

Living members of the Order Carnivora are characterized by a particular arrangement of the cheekteeth dentition called the **Carnassial Pair**, abbreviated as P^4/M_1 , for the fourth upper premolar and first lower molar. In action, on either side of the jaw, the P^4 passes outside the M_1 , so as to create a shearing surface that assists in the processing of the primordial carnivore diet, flesh. The P^4/M_1 form has been variously modified in various lineages, including reduced dentition in the insectivorous aardwolf (*Proteles cristata*) or crabeater seals (*Lobodon carcinophaga*), or as a massive crushing surface in hyaenas. Other taxa with less dependence on flesh have more or less flattened crowns.

Living **Carnivora** occur in two recognizably distinct lineages, the dog-like **Caniformia / Canoidea** and the cat-like **Feliformia / Feloidea**. The former classically comprises seven families: **Canidae** (dogs), **Ursidae** (bears), **Mustelidae** (weasels & relatives), **Procyonidae** (raccoons & relatives), and the three marine forms, **Phocidae** (true or earless seals), **Otariidae** (eared seals), **Odobenidae** (walrus). The latter classically comprises three forms, **Felidae** (cats), **Hyaenidae** (hyaenas), and **Viverridae** (mongooses & civets). The last named was the least known, and has undergone extensive revision in the light of new molecular evidence. The Felidae as well have been controversial, as to whether (seeming) morphological uniformity is reflected in single genera of small (*Felis*) and large (*Panthera*), or some other arrangement. [Lawlor *et al.* 2021 provides a compact overview, Wilson & Reeder (2005) an exhaustive review].

In this lab we will apply three phylogenetic approaches to molecular data, in this case the complete **mitochondrial DNA (mtDNA)** protein coding regions (ca. 12 kbp) of multiple species of carnivores. The methods are those available through the program **MEGA (Molecular Evolutionary Genetics Analysis, 10th ed.)**, Neighbor Joining (**NJ**), Maximum Parsimony (**MP**), and Maximum Likelihood (**ML**). The first is not strictly speaking a phylogenetic method (as discussed in lecture), but will be referred to as such for uniformity.

The base data are arranged in a MEGA file of 13 protein coding regions from multiple carnivore species, so as to preserve the first **5'-3'** reading frame on the Sense (Heavy) Strand of the first region (**ND1**) consistent through the end of the last (**Cytb**), including reverse complements [Why?] of regions that occur on the Light Strand.

We will use these data and methods to explore several classic and contemporary problems of Carnivore systematics and taxonomy. Each exercise will involve a subset of the sequence in the base file, plus several sequences to be identified in and retrieved from the **GenBank** library of DNA sequence data as arranged taxonomically. The imported data will be imported into the MEGA files as above, and the analyses done so as to investigate the questions

Methods

1. Open **Genbank** at the [root of the Carnivora tree](#). Tick the box 'genomes' and click on 'Go'. The update screen will indicate the presence of a complete mtDNA genome and (or) a whole-genome project, as '1'.
 - a. Click on the '1': typically a window with 'See also 1 organelle- and plasmid-only records matching your search' will open. [If not, consult the instructors].
 - b. Click on that link. An **Organism Overview** will be presented, with a **RefSeq** link to a sequence file, typically beginning '**NC**' and followed by six digits. Click the link.
 - c. This will take you to a **Reference File** for the mtDNA genome of interest, annotated in GenBank format with coded descriptions of the genome contents. In the white panel to the upper right, click on '**send to**'. In the drop down menu choose '**Coding Sequences**' then '**Create File**'
 - d. A new '**Downloads**' menu will open and ask what you want to do. Click on '**Save as**', which will bring up a Windows box with '**sequence(1).txt**'. OVERWRITE this with the NC123456 code and the genus and species name of the genome source: SPELL CAREFULLY. Make sure it goes to the correct folder.
 - e. Proceed as above to any additional GenBank records as required. You can back click to the root menu, or start over above.
 - f. **Example:** at the [root](#), click on the '1' for *Conepatus*, then click on 'see', then on '[NC_042596](#)'. Proceed as in (d) and download 'NC042596 Conepatus chinga.txt'
2. **Edit** the file 'NC042596 Conepatus chinga.txt' as indicated on the next three sheets.

