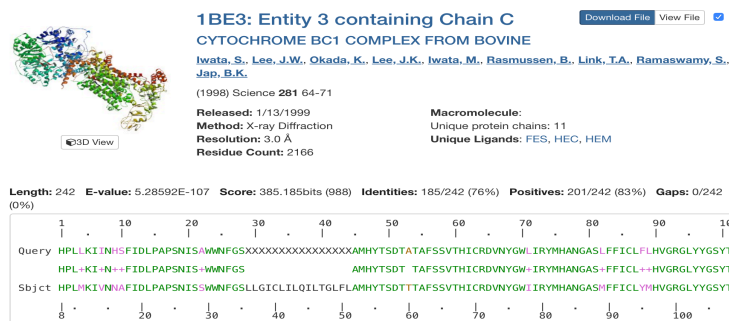
Sequence 2: (found very deep in a glacier crevice near some fossilized Quercus garryana acorn residue)

```
cccactgcttaaaattgttaactactcttttattgacctactgcccccttaacatctcagcctgatgaaactttggctccctccttgactttgcct
tctaataccaaattctaaccggcctatttctagctatacactacacatctgatacagcaacagctttctcctcgtgactcacatctgccgagacgt
aaactatggctgactgattcgatacatgcatgctaacggcgccctctctattttcatttgcttattcctcatgtgggtcgagggctttactatgggt
cctacacttattttgaaacctgaaacgttggagtattcttcttttgcagtaatagcaactgctttataggctatgtcctaccgtgaggacaaatat
catttgagggtgcaacagtcacccaacctcttatcagctatcccctatattggcactacttagtagaatgaatctgaggtgggttctcagtcg
acaaagcaactttaacacgatttttcgctttcacttcattctgcttttatcgttgacgctttagtcatagtccactattatttccacgaaacagg
atccaacaacccgtccggcctaattctctgattctgataagatccatttcacccctactacacaattaaagatgtattaggtgttctccttctattac
tcttatttatgattctagtctattttcccc
```

Translated Protein Sequence 2.

```
HPLLKIINHSDIDLPAPSNISAWWNFGSLLGLCLLIQILTGLFLAMHYTSDTATAFSSVTHI
CRDVNYGWLIRYMHANGASLFFICLFLHVGRGLYYGSYTYFETWNVGVILLFAVMATA
FMGYVLPWQGMSFWGATVITNLLSAIPYIGTTLVEWIWGGFSVDKATLTRFFAFHFILPF
IVAALVMVHLLFLHETGSNNPSGLISDSDKIPFHPYYTIKDVLGVLILLLLLFMILVLFSPD
L
```

Protein Structures related to these sequences are: 1BE3 - PDB Data shown below. Organism: Bos Taurus.





Deep view Figure 3: represent the part of Cytochrome B – c1 subunit - which is closely related the translated sequence 2. 7 alpha helices and 2 beta sheets.



Deep View Figure 4: represent the part of Cytochrome C oxidase subunit II which is closely related the translated sequence 1 but with removed missing links after alignment with 1OCC. As we can see actual translated sequence is incomplete and we see distorted/broken alpha helices but beta sheets are intact.

Now, Sequence 1 to encode for Cytochrome C oxidase subunit II, whereas Sequence 2 encoded for Cytochrome B – c1 subunit. Both encoded proteins are not independently functional as they are part of a larger complex.

References:

1. GenBank Overview <https://www.ncbi.nlm.nih.gov/genbank/>
2. Nucleotide BLAST: from <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
3. Figtree : <http://tree.bio.ed.ac.uk/software/figtree/>
4. Deep View : https://spdbv.vital-it.ch/download_prerelease.html
5. Protein Data Bank : https://spdbv.vital-it.ch/download_prerelease.html