# *Computational Genomics and Bioinformatics*

# *Final Assessed Coursework 2021*

*Sam Spedding, Student Number 1700525*

## Is *Orcinus orca,* also known as the killer whale, a whale, a dolphin, or something else? What is its closest cousin?

ABSTRACT: A collection of dolphin and whale species, alongside an outgroup, is used to determine whether *Orcinus orca*, is a dolphin or a whale. The results show it is more closely related to the species of dolphin despite its common name 'killer whale'. Further, we attempt to determine which of the dolphin species *Orcinus orca* is closest to. Phylogenetic analysis of CYTB and COX1 proteins shows *Lagenorhynchus obliquidens* (Pacific white-sided dolphin) is the closest relative, and this agrees with the literature.

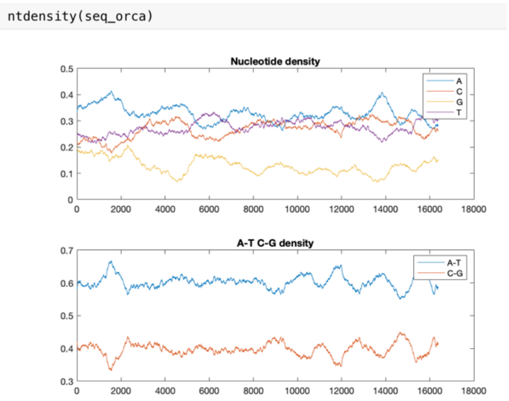
## Task 1

Here we study the killer whale, Latin name *Orcinus orca*. The accession number of the complete mitochondrial DNA (mtDNA) nucleotide sequence is `NC_023889`.

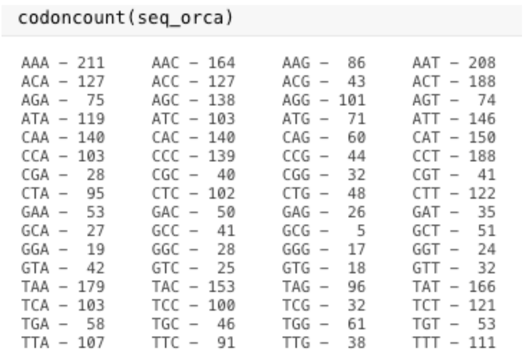
```
seq_orca = getgenbank('NC_023889','sequenceonly','true');
```

Some basic statistical descriptions and features of the data:

- The length of the sequence is 16,386 base pairs.
- The nucleotide density distribution and codon count are as follows:
- `basecount(seq_orca)` tells us there are 5384 A's, 4399 C's, 2106 G's and 4496 T's.
- We also check the nucleotide and AT-CG density distributions, and codon counts:



```
ntdensity(seq_orca)
```



```
codoncount(seq_orca)

AAA - 211     AAC - 164     AAG -  86     AAT - 208
ACA - 127     ACC - 127     ACG -  43     ACT - 188
AGA -  75     AGC - 138     AGG - 101     AGT -  74
ATA - 119     ATC - 103     ATG -  71     ATT - 146
CAA - 140     CAC - 140     CAG -  60     CAT - 150
CCA - 103     CCC - 139     CCG -  44     CCT - 188
CGA -  28     CGC -  40     CGG -  32     CGT -  41
CTA -  95     CTC - 102     CTG -  48     CTT - 122
GAA -  53     GAC -  50     GAG -  26     GAT -  35
GCA -  27     GCC -  41     GCG -   5     GCT -  51
GGA -  19     GGC -  28     GGG -  17     GGT -  24
GTA -  42     GTC -  25     GTG -  18     GTT -  32
TAA - 179     TAC - 153     TAG -  96     TAT - 166
TCA - 103     TCC - 100     TCG -  32     TCT - 121
TGA -  58     TGC -  46     TGG -  61     TGT -  53
TTA - 107     TTC -  91     TTG -  38     TTT - 111
```

We detect the open reading frames above a minimum length threshold given by the 99th percentile for the lengths of ORFs generated by randomly permuting the mtDNA sequence. We calculate the upper 1% threshold to be 186. Then we search for ORFs in the original sequence as follows:

```
sig_level = 0.01;
critical_value = prctile(ORF_lengths_perm, 100*(1-sig_level))
```

```
critical_value = 186
```

```
% ORF lengths for orca:

ORFs_orca = seqshoworfs(seq_orca, 'MinimumLength', critical_value, ...
                        'nodisplay','true', 'GeneticCode', 2, 'frames', 'all');
ORF_lengths_orca = [];
for frame = 1:6
    start = ORFs_orca(frame).Start;
    stop = ORFs_orca(frame).Stop;
    start = start(1:length(stop));

    ORF_lengths_orca = [ORF_lengths_orca; (stop - start)'];
end
```

How many ORFs are there above the critical length?

```
length(ORF_lengths_orca)
```

```
ans = 9
```

Even the shortest has length well above the threshold:

```
min(ORF_lengths_orca)
```

```
ans = 678
```

We use genbank to identify two protein-coding genes. The Cytochrome B (CYTB, genbank Gene ID: 18982992) sits in the base pair range 14192 – 15331, which we extract and translate to amino acid. The Cytochrome C Oxidase Subunit I (COX1, genbank Gene ID: 18982982) sits in the base pair range 5360 – 6910, which we also extract and translate:

```
CYTB_orca_NT = seq_orca(14192:15331);
CYTB_orca_AA = nt2aa(CYTB_orca_NT, 'ACGTOnly', 'false');
COX1_orca_NT = seq_orca(5360:6910);
COX1_orca_AA = nt2aa(COX1_orca_NT, 'ACGTOnly', 'false');
```

## Task 2

BLASTing the translated CYTB, of the top 25 results, 19 are of the same species *Orcinus orca*. Also present lower down the list, in decreasing order of alignment score, are the species *Stenella longirostris* (spinner dolphin), *Lagenorhynchus obliquidens* (Pacific white-sided dolphin), *Lagenorhynchus obscurus* (dusky dolphin), *Tursiops truncatus* (common bottlenose dolphin), and *Delphinus delphis* (saddleback dolphin), all of which have a 99% query coverage. This is strong evidence to suggest that *Orcinus orca*, despite its common name, ‘killer whale’, is more closely related to dolphins than it is to whales.

We import CYTB and COX1 proteins from genbank just as we did in task 1:

```
% Stenella longirostris:
seq_dolphin1 = getgenbank('NC_032301','sequenceonly','true');
CYTB_dolphin1_NT = seq_dolphin1(14185:15324);
COX1_dolphin1_NT = seq_dolphin1(5353:6903);
```

and the rest are imported identically. We then store these in a data structure:

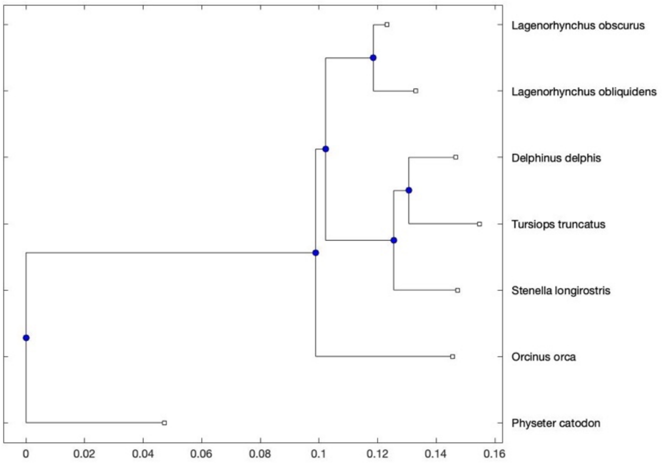
```
dolphins_struct.CYTB = {CYTB_orca_NT, CYTB_dolphin1_NT, CYTB_dolphin2_NT, ...
                        CYTB_dolphin3_NT, CYTB_dolphin4_NT, CYTB_dolphin5_NT, CYTB_whale_NT};
dolphins_struct.Names = {'Orcinus orca', 'Stenella longirostris', 'Lagenorhynchus obliquidens', ...
                        'Lagenorhynchus obscurus', 'Tursiops truncatus', 'Delphinus delphis', 'Physeter catodon'};
```

## Task 3

Here we generate a rooted phylogenetic tree for the above species of dolphin, as well as the killer whale. We also include *Physeter catodon* (sperm whale) as an outgroup and re-root the tree by this outgroup species. We use the following commands:

```
distance_matrix = squareform(seqpdist(dolphins_struct.CYTB, 'method', 'jukes-cantor'));
tree = seqneighjoin(distance_matrix, 'equivar',dolphins_struct.Names);
tree = reroot(tree, 7);
phytreeviewer(tree)
```

which gives the following for the phylogenetic tree:



We see that the species closest to *Orcinus orca* according to the tree is *Stenella longirostris*, which was the closest on the list according to the results from BLAST. P*hyseter catodon*, which was used as an outgroup to re-root the tree, is the furthest from *Orcinus orca* and is the only non-dolphin of the group. This suggests that *Orcinus orca* shares more evolutionary history with dolphins than it does with *Physeter catodon*, which is the species that is genetically furthest from the group. However, the other dolphin species form their own **monophyletic group**, or **clade**, which suggests *Orcinus orca* may have diverged from the other dolphins at some point in their evolutionary history. It makes sense at this point to broaden our study to include other whales and dolphins to determine which species *Orcinus orca* is closest to.

## Task 4

We now consider a wider variety of aquatic mammals. We include *Balaenoptera musculus* (blue whale), *Monodon monoceros* (narwhal), *Delphinapterus leucas* (beluga whale), and as an outgroup we use *Odobenus rosmarus rosmarus* (Atlantic walrus), which we will abbreviate by dropping a *rosmarus*. We leave behind *Lagenorhynchus obscurus* and *Physeter catodon* at this point because these organisms do not have COX1 proteins available on genbank. We again use `getgenbank` to import the relevant proteins and store them in a data structure called `species_struct`, just as we did in Task 3.

When importing the full mitochondrial genome from genbank, there is a breakdown of each protein included in the genome. We use this to locate the CYTB and COX1 proteins for each of our species (we now have nine) just as we did in Task 1. We use this method for finding the proteins as it is the most efficient and reliable method for species with full mtDNA sequences available. For each of these two protein-coding sequence types, we compute a pairwise genetic distance matrix between every species, using the Jukes-Cantor correction method to account for the possibility of multiple substitutions at the at the same nucleotide position between species.

```
CYTB_dist_matrix = squareform(seqpdist(species_struct.CYTB_NT, 'method', 'jukes-cantor'));
COX1_dist_matrix = squareform(seqpdist(species_struct.COX1_NT, 'method', 'jukes-cantor'));
```

We use the `squareform` function to create a symmetric square-format matrix *Z* such that *Z(i,j)* represents the distance between genetic code strings *i* and *j*. The results are as follows.

CYTB distance matrix:

| Orcinus orca | Stenella longirostris | Lagenorhynchus obliquidens | Tursiops truncatus | Delphinus delphis | Balaenoptera musculus | Monodon monoceros | Delphinapterus leucas | Odobenus rosmarus |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.0969 | 0.0796 | 0.1057 | 0.093 | 0.1886 | 0.1396 | 0.1406 | 0.2889 |
| 0.0969 | 0 | 0.0777 | 0.0486 | 0.0458 | 0.1748 | 0.1386 | 0.1498 | 0.283 |
| 0.0796 | 0.0777 | 0 | 0.0853 | 0.0767 | 0.1706 | 0.1185 | 0.1315 | 0.2771 |
| 0.1057 | 0.0486 | 0.0853 | 0 | 0.0403 | 0.1664 | 0.1325 | 0.1529 | 0.2783 |
| 0.093 | 0.0458 | 0.0767 | 0.0403 | 0 | 0.1748 | 0.1376 | 0.1498 | 0.283 |
| 0.1886 | 0.1748 | 0.1706 | 0.1664 | 0.1748 | 0 | 0.1801 | 0.1737 | 0.2678 |
| 0.1396 | 0.1386 | 0.1185 | 0.1325 | 0.1376 | 0.1801 | 0 | 0.072 | 0.2818 |
| 0.1406 | 0.1498 | 0.1315 | 0.1529 | 0.1498 | 0.1737 | 0.072 | 0 | 0.3033 |
| 0.2889 | 0.283 | 0.2771 | 0.2783 | 0.283 | 0.2678 | 0.2818 | 0.3033 | 0 |

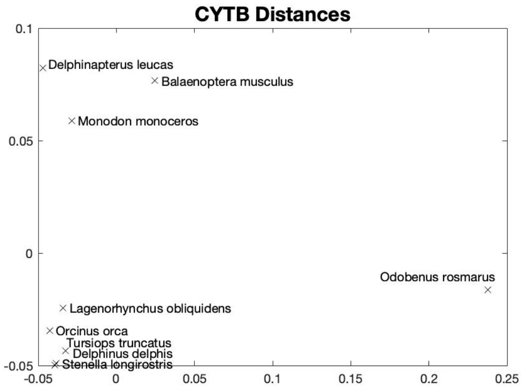
COX1 distance matrix:

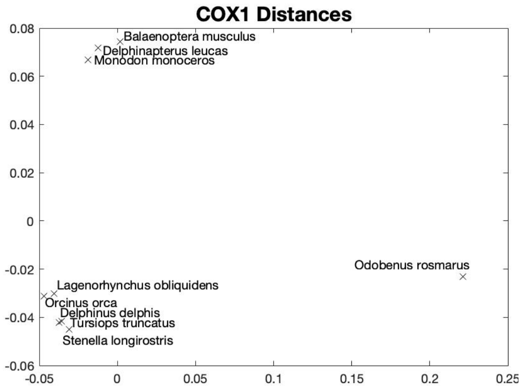
| Orcinus orca | Stenella longirostris | Lagenorhynchus obliquidens | Tursiops truncatus | Delphinus delphis | Balaenoptera musculus | Monodon monoceros | Delphinapterus leucas | Odobenus rosmarus |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.0786 | 0.0675 | 0.0779 | 0.0654 | 0.1621 | 0.1326 | 0.1402 | 0.2638 |
| 0.0786 | 0 | 0.0612 | 0.0321 | 0.0248 | 0.1689 | 0.1319 | 0.1326 | 0.2497 |
| 0.0675 | 0.0612 | 0 | 0.0626 | 0.0537 | 0.1582 | 0.1253 | 0.1326 | 0.2568 |
| 0.0779 | 0.0321 | 0.0626 | 0 | 0.0242 | 0.1582 | 0.1312 | 0.1438 | 0.2564 |
| 0.0654 | 0.0248 | 0.0537 | 0.0242 | 0 | 0.1605 | 0.1275 | 0.1356 | 0.253 |
| 0.1621 | 0.1689 | 0.1582 | 0.1582 | 0.1605 | 0 | 0.162 | 0.1674 | 0.2568 |
| 0.1326 | 0.1319 | 0.1253 | 0.1312 | 0.1275 | 0.162 | 0 | 0.0702 | 0.2573 |
| 0.1402 | 0.1326 | 0.1326 | 0.1438 | 0.1356 | 0.1674 | 0.0702 | 0 | 0.2532 |
| 0.2638 | 0.2497 | 0.2568 | 0.2564 | 0.253 | 0.2568 | 0.2573 | 0.2532 | 0 |

We can use the command `cmdscale` to visualise these distances as follows:

```
distance_matrices = {CYTB_NT_dist_matrix, COX1_NT_dist_matrix};
titles = {'CYTB Distances', 'COX1 Distances'};

for m = 1:2
    ax(m) = nexttile(tlo);
    points = cmdscale(distance_matrices{m});
    plot(ax(m), points(:,1),points(:,2), 'x', 'color', 'k')
    xlimits=get(ax(m),'xlim');
    text(ax(m), points(:,1) + (xlimits(2)-xlimits(1))/75, points(:,2), species_struct.Names)
    title(ax(m), titles{m}, 'FontSize', 16)
end
```





We see that, for the distances associated with both protein-coding regions, the whales and dolphins are in two distinct groups, with both the CYTB and COX1 proteins for the *Orcinus orca* indisputably **orthologous** to those of each species of dolphin. The outgroup species *Odobenus rosmarus* used as a control sits distinctly separate from both groups of species.

The species most closely related to *orcinus orca* is not as clear-cut between the two proteins, with *Lagenorhynchus obliquidens* being closest in its CYTB protein, while *Delphinus delphis* is closest according to COX1 (highlighted in yellow in the matrix tables), although the differences are marginal in each. Furthermore, those same two species score second closest in the protein in which they aren't the closest. It appears that *orcinus orca* is very closely related to more than one species of dolphin, and we can conclude that it is indeed a dolphin.

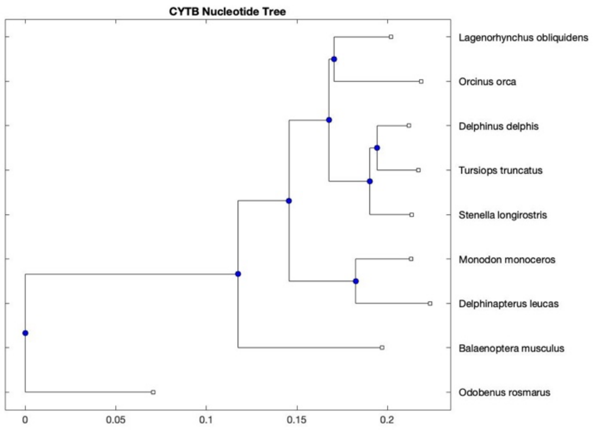
**We now ask the question: which, out of *Lagenorhynchus obliquidens* and *Delphinus delphis*, is *Orcinus orca* most closely related to?**

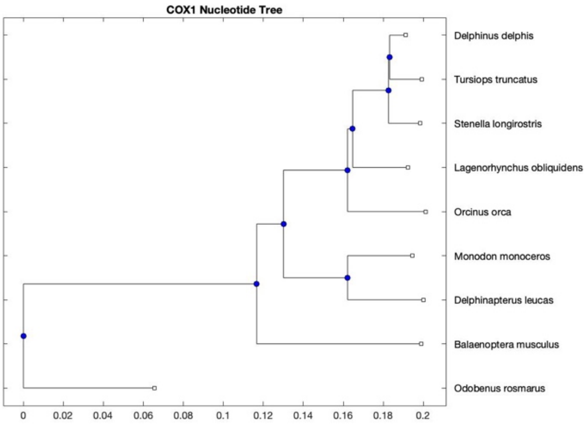
## Tasks 5 and 6

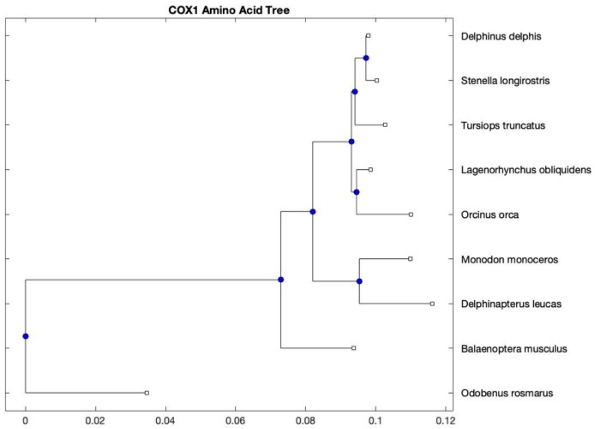
We compute phylogenetic trees for the CYTB and COX1 nucleotide sequences, and also for the CYTB and COX1 sequences translated into amino acid, and reroot each by the outgroup in each case. The code for CYTB is shown below:

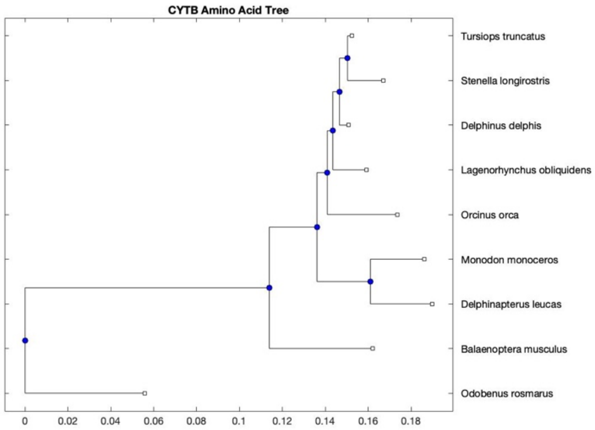
```
% Translate to amino acids:
for s = 1:n_species
    species_struct.CYTB_AA{s} = nt2aa(species_struct.CYTB_NT{s}, 'ACGTOnly', 'false');
end
% tree for NT CYTB:
tree_CYTB_NT = seqneighjoin(CYTB_NT_dist_matrix, 'equivar',species_struct.Names);
tree_CYTB_NT = reroot(tree_CYTB_NT, outgroup_num);
% tree for AA CYTB:
CYTB_AA_dist_matrix = squareform(seqpdist(species_struct.CYTB_AA, 'method', 'jukes-cantor'));
tree_CYTB_AA = seqneighjoin(CYTB_AA_dist_matrix, 'equivar',species_struct.Names);
tree_CYTB_AA = reroot(tree_CYTB_AA, outgroup_num);
```

The results are as follows:


CYTB Nucleotide Tree


COX1 Nucleotide Tree


COX1 Amino Acid Tree


CYTB Amino Acid Tree

We see that the bottom four species are in the same place and with the same structure in each tree. The outgroup *Odobenus rosmarus* is distinctly separated from the other groups, while *Balaenoptera musculus*, *Monodon monoceros*, and *Delphinapterus leucas*, being the three other non-dolphin species, are consistently in a separate clade. The remaining five species also consistently form their own clade across the four trees, although the structure within that clade differs between each tree. This is additional strong evidence to suggest that *Orcinus orca* is a dolphin, but the question of whether it is closer to *Lagenorhynchus obliquidens* or *Delphinus delphis* it is closer to remains unanswered.

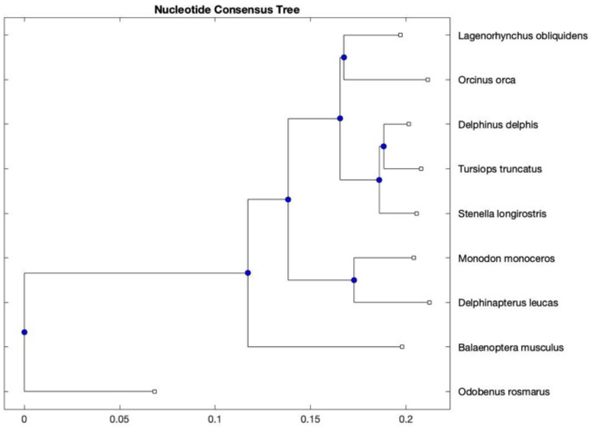
We now generate two **consensus trees**, one for nucleotides and one for amino acids, by performing a weighted average of the genetic distance matrices used to generate the trees for each protein-coding region, as described in [1].

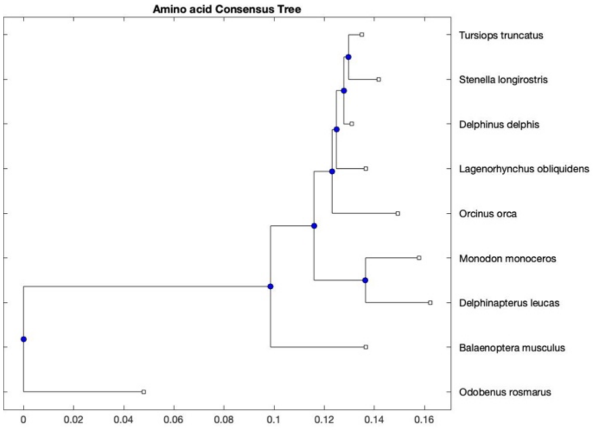
```
% consensus tree NT:
weights_NT = [sum(CYTB_NT_dist_matrix, 'all'), sum(COX1_NT_dist_matrix, 'all')];
weights_NT = weights_NT/sum(weights_NT);
consensus_NT_dist_matrix = weights_NT(1)*CYTB_NT_dist_matrix ...
                         + weights_NT(2)*COX1_NT_dist_matrix;
tree_consensus_NT = seqneighjoin(consensus_NT_dist_matrix, 'equivar',species_struct.Names);
tree_consensus_NT = reroot(tree_consensus_NT, outgroup_num);


% consensus tree AA:
weights_AA = [sum(CYTB_AA_dist_matrix, 'all'), sum(COX1_AA_dist_matrix, 'all')];
weights_AA = weights_AA/sum(weights_AA);
```

```
consensus_AA_dist_matrix = weights_AA(1)*CYTB_AA_dist_matrix ...
                         + weights_AA(2)*COX1_AA_dist_matrix;
tree_consensus_AA = seqneighjoin(consensus_AA_dist_matrix, 'equivar',species_struct.Names);
tree_consensus_AA = reroot(tree_consensus_AA, outgroup_num);
```

The results are as follows:





From the nucleotide consensus tree the species most closely related to *Orcinus orca* is *Lagenorhynchus obliquidens*, since they lie separate from the other dolphins in their own clade. However, in the amino acid consensus tree the structure is not as clear; here *Orcinus orca* is the first to branch from the other dolphins, which form their own subclade of four. However, of these four, again *Lagenorhynchus obliquidens* comes out as closest. **We make a preliminary conclusion that *Lagenorhynchus obliquidens* is the closest relation (of the dolphins we are studying) to *Orcinus orca*.**

For our dolphins of the family Delphinidae, the phylogenetic structures break down at the genus level. *Orcinus orca* is the sole (extant) member of the genus *Orcinus*, while the other dolphins are each from separate genera, as is clear from their Latin names. Thus, we cannot compare the above result to the genetic trees found on the Taxonomy Browser [2].

However, the main branches of our consensus trees we have found are in line with the phylogenetic relations given on Taxonomy Browser **up to and including the rank of family**. Our five dolphin species (including *Orcinus orca*) are all members of the family Delphinidae; *Monodon monoceros* and *Delphinapterus leucas* (despite its genus name) are of the family Monodontidae; and *Balaenoptera musculus* is a member of the family Balaenopteridae. The outgroup is from the order Carnivora, which contains the placental flesh-eating mammals, such as cats, dogs and bears, while the other species we are studying hail from the order Artiodactyla, which contains whales, dolphins, hippos and pigs [2].

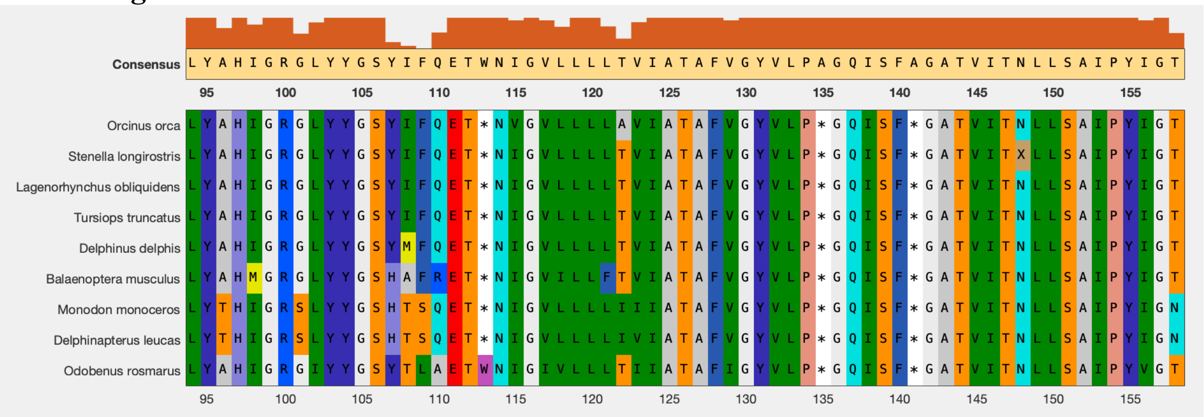
## Task 7

We perform two multiple alignments across the nine species for the amino acid sequences (since these are a third of the length of the DNA sequences so polymorphic sites are easier to spot) of each protein using the command `multialign`, as seen below.

```
for s = 1:n_species
    CYTB_struct_vector(s) = struct('Sequence',species_struct.CYTB_AA{s},...
        'Header',species_struct.Names{s});
    COX1_struct_vector(s) = struct('Sequence',species_struct.COX1_AA{s},...
        'Header',species_struct.Names{s});
end
ma_CYTB = multialign(CYTB_struct_vector);
seqalignviewer(ma_CYTB, 'alphabet', 'AA');
ma_COX1 = multialign(COX1_struct_vector);
seqalignviewer(ma_COX1, 'alphabet', 'AA');
```

The brown bar chart at the top shows the level of homeomorphism across all species, and we see from inspection that in the majority of positions all nine species are perfectly aligned. We also see that the majority of polymorphisms occur in the final row, which represents the outgroup species. Screenshots of some interesting points from the alignment of each protein are shown below. We use the RASMOL amino acid colour scheme which uses a larger range of colours than the default setting, so a more precise level of alignment may be inferred.

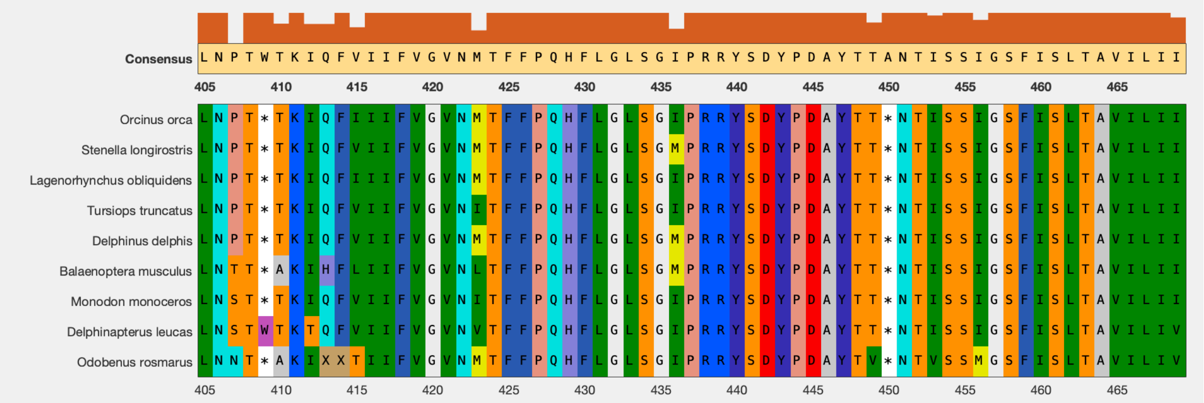
**CYTB alignment:**



```
Consensus                   L Y A H I G R G L Y Y G S Y I F Q E T W N I G V L L L L T V I A T A F V G Y V L P A G Q I S F A G A T V I T N L L S A I P Y I G T
Orcinus orca                L Y A H I G R G L Y Y G S Y I F Q E T * N V G V L L L L A V I A T A F V G Y V L P * G Q I S F * G A T V I T N L L S A I P Y I G T
Stenella longirostris       L Y A H I G R G L Y Y G S Y I F Q E T * N I G V L L L L T V I A T A F V G Y V L P * G Q I S F * G A T V I T X L L S A I P Y I G T
Lagenorhynchus obliquidens  L Y A H I G R G L Y Y G S Y I F Q E T * N I G V L L L L T V I A T A F V G Y V L P * G Q I S F * G A T V I T N L L S A I P Y I G T
Tursiops truncatus          L Y A H I G R G L Y Y G S Y I F Q E T * N I G V L L L L T V I A T A F V G Y V L P * G Q I S F * G A T V I T N L L S A I P Y I G T
Delphinus delphis           L Y A H I G R G L Y Y G S Y M F Q E T * N I G V L L L L T V I A T A F V G Y V L P * G Q I S F * G A T V I T N L L S A I P Y I G T
Balaenoptera musculus       L Y A H M G R G L Y Y G S H A F R E T * N I G V I L L F T V I A T A F V G Y V L P * G Q I S F * G A T V I T N L L S A I P Y I G T
Monodon monoceros           L Y T H I G R S L Y Y G S H T S Q E T * N I G V L L L L I I I A T A F V G Y V L P * G Q I S F * G A T V I T N L L S A I P Y I G N
Delphinapterus leucas       L Y T H I G R S L Y Y G S H T S Q E T * N I G V L L L L I V I A T A F V G Y V L P * G Q I S F * G A T V I T N L L S A I P Y I G N
Odobenus rosmarus           L Y A H I G R G I Y Y G S Y T L A E T W N I G I V L L L T I I A T A F I G Y V L P * G Q I S F * G A T V I T N L L S A I P Y V G T
```

There is a polymorphism in position 107 in which we see that the five members of the family Delphinidae are aligned with each other and with the outgroup (from Y = TAC), while the other three Artiodactyla are aligned with each other (to H = CAC, making this a single nucleotide polymorphism, or **SNP**). In position 122, polymorphism occurs from the members of Monodontidae, and additionally from *Orcinus orca*, while the other species are aligned.

**COX1 alignment:**



```
Consensus                   L N P T W T K I Q F V I I F V G V N M T F F P Q H F L G L S G I P R R Y S D Y P D A Y T T A N T I S S I G S F I S L T A V I L I I
Orcinus orca                L N P T * T K I Q F I I I F V G V N M T F F P Q H F L G L S G I P R R Y S D Y P D A Y T T * N T I S S I G S F I S L T A V I L I I
Stenella longirostris       L N P T * T K I Q F V I I F V G V N M T F F P Q H F L G L S G M P R R Y S D Y P D A Y T T * N T I S S I G S F I S L T A V I L I I
Lagenorhynchus obliquidens  L N P T * T K I Q F I I I F V G V N M T F F P Q H F L G L S G I P R R Y S D Y P D A Y T T * N T I S S I G S F I S L T A V I L I I
Tursiops truncatus          L N P T * T K I Q F V I I F V G V N I T F F P Q H F L G L S G I P R R Y S D Y P D A Y T T * N T I S S I G S F I S L T A V I L I I
Delphinus delphis           L N P T * T K I Q F V I I F V G V N M T F F P Q H F L G L S G M P R R Y S D Y P D A Y T T * N T I S S I G S F I S L T A V I L I I
Balaenoptera musculus       L N T T * A K I H F L I I F V G V N L T F F P Q H F L G L S G M P R R Y S D Y P D A Y T T * N T I S S I G S F I S L T A V I L I I
Monodon monoceros           L N S T * T K I Q F V I I F V G V N I T F F P Q H F L G L S G I P R R Y S D Y P D A Y T T * N T I S S I G S F I S L T A V I L I I
Delphinapterus leucas       L N S T W T K T Q F V I I F V G V N V T F F P Q H F L G L S G I P R R Y S D Y P D A Y T T * N T I S S I G S F I S L T A V I L I V
Odobenus rosmarus           L N N T * A K I X X T I I F V G V N M T F F P Q H F L G L S G I P R R Y S D Y P D A Y T V * N T V S S M G S F I S L T A V I L I V
```

This is a slice in which the only two species for which the entire corresponding subsequences are 100% aligned are *Orcinus orca* and *Lagenorhynchus obliquidens* (first and third rows). This was the **longest subsequence between two species with no amino acid polymorphisms** I could find, and it runs between positions 194 – 487 out of a total length of 517. This is very strong evidence to support our claim that the closest relation to *Orcinus orca* is indeed *Lagenorhynchus obliquidens*.

## Task 8

Several studies have attempted to conform the species of the family Delphinidae. As discussed in [3], the assignment of subfamilies to group the constituent genera of the family into disjoint clades has been proposed multiple times but no consensus has been reached. However, the article presents a novel tree topology different from previous publications from which some intriguing conclusions may be inferred.

The article analyses complete mitochondrial genomes using the G-INS-i algorithm, a global pairwise adaptation of the Needleman-Wunsch algorithm, which contrasts with our method in which we only aligned parts of the mtDNA from the same two protein-coding regions. A global alignment method will provide more trustworthy results as the full genome is taken into consideration. In our method, the majority of the genome is neglected, and so a limitation of this method is that this could lead to results that differ depending on what protein-coding region we analyse; perhaps some proteins are more strongly aligned between species than others.

The phylogenetic tree shown in the article shows the species *Lagenorhynchus albirostris* diverged individually prior to the other members of Delphinidae, while other members of the same genus, namely *Lagenorhynchus cruciger*, *Lagenorhynchus australis* and *Lagenorhynchus obsurus* are contained in the proposed subfamily Lissodelphininae alongside *Orcinus orca*. Since our species in question, *Lagenorhynchus obliquidens*, was not included in the study, we are left with two possibilities: either *Lagenorhynchus obliquidens* diverged many years ago alongside its cousin *Lagenorhynchus albirostris*, or it is closely related to the other members of its genus, and thus its subfamily. Our findings from this report provide strong evidence for the latter, which in turn backs up our claim that *Lagenorhynchus obliquidens* is the closest relative of *Orcinus orca* from our chosen dolphin species.

The article goes on to say "The *Orcinus* group was a sister to the genetic clade including the genera *Cephalorhynchus* and *Lagenorhynchus*, which could be considered the Lissodelphininae subfamily." Comparing this with our consensus tree at the DNA sequence level (nucleotide sequences), this result is consistent with our findings. And while our amino acid consensus tree presents *Orcinus Orca* in a clade separate from the other dolphins, which contradicts this result, I would propose that, on the whole, our findings agree with the results of this article based on the analysis from the nucleotide consensus tree and multi-alignments studied in the previous task.

### References:

[1]: Mathworks Help Centre example*, Analyzing the Origin of the Human Immunodeficiency Virus*
https://uk.mathworks.com/help/bioinfo/ug/analyzing-the-origin-of-the-human-immunodeficiency-virus.html

[2]: National Centre for Biotechnology Information – Taxonomy Browser
https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi

[3]: Jose L. Horreo, *New insights into the phylogenetic relationships among the oceanic dolphins (Cetacea: Delphinidae)* – Department of Biodiversity and Evolutionary Biology, National Museum of Sciences (MNCN-CSIC), Madrid, Spain
https://doi.org/10.1111/jzs.12255