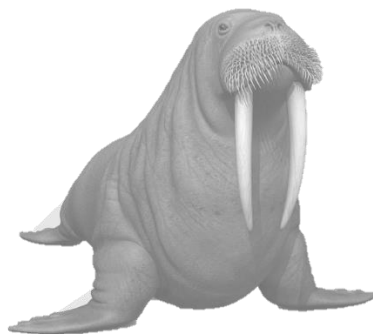




**Computational Genomics
and Bioinformatics**

EMATM0004

***A Phylogenetic Analysis of the
Obodenus Rosmarus***



Rabeeah Masood
1982306
MSc Biomedical Engineering

19th August 2020

Introduction

The modern walrus, *Odobenus rosmarus*, is a relict species, the lone survivor of a formerly diverse group of *Odobenidae Pinnipeds*. Walruses had evolved in the North Pacific by at least the middle Miocene and were moderately diverse (five sympatric species) by the late Miocene. This project investigates the evolutionary history of *Odobenus rosmarus* within the enigmatic and disputed subspecies of the Pinnipedia (i.e. walruses - *Odobenidae*, eared seals, including sea lions and fur seals - *Otariidae* & earless seals - *Phocidae*).

A majority of authors support a monophyletic origin of the Pinnipeds from a *Caniformia*, which are “dog-like” carnivorans. However, there are others who suggest a di-phyletic origin with the *Phocidae* being related to the *Mustelids* (a disputed family on itself). (Andersen, et al., 1998).

Data Description

The primary species for this investigation will be *Odobenus rosmarus rosmarus*; The accession number for the complete mtDNA is NC_004029(.2). The walrus' phylogeny will be computed in relation to the bearded Seal (*Erignathus barbatus* representing *Phocidae*), the California Sea Lion (*Zalophus californianus* representing *Otariidae*), the Polar bear (*Ursusmaritimus* representing *Caniformia*) and the Wolverine (*Gulo gulo* representing *mustelids*). As an out group to root the phylogenetic trees, *Homo sapiens* will be used. Appendix A provides the full table of accession numbers of species used.

Sequence Statistics

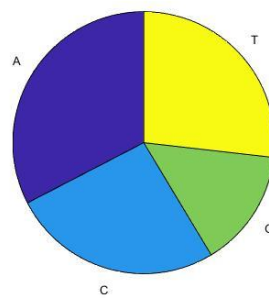
An in-depth analysis of the mtDNA of the *Odobenus Rosmarus* provided us with interesting statistics and information.

The length of the genome was found to be 16565 bp long. Out of which the division of specific base pairs was found to be somewhat odd.

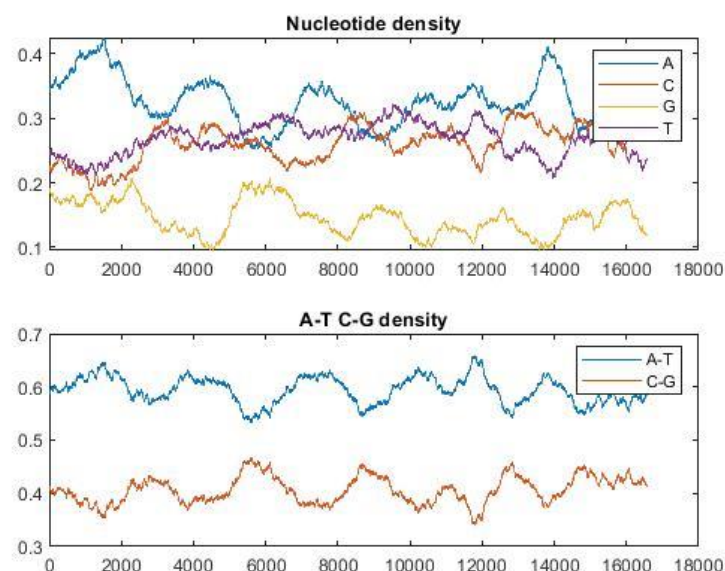
Figure 1: Distribution of nucleotide bases

	Density of bases	Percentage of bases
A	5401	32.6%
C	4310	26.0%
G	2414	14.6%
T	4440	26.8%

Distribution of Nucleotide bases for Walrus mitochondrion



As shown in Figure 1 above, we can see that the distribution of T and C are relatively the same whereas there are twice as many A's as there are G's. This is an interesting change from the norm of having a similar number of complementary pairs (A & T and C & G). To further investigate the local fluctuations in the nucleotide frequency a graph is plotted.



Plotting the nucleotide density shows clear peaks and troughs indicating the nucleotides are not drawn from an independent and evenly distributed probability distribution as the distribution is seen to be changing along the genome. Plotting the GC-content further verifies this statement as it displays six distinct waves of variation in both the AT and GC content.

Next, we employ a method to find protein coding regions. This is done by identifying all Open Reading frames (ORFs) within a certain threshold. Carefully setting the genetic code for Vertebrate Mitochondria and arbitrarily picking a threshold value of 75 bp, we find 13 protein coding regions. A simple CDS search of the whole genome confirms this find by revealing the 13 proteins present within the genome. Out of these, we choose and translate CYTB and COX2 to continue our evolutionary investigation.

Using the protein blast on NCBI, we find the closest related species containing the same proteins.

Results of CYTB Blast

Table 2: Closest related species found by CYTB blast

Rank	Latin Name	Common Name	Genetic Family	Total Score (Out of 760)
1	Halichoerus Grypus	Grey Seal	Phocidae	681
2	Gulo Gulo	Wolverine	Mustelidae	680
3	Phtulina Vitulina Stejnegeri	Harbour Seal	Phocidae	679
4	Erignathus Barbatus	Bearded Seal	Phocidae	679
5	Mustela Frenata	Long-tailed weasel	Mustelidae	674

Looking at some of the closest relatives found (Displayed in Table 2), the lack of *Otariidae*s (Sea lion and fur seals) suggests an interesting diphyletic origin of *Pinnipedia* with *Odobenidae*, *Phocidae* and *Mustelids* on one branch with *Otariidae*s on another.

Similarly, we used CYTC to find the closest relatives and reinforce what was found from CYTB.

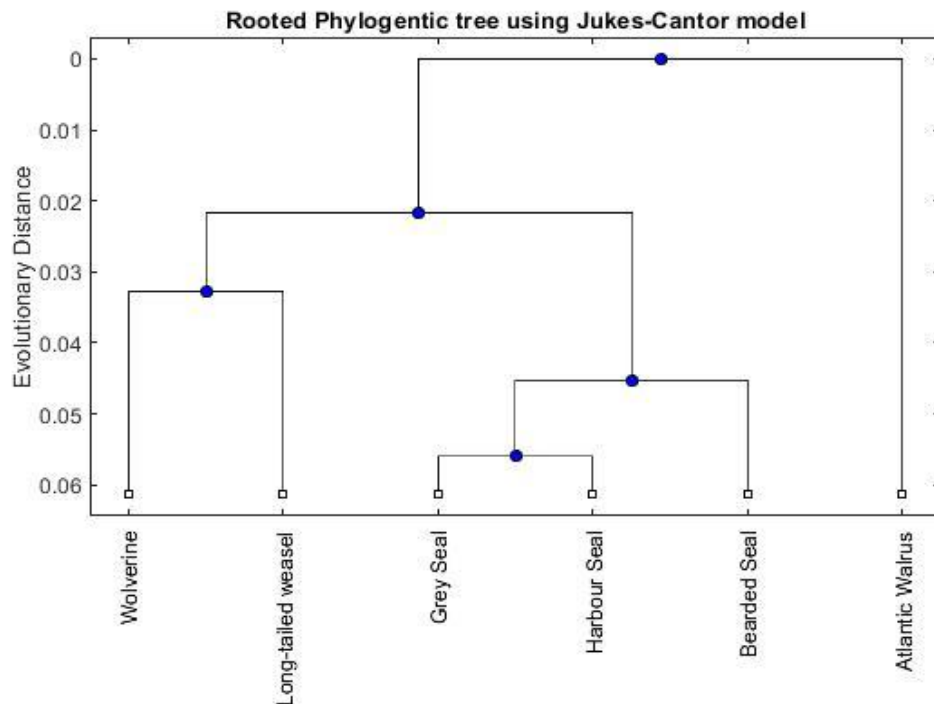
Results of CYTC Blast

Table 3: Closest species found through CYTC blasting

Rank	Latin Name	Common Name	Genetic Family	Total Score (Out of 466)
1	Tremarctos Ornatus	Spectacled Bear	Caniformia	447
2	Otaria byroni	South American Sea Lion	Otariidae	446
3	Arctocephalus Townsendi	Guadalupe Fur Seal	Otariidae	445
4	Neophoca Cinerea	Australian sea Lion	Otariidae	444
5	Callorhinus ursinus	Northern fur Seal	Otariidae	444

The results found by analysing and comparing the closest species through CYTC blasting does not support the results found through CYTB blasting. The lack of *Phocidae* and *Mustelids* results in two plausible theories; A monophyletic origin or a dyphyletic origin with *Otariidae* and *Odobenidae* on the same branch. This means we need to further analyse the genome to uncover evolutionary history to find the true phylogenetics.

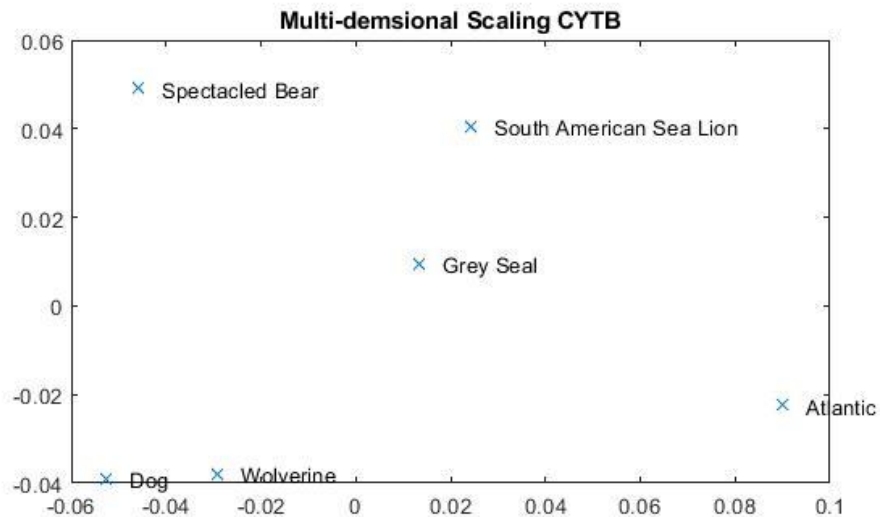
Generating a rooted phylogenetic tree for the organisms found by blasting CYTB we can discover the ancestral background of the species.



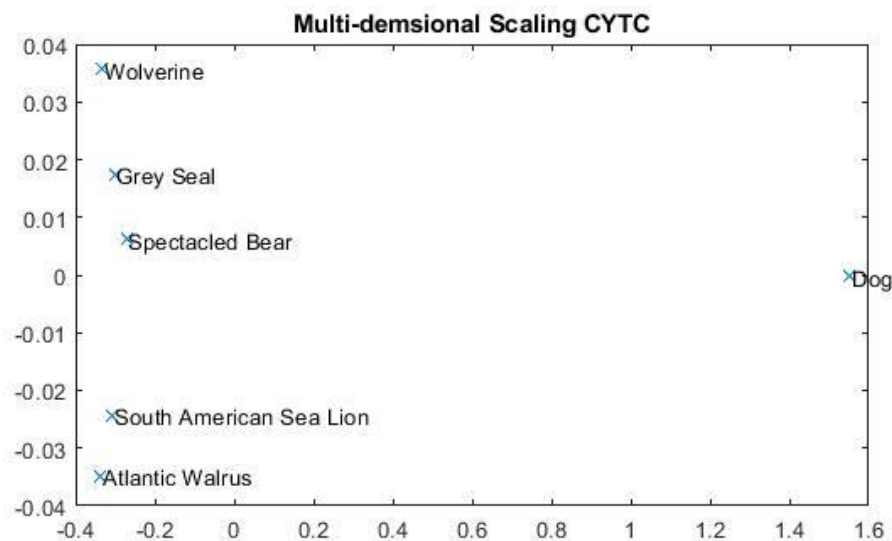
Using the Jukes-Cantor method, we have created the rooted Phylogenetic tree above based on the genetic distance between selected species. Analysing the tree shows us that the Atlantic Walrus (*Odobenidae*) stems from the original *pinniped* family and is not an ancestral family to the others. But an interesting find is that the *Phocidae* and the *Mustelids* are more closely related to each other than to the *Odobenidae* with both branching out from a single root. Within the family of the *Phocidae* we find a subdivision between the bearded seal from the grey seal and the harbour seal, which are on the same subbranch.

Since we had only used the results found from our CYTB blast, this leaves room for further investigation into the species found by our CYTC blast which contradicted each other. Computing the Genetic distance, using the Jukes-Cantor method again, between selected species reported in the tables above, and others we would like to investigate, would help us better understand the evolutionary history and relation of these animal families.

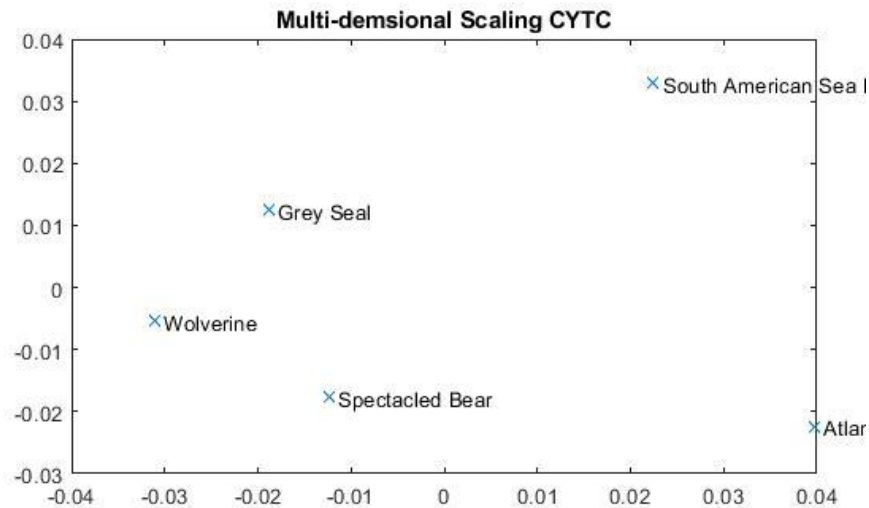
We found the distance between each CYTB and CYTC protein of each species in both the Amino Acid and Nucleotide sequence (Shown in Matlab Code). But to visualize the distance between species, we use multidimensional scaling to find the distance between all the CYTB and CYTC, results of which are shown below.



The additional species chosen was a dog (from the family *Canidae*). Looking at the image. We see that the dog and wolverine seem to be the closest in relation. With the sea lion and grey seal also seeming close. The Atlantic Walrus seems almost separate from the group with the grey seal almost mid way between the sea lion and walrus. The spectacled bear seems only somewhat close to the sea lion.

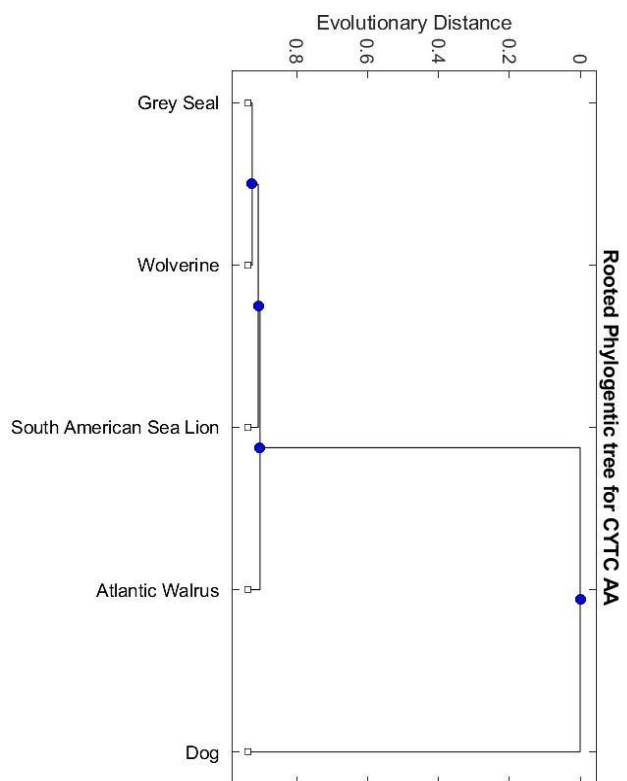
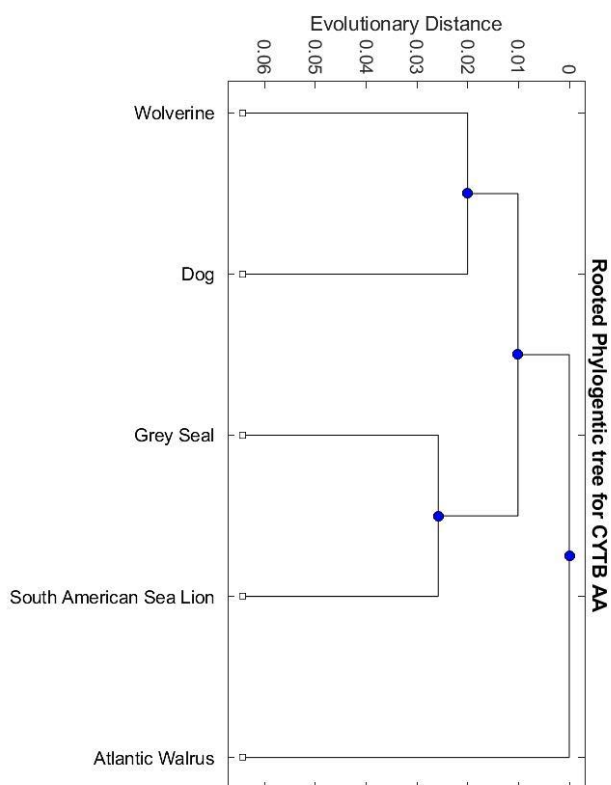


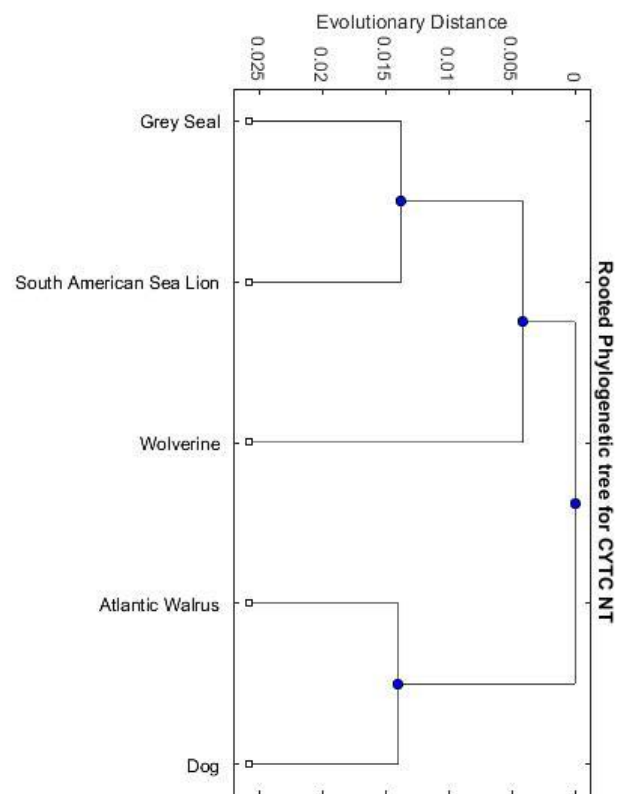
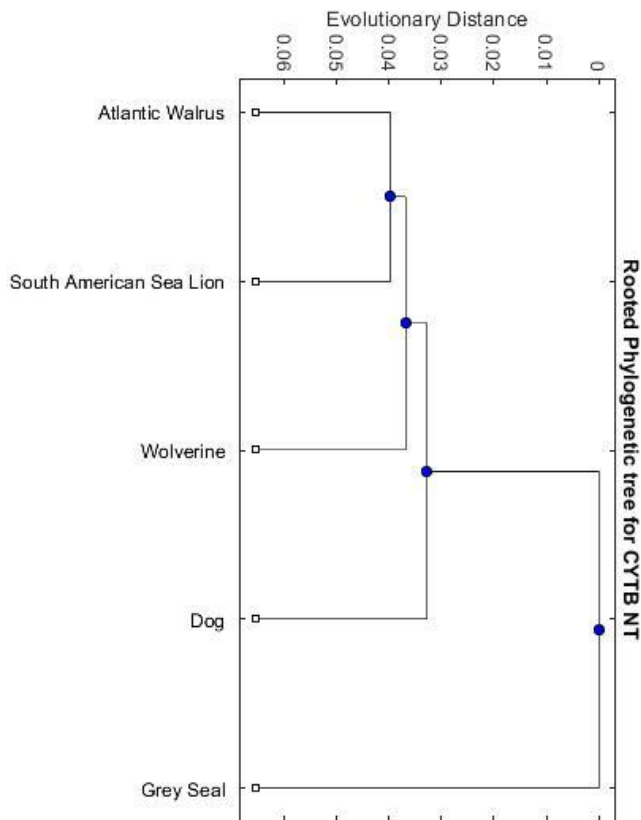
The results found from the multidimensional scaling of the CYTC show contradicting results, with the dog an outlier. This graph appears to relate the walrus closer to the sea lion than the grey seal. With the grey seal close to the spectacled bear and wolverine.



Removing the outlier, we can see some consensus in that the Atlantic walrus is once again a separate entity. While compared to CYTB, the grey seal seems closer to the wolverine than the sea lion, although it is still closest out of all with the sea lion almost an outlier as well. This was something we witnessed earlier with the phylogenetic tree connecting the seals (*Phocidae*) and the wolverine (*Mustelids*) to a single ancestor.

Next forming phylogenetic trees at both the amino acid and nucleotide level we can possibly find some consensus within the data.





The results, once again, present a confusing picture with contradictory patterns within each tree. (The spectacled bear was removed from this study since we could not find the CYTC NT.) There are some similarities across some of the trees. Like the CYTC NT, CYTC AA and CYTB AA all show the *Phocidae* and the *Mustelids* being closer in relation to each other than to the *Odobenidae*. CYTC AA tree supports the notion that the *Odobenidae* are an earlier ancestor of the Pinnipeds. But both the CYTC NT and the CYTB AA tree show the *Odobenidae* stemming from a different branch all together. There seems to be a major contrast in the placement of the Dog (*Canidae*), but can be assumed to be closely related to the *Mustelids* based on the evidence presented by Castresana (J., 2001) presenting research to suggest CYTB is more reliable for contrasting trees at the genus and family level, and indicate the true phylogeny. All these results seem to have contrast results from the study done by (Andersen, et al., 1998) which shows an early split from the *Phocidae* and later split between the *Odobenidae* and *Otariidae*. (Lento, 1995) however, does present offer that could be seen somewhat consensual to our results with the *Odobenidae* being an early split from the common ancestor, *pinniped*.

The Online tax browser (NCBI) shows the *Odobenidae*, *Otariidae*, *Phocidae* and *Mustelids* as distinct families within the *Caniformia* and no group being an ancestor to another.

Conclusion

The evolutionary history of the Odobenidae seems to be far more complicated than expected, with few similarities found to previous studies, a variety of hypothesis still in contention. It can be determined that further in-depth work needs to be done to find the true phylogeny of the Odobenidae with reliable results. The multiple alignment performed is shown in Appendix A.

Multiple Alignment Performed.

PYYTIKIDILGLIILILILMLLVLFSPDLLGDPDNYTPANPLSTPPH IKPEWYFFLAYAILRSIP
 PYYTIKIDILGALLILVLTLVLVFSPLLGLGDPDNYTPANPLSTPPH IKPEWYFFLAYAILRSIP
 PYYTIKIDILGALFLAIVLVMVLVLFSPDLLGDPDNYTPANPLSTPPH IKPEWYFFLAYAILRSIP
 PYYTIKIDILGALLLILMSLVLFSPDLLGDPDNYTPANPLNTPPH IKPEWYFFLAYAILRSIP
 PYYTIKIDILGVLLILALVTLVLVFSPLLGLGDPDNYTPANPVS TPPH IKPEWYFFLAYAILRSIP

Below are all the accession numbers used for each type of protein and protein coding sequence.

%CYTB AA results

% Species Description	Accession Number
'Atlantic Walrus'	'NP_659349';
'Grey Seal'	'ACZ28998';
'Wolverine'	'YP_001382271';
'Dog'	'ABY61049';
'Spectacled Bear'	'AAB50570';
'South American Sea Lion'	'AAQ95107';

%CYTC AA results

%Species Description	Accession Number
'Atlantic Walrus'	'NP_659340.3';
'Grey Seal'	'NP_007072';
'Wolverine'	'YP_001382262';
'Dog'	'NP_001183974';
'Spectacled Bear'	'YP_001542732';
'South American Sea Lion'	'AAR00312';

% CYTB NT Results

%Species Description	Accession Number
'Atlantic Walrus'	'X82299';
'Grey Seal'	'GU733679';
'Wolverine'	'DQ206375';
'Dog'	'EU352854';
'Spectacled Bear'	'U23554.1';
'South American Sea Lion'	'AY377328.1';

%Results for CYTC NT

%Could not find spectacled bear CYTC NT on NCBI

%Species Description	Accession Number
'Atlantic Walrus'	'AY377171';
'Grey Seal'	'GU733706';
'Wolverine'	'AY377174';
'Dog'	'NM_001197045.1';
'South American Sea Lion'	'AY377172.1';

Bibliography

Andersen, L. W. et al., 1998. Population Structure and gene flow of the Atlantic Walrus (*Odobenus rosmarus rosmarus*) in the eastern Atlantic Arctic based on mitochondrial DNA and microsatellite variation. In: *Molecular Ecology Vol.7*. s.l.:s.n., pp. 1323-1336.

J., C., 2001. Cytochrome b Phylogeny and the Taxonomy of Great Apes and Mammals . In: *Molecular biology and Evolution(18)*. s.l.:s.n., pp. 465-471.

Lento, G. & H. R. & C. G. & P. D., 1995. Use of Spectral Analysis to Test Hypotheses on the Origin of Pinnipeds. In: *Molecular Biology and Evolution Vol.12*. s.l.:s.n., pp. 28-52.