3. Download and install **MEGA-X** from the internet. Open the base **MEGA** data file, something like **4250 Carnivore.mas** where ***.mas** is the alignment format.
 - a. Choose **Analyze**. Choose **Vertebrate mtDNA** as the genetic code, if not already chosen.
 - b. Move the slider all the way to the left. Notice that positions 1-3 are the **ATG start triplet** [not a codon, which occurs in mRNA only] for the first coding region, **ND1**.
 - c. Scan the slider forward to position **2014**. Notice that position 2011-2013 are Stop triplets, that there is a 12-space gap corresponding to exactly 4 x 3bp blank triplets, and the second coding region **ND2** begins with a **ATG** triplet at position **2026**.
 - i. Scanning forward through the rest of the file will show that all pre-existing sequences have been aligned so as to keep the first reading frame constant. Gaps are not consistently 12 or other multiples of three, for various reasons.
 - d. **Import your new files, edited as in Step 2, as follows**
 - i. Under the second MEGA pull-down menu **Edit**, choose '**Insert sequence from file**'
 - ii. Choose one or all new sequences to be inserted.
 - iii. The new files will show up more or less at random among the existing files.: gather them at the bottom
 - iv. Scan along up to position 2014: it should be apparent that homologous **ND1** sequences among carnivores are typically quite similar, with major differences between caniform and feliform species.
 - v. In most cases, the end of the new sequence will terminate at position 2013: it may be necessary to pad the end with an 'a' or 'ag' to conform to the other sequences [mtDNA Stop Triplets are frequently formed during *polyadenylation* of the mRNA with a string of 'aaaa's, corresponding to a terminal 'g' on the Sense Strand]. Indicates edits as lower-case letters.
 - vi. In most cases, the beginnings of successive **Open Reading Frames (ORFs)** will begin in about the same spot, but may shift. Sequences are subject to insertion – deletion (**indel**) events with respect to other related sequences
 - vii. **In all cases, micro-tune the alignments between your new sequences data and the old. This will typically involve insertion of a space (type a hyphen) or removal of one (place cursor on NEXT letter, backspace). There are numerous editing tricks that can be learned as you go on; the instructor will indicate some of these.**
 1. **Notably: the DNA data can be displaced as one-letter amino acid translations: this often makes adjustment of homologous region easier.**
 - viii. **SAVE** the intermediate results at frequent intervals: MEGA is buggy and crashes a lot.
4. Export the ***.mas** alignment to a ***.meg** file for analysis.
 - a. Analysis can also be done within the *.mas file, but annotation and organization of files is more flexible and annotatable in *.meg. The only drawback is that you have to go back to *.mas to modify the file content.
 - b. Under '**Data**' / '**Export alignment**' / '**MEGA format**'
 - i. If you're luck, this will go smoothly.
 - ii. If you're not, some silly formatting error can hang you up. Consult instructors.
 - c. You will have choice of **NJ**, **MP**, and **ML** options, with various choices of parameters within these modes. Initial trees can be obtained with corroboration from Bootstrap analysis (see lecture). These trees can be arranged, drawn, & exported as ***.tiff** figures or as ***.pdf** files for incorporation in your lab reports.

5. Systematic investigations:

- a. **Relationships of Giant Panda and Lesser Panda to other caniform taxa.** The Giant Panda (*Ailurus melanoleuca*) differs from typical bears in several respects, including a diet that comprises exclusively bamboo, a highly modified cheekteeth and jaw structure consistent with this diet, and a biogeographic distribution limited to continental China. The Giant Panda also lacks specific skull characteristics otherwise characteristic of bears. This, along with its biogeographic and dietary similarities to the Lesser or Red Panda (*Proailurus fulgens*) have historically been taken to suggest that *Ailurus* and *Proailurus* are closer relatives of the Procyonidae than of Ursidae. **Test this hypothesis** by an appropriate selection of ursid, procyonid, and other caniform sequences, along with those of *Ailurus* and *Proailurus*.
- b. **Differentiation of feliform and caniform morphology.** The occurrence of only three feliform families *versus* seven caniform families has been taken to indicate greater morphological divergence in the latter, and to predict (as in the comparison of Great Apes) greater molecular differentiation. **Test this hypothesis** by comparing representative species for each family. Choose the type genus where possible, ie, *Canis* from Canidae, *Felis* from Felidae, and so on.
- c. **Molecular diversity within the Felidae.** In comparison to other carnivore families, it is often stated that all felids look pretty much alike, the only notable exception being the cheetah (*Acionyx jubata*), which has non-retractile claws. Otherwise, cats have often been separated into two genera, smaller (*Felis*) and larger (*Panthera*). **Test this hypothesis** by an appropriate selection of felid species, including *Acionyx* and a variety of larger and smaller species.
- d. **Molecular diversity within viverrid mongooses & relatives.** Unlike carnivore families with Nearctic and Palearctic distributions, relationships among the south Asian and African species of cat-like carnivores known as mongooses and civets (among other local names) were poorly understood until the advent of molecular data. Textbooks treated them as a single family, Viverridae. Recent work suggests otherwise. **Test the hypothesis** of a single monophyletic family; if not, suggest an alternative arrangement of viverrid taxa, and their relationship to the other two feliform families, Felidae and Hyaenidae
- e. **Evolution of pagophilic behavior in Phocidae.** Based on a small region of the mtDNA Cytb gene, [Perry et al. \(1995\)](#) proposed a theory of the origins of pagophilic (“ice-loving”) breeding behavior in earless seals of the North Atlantic. Complete mitogenome sequences for all species in that study. **Test this hypothesis** with the new data from the Phocidae.

1) NC042596 *Conepatus chinga.txt* file looks like:

```
>lcl|NC_042596.1_cds_YP_009646897.1_1 [gene=ND1] [locus_tag=FI654_mgp13] [db_xref=GeneID:40412674]
[protein=NADH dehydrogenase subunit 1] [transl_except=(pos:955..955,aa:TERM)]
[protein_id=YP_009646897.1] [location=2758..3712] [gbkey=CDS]
ATGCTAATTATTAACATTATTTCTCTAATCGTACCTATTCTCCTGGCTGTAGCTTTCTTGACCCTAGTAG
AACGTAAGTACTAGGTTATATACAGCTTCGAAAAGGTCCTAACATTGTAGGACCTTACGGACTACTTCA
ACCCATCGCAGACGCTGTAAAACCTATTCACAAAAGAACCTACGACCCCTCACATCCTCTATATCCATA
TTTATTTTAGCCCCATTCTTGCCCTAACACTAGCCCTAACAAATATGAATTCCACTGCCAATACCATAACC
CACTTACCAACATAAACCTGAGTATTCTCTTCTACTAGCCATATCAAGCCTCGCTGTTTACTCAATTCT
ATGATCAGGTTGATCTTCAAACCTAAAATATGCTCTAATTGGAGCCTTACGAGCTGTAGCCCAACAATC
TCCTACGAAGTAACACTAGCCATTATCCTATTATCAGTATTATTAATAAACGGATCATTCACTCTATCTA
CTTTAATCACTACACAAGAACACTTATGATTAATCTTCCCCGCATGACCACTGGCCATAATATGATTCAT
CTCAACCCTAGCAGAACTAATCGAGCACCATTTGATTTAACAGAAGGAGAATCAGAAGTACTAGTATCAGGA
TTTAAATGTAGAATACGCAGCTGGCCCTTCGCACATTTCTTTTTAGCAGAATATGCTAACATCATTATAA
TAAATATCCTTACTACCATTCTATCTTCCGGCGATATCATAATCCATTAATGCCGGAACCTATATCAAT
TAACTTCACTACTAAAACACTACTCCTAACAAATTTCTTTTCTGTGAATCCGAGCATCATACCCCCGATTC
CGCTATGATCAACTTATACACTTATATGAAAAATTTTTTACCCTTGACACTAGCCCTTTGCATATGAC
ACATAGCCCTACCTATTATCACAGCAAGCATCCCTCCACAAACAT
>lcl|NC_042596.1_cds_YP_009646898.1_2 [gene=ND2] [locus_tag=FI654_mgp12] [db_xref=GeneID:40412677]
[protein=NADH dehydrogenase subunit 2] [transl_except=(pos:1042..1042,aa:TERM)]
[protein_id=YP_009646898.1] [location=3922..4963] [gbkey=CDS]
ATAAAACCCACTATCATAATCGTTATCTCATCAACCGTTATCATAGGAACCATAATAGTAATAATAAGCT
CCCCTGATTATTAATCTGAATGGATTTGAAATAAATACTAACAATTATCCCCATTCTAATAAAAAA
TTTTAACCCACGAGCCATAGAAGCATCAACAAAATATTTTCTAACACAGGCTACAGCATCAATAATCCTA
ATACTAGGCATTGCCACAAATCTACTACTAACAGGACAATGAACAATGCTAAACATGCTAAACCCAATTG
CATCAAACATCACTACCATTGCCCTAGCAATGAAACTCGGACTCTCCCCTTTCCACTTCTGAGTACCCGA
AGTAACCCAAGGAGTCTCTATATTTTCAGGAACAGTACTACTAACCTGACAAAAAATCGCACCATATCA
GTCTTATACCAAATTTCAACATCCATTAACCCCTAGCCCTACTAACCATTTCTAGCCATTATATCAGTACTAG
TAGGAGGATGAGGAGGACTAAACCAAACCTCAGCTCCGAAAAATCCTGGCTTACTCATCCATCTCCCATAT
AGGATGAATAACAGCAGTATTAACATATAATCCAACCTTAACAACCCCTAACCTACTAATTTATATTTATA
ATAACAACAGGAACCTTTTATCTACTTATTCTAAATCTTCCACCACCACACTATCACTATCTCACACAT
GAAACAAGTTACCCTTAAGCTACTCCTAGTACTGATACTCATACTATCCATCGGAGGCTACCACCCCT
ATCAGTTTTTATACCAAAATGAATAATCATTCAAGAACTCACAAAAAACGACCTAATTATGATACCAACA
CTCATAGCAATAGCAGCCCTACTAAATCTATATTTTTACATACGACTTATATACACAACAACACTCACAA
TATTTCCCTCATCAAAACAACATAAAAAATAAATGACAGTTTCGAAAACCTCAAAAATAATACCCCTTCTACC
ACCCTAATTTATCTATCAACCATATCACTACCCTCTCCCAATACTAATAATTTCTACTACT
>lcl|NC_042596.1_cds_YP_009646899.1_3 [gene=COX1] [locus_tag=FI654_mgp11] [db_xref=GeneID:40412682]
[protein=cytochrome c oxidase subunit I] [protein_id=YP_009646899.1] [location=5354..6898]
[gbkey=CDS]
ATGTTTCATCAACCGATGGCTATTTTCTACAAATCACAAAGACATCGGCACCTTTTACCTCCTATTCGGGG
CATGAGCCGGAATAGCTGGAACAGCCCTTAGTCTTCTAATTCGAGCTGAATTAGGACAACCTGGAACCCCT
ATTAGGAGATGACCAGATTTACAATGTAGTAGTCACAGCCCATGCATTCATCATAATCTTCTTCATAGTA
ATACCAATCATAAATTGGAGGCTTTGAAACTGACTAGTCCCATTAAATGATTGGAGCCCCAGATATAGCAT
TTCCACGAATAAATAACATAAGCTTCTGATTATTTACCCCTCCTTCTACTTATAGCTTCCCTCCAT
GGTAGAAGCAGGTGCAGGAACCTGGTGAACAGTATATCCTCCACTAGCAGGTAACCTAGCCCACGCAGGA
GCATCCGTAGATCTAACAAATCTTCTCCTTACATCTAGCAGGTGTATCCTCAATTTTAGGGGCTATTAAC
TCATTACTACAATTATTAATATAAAACCCCTGCAATATCCCAATATCAAACCTCCCCTGTTTCGTATGATC
TGTACTAATTACAGCAGTCTTCTTACTCTTATCCCTACCAGTATTGGCAGCCGGCATTACTATATTACTC
ACAGATCGAAACCTAAATACAACCTTCTTTGACCCAGCTGGAGGAGGAGATCCAATCCTATATCAACACT
TGTTCTGATTCTTTGGACATCCAGAAGTCTATATTTCTAATTTTACCAGGATTTGGTATAATTTACATAT
CGTAACTTATTACTCAGGCAAAAAAGAACCATTTGGATATATGGGTATAGTATGAGCAATAATGTCTATT
GGCTTCTTAGGCTTTATTTGTATGAGCCACCATATATTTACTGTAGGTATAGACGTAGATACACGAGCCT
ACTTTACCTCAGCTACCATAATCATTGCAATTTCAAACCTGGAGTAAAAGTATTTAGTTGACTAGCTACTTT
GCACGGAGGCAATATCAAATGATCCCCTGCCATACTATGAGCACTAGGCTTCATTTTTTTTATTCACAGTT
GGTGGCCTTACAGGTATTGTATTATCCAATTTCTACTAGACATTTGACTTATGATACGTACTATGTAG
TAGCCCACTTCCACTATGTGTTTATCAATGGGAGCAGTATTTGCAATCATAGGAGGATTTGTTTCATTGATT
CCCCCTATTTTTAGGTTATACACTCAATGATGCATGAGCAAAAAATCCACTTCACAATTATATTTGTAGGA
GTAAACATAACATTTTTCCACAACATTTTCTAGGTCTATCAGGAATACCCGACGTTATTCAGATTACC
CAGACGCTTATACAACATGAAACACAGTATCTTCTATAGGCTCTTTTATCTCGCTAACAGCAGTAATACT
AATAATCTTTATAATCTGAGAAGCCTTCGCATCCAAACGAGAAGTTCTAGTAGTAAATTAACCACCACT
AATATTGAATGACTACATGGATGTCTCCCCCATATCACACATTTGAAGAACCCTTATGTAATACTAA
AATAA
```

And so on through thirteen blocks

2) Add twelve (12) hyphen strings '-----' after every sequence block except the last;

DELETE >lcl|... blocks on all lines except the first

>lcl|NC_042596.1_eds_YP_009646897.1_1 [~~gene=ND1~~] [~~locus_tag=FI654_mgp13~~] [~~db_xref=GeneID:40412674~~]
[~~protein=NADH dehydrogenase subunit 1~~] [~~transl_except=(pos:955..955,aa:TERM)~~]
[~~protein_id=YP_009646897.1~~] [~~location=2758..3712~~] [~~gbkey=CDS~~]

ATGCTAATTATTAACATTATTTCTTAATCGTACCTATTCTCCTGGCTGTAGCTTTCTTGACCCTAGTAG
AACGTAAGACTAGTTATATATACAGCTTCGAAAAGGTCCTAACATGTAGGACCTACGGACTACTTCA
ACCCATCGCAGACGCTGTAAAACATATTCACAAAAGAACCTACGACCCCTCACATCCTCTATATCCATA
TTTTTTTTAGCCCCATTCTTGCCCTAACACTAGCCCTAACAAATATGAATTCCACTGCCAATACCATAACC
CACTTACCAACATAAACCTGAGTATTCTTCTACTAGCCATATCAAGCCTCGCTGTTTACTCAATTCT
ATGATCAGGTTGATCTTCAAACCTAAAATATGCTCTAATTGGAGCCTTACGAGCTGTAGCCCAAACAATC
TCCTACGAAGTAACACTAGCCATTATCCTATTATCAGTATTATTAATAAACGGATCATTCACTCTATCTA
CTTTAATCACTACACAAGAACACTTATGATTAATCTTCCCCGCATGACCACTGGCCATAATATGATTCAT
CTCAACCCCTAGCAGAACTAATCGAGCACCATTTGATTTAACAGAGGAGAATCAGAACTAGTATCAGGA
TTTAATGTAGAATACGCAGCTGGCCCTTCGCACATTTCTTTTTAGCAGAATATGCTAACATCATATATAA
TAAATATCCTTACTACCATTCTATTCTTCGGCGCATATCATAATCCATTAATGCCGGAACCTATATTCAT
TAACTTCACTACTAAAACACTACTCCTAACAAATTTCTTCTGTGAATCCGAGCATCATACCCCCGATTC
CGCTATGATCAACTTATACACTTATATGAAAAAATTTTTTACCCTGACACTAGCCCTTTGCATATGAC
ACATAGCCCTACTTATACACAGCAAGCATCCCTCCAAACAT-----

>lcl|NC_042596.1_eds_YP_009646898.1_2 [~~gene=ND2~~] [~~locus_tag=FI654_mgp12~~] [~~db_xref=GeneID:40412677~~]
[~~protein=NADH dehydrogenase subunit 2~~] [~~transl_except=(pos:1042..1042,aa:TERM)~~]
[~~protein_id=YP_009646898.1~~] [~~location=3922..4963~~] [~~gbkey=CDS~~]

ATAAAACCCACTATCATAATCGTTATCTCATCAACCGTTATCATAGGAACCATAATAGTAATAATAAGCT
CCCCTGATTATTAATCTGAATTGGATTTGAAATAAATACTAACAATTATCCCCATTCTAATAAAAAA
TTTTAACCCACGAGCCATAGAAGCATCAACAAAATATTTTCTAACACAGGCTACAGCATCAATAATCCTA
ATACTAGGCATTGCCACAATCTACTACTAACAGGACAATGAACAATGCTAAACATGCTAAACCCAATTG
CATCAAACATCACTACCATTGCCCTAGCAATGAACTCGGACTCTCCCCTTTCCACTTCTGAGTACCCGA
AGTAACCCAAGGAGTCTCTATATTTTCAGGAACAGTACTACTAACCTGACAAAAAATCGCACCATTATCA
GTCTTATACCAAATTTACCATCCATTAACCCCTAGCCTACTAACCATTTAGCCATTATATCAGTACTAG
TAGGAGGATGAGGAGGACTAAACCAAACCTCAGCTCCGAAAAATCCTGGCTTACTCATCCATCTCCCATAT
AGGATGAATAACAGCAGTATTAATCAATAAATCCAACTTAACAACCCCTAACCTACTAATTTATATATA
ATAACAACAGGAACCTTTTTTACTTATTCTAAATTTCTCCACCACCACACTATCACTATCTCACACAT
GAAACAAGTTACCCCTAAGCACACTCTAGTACTGATACTCATACTATCCATCGGAGGCTACCACCCCT
ATCAGGTTTTATACAAAATGAATAATCATTCAAGAATCACAAAAACGACCTAATTATGATACCAACA
CTCATAGCAATAGCAGCCCTACTAAATCTATATTTTACATACGACTTATATACACAACAACACTCACAA
TATTTCCCTCATCAAAACAATAAAAAATAAATGACAGTTCGAAAACCTCAAAAATAATACCCCTTCTACC
ACCCTAATTATTTCTATCAACCATATCACTACCCTCTCCCCAATACTAATAATTTCTACT-----

>lcl|NC_042596.1_eds_YP_009646899.1_3 [~~gene=COX1~~] [~~locus_tag=FI654_mgp11~~] [~~db_xref=GeneID:40412682~~]
[~~protein=cytochrome c oxidase subunit I~~] [~~protein_id=YP_009646899.1~~] [~~location=5354..6998~~]
[~~gbkey=CDS~~]

ATGTTTCATCAACCGATGGCTATTTTCTACAAATCACAAAGACATCGGCACTCTTTACCTCCTATTCGGGG
CATGAGCCGGAATAGCTGGAACAGCCCTTAGTCTTCTAATTCGAGCTGAATTAGGACAACCTGGAACCT
ATTAGGAGATGACCAGATTACAATGTAGTAGTACAGCCCATGCATTCATCAATCTTCTCATAGTA
ATACCAATCATAAATTGGAGGCTTTGGAACTGACTAGTGCATTAATGATGAGGCCCATATAGCAT
TTCCACGAATAAATAACATAAGCTTCTGATTATACCCCATCCCTCCTTCTACTTATAGCTTCTCCAT
GGTAGAAGCAGGTGCAGGAACCTGGGTGAACAGTATATCCTCCACTAGCAGGTAACCTAGCCCACGCAGGA
GCATCCGTAGATCTAACAAATCTTCTCCTTACATCTAGCAGGTGTATCCTCAATTTTAGGGCTATTAAC
TCATTACTACAATTATTAATATAAAACCCCTGCAATATCCCAATATCAAACCTCCCTGTTTCGTATGATC
TGTACTAATTACAGCAGTCTTACTCTTATCCCTACCAGTATTGGCAGCCGGCATTACTATATFACTC
ACAGATCGAAACCTAAATACAACCTTCTTTGACCAGCTGGAGGAGGAGATCCAATCCTATATCAACACT
TGTTCTGATTCTTTGGACATCCAGAAGTCTATATTTCTAATTTTACCAGGATTTGGTATAATTTACATAT
CGTAACTTATTACTCAGGCAAAAAAGAACCATTTGGATATATGGGTATAGTATGAGCAATAATGTCTATT
GGCTTCTTAGGCTTTATTGTATGAGCCACCATATATTTACTGTAGGTATAGACGTAGATACACGAGCCT
ACTTTACCTCAGCTACCATAATCATTGCAATTTCAACTGGAGTAAAAGTATTTAGTTGACTAGCTACTTT
GCACGGAGGCAATATCAAATGATCCCTGCCATACTATGAGCCTAGGCTTCATTTTTTTTATTACAGTT
GGTGGCCTTACAGGTATTGTATTTATCCAAATCTTACTAGACATTGTACTCATGATACGTACTATGTAG
TAGCCCACTTCCACTATGTGTTATCAATGGGAGCAGTATTTGCAATCATAGGAGGATTTGTTTCATGATT
CCCCTTATTTTTCAGGTTATACACTCAATGATGCATGAGCAAAAAATCCACTTCACAAATATATTTGTAGGA
GTAAACATAACATTTTTCCACAACATTTTCTAGGTCTATCAGGAATACCCCGACGTTATTCAGATTACC
CAGACGCTTATACAACATGAAACACAGTATCTTCTATAGGCTCTTTTATCTCGCTAACAGCAGTAATACT
AATAATCTTTATAATCTGAGAAGCCTTCGCATCCAAACGAGAAGTTCTAGTAGTAAATTAACCACCACT
AATATTGAATGACTACATGGATGTCTCCCCATATCACACATTTGAAGAACCCTTATGTAATACTAA
AATAA-----

And so on through thirteen blocks

3) Delete '-' after NC code, copy species name from original file. Text file now look like this; hyphen strings are not in red. Save edited file as 'NC042596 *Conepatus chinga* string.txt' where string indicates continuous file string

```
>lcl|NC042596 Conepatus chinga
ATGCTAATTATTAACATTATTTCTCTAATCGTACCTATTCTCCTGGCTGTAGCTTTCTTGACCCTAGTAG
AACGTAAAGTACTAGGTTATATACAGCTTCGAAAAGGTCCTAACATTGTAGGACCTTACGGACTACTTCA
ACCCATCGCAGACGCTGTAAAACCTATTCACAAAAGAACCCTACGACCCCTCACATCCTCTATATCCATA
TTTTATTTAGCCCCATTCTTGCCCTAACACTAGCCCTAACAAATGAATTCCACTGCCAATACCATAACC
CACTTACCAACATAAACCTGAGTATTCTCTTCACTAGCCATATCAAGCCTCGCTGTTTACTCAATTCT
ATGATCAGGTTGATCTCAAACCTCAAATATGCTCTAATTGGAGCCTTACGAGCTGTAGCCCAAACAATC
TCCTACGAAGTAACACTAGCCATTATCCTATTATCAGTATTATTAATAAACGGATCATTCACTCTATCTA
CTTTAATCACTACACAAGAACACTTATGATTAATCTTCCCCGCATGACCACTGGCCATAATATGATTCAT
CTCAACCTTAGCAGAACTAATCGAGCACCATTTGATTTAACAGAAGGAGAATCAGAACTAGTATCAGGA
TTTAAATGTAGAATACGCAGCTGGCCCTTCGCACATATTCTTTTGTAGCAGAATATGCTAACATCATTATAA
TAAATATCCTTACTACCATTCTATTCTTCGGCGCATATCATAATCCATTAATGCCGGAACCTATATCAAT
TAACCTCACTACTAAAACACTACTCCTAACAAATTTCTTCTGTGAATCCGAGCATCATACCCCGATTC
CGCTATGATCAACTATACACTATATATGAAAAAATTTTTACCCTGACACTAGCCCTTTGCATATGAC
ACATAGCCCTACTTATATCAGCAAGCATCCCTCCAAAACAT-----
ATAAAACCCACTATCATAATCGTTATCTCATCAACCGTTATCATAGGAACCATAATAGTAATAATAAGCT
CCCCTGATTATTAATCTGAATTGGATTTGAAATAAATACTAACAATTATCCCCATTCTAATAAAAAA
TTTTAACCCACGAGCCATAGAAGCATCAACAAAATATTTCTAACACAGGCTACAGCATCAATAATCCTA
ATACTAGGCATTGCCACAATCTACTACTAACAGGACAATGAACAATGCTAAACATGCTAAACCCAAATTG
CATCAAACATCACTACCATTGCCCTAGCAATGAACTCGGACTCTCCCTTTCCACTTCTGAGTACCCGA
AGTAAACCAAGGAGTCTCTATATTTTCAGGAACAGTACTACTAACCTGACAAAAAATCGCACCATTATCA
GTCCTATACCAAATTCACCATCCATTAACCCCTAGCCTACTAACCTTCTAGCCATTATATCAGTACTAG
TAGGAGGATGAGGAGGACTAAACCAAACCTCAGCTCCGAAAAATCCTGGCTTACTCATCCATCTCCCATAT
AGGATGAATAACAGCAGTATTAACATATAATCCAACCTTAACAACCCCTAACCTACTAATTTATATATA
ATAACAACAGGAACCTTTTTATTACTTATTCTAAATTTCTCCACCACCACACTATCACTATCTCACACAT
GAAACAAGTTACCCCTAAGCACACTCCTAGTACTGATACTCATCCATCGGAGGCCTACCACCCCT
ATCAGTTTTTATACCAAATGAATAACTATTCAAGAACTCACAAAAAACGACCTAATTTATGATACCAACA
CTCATAGCAATAGCAGCCCTACTAAATCTATATTTTACATACGACTTATATACACAACAACACTCACAA
TATTTCCCTCATCAAAACAACATAAAAAATAAATGACAGTTCGAAAACCTCAAAAAATAATACCCCTTCTACC
ACCCTAATTTATCTATCAACCATATCACTACCCTCTCCCAATACTAATAATTTCTACT-----
ATGTTTCATCAACCGATGGCTATTTTCTACAAATCACAAAGACATCGGCCTCTTTACCTCCTATTCGGGG
CATGAGCCGGAATAGCTGGAACAGCCCTTAGTCTTCTAATTCGAGCTGAATTAGGACAACCTGGAACCCCT
ATTAGGAGATGACCAGATTTACAATGTAGTAGTACAGCCCATGCATTCATCATAATCTTCTTCTATAGTA
ATACCAATCATAATGGAGGCTTTGAAAACCTGACTAGTGCCATTAATGATTGGAGCCCCAGATATAGCAT
TTCCACGAATAAATAACATAAGCTTCTGATTATACCCCATCCCTCCTTCTACTTATAGCTTCTCCAT
GGTAGAAGCAGGTGCAGGAACCTGGGTGAACAGTATATCCTCCACTAGCAGGTAACCTAGCCCACGCAGGA
GCATCCGTAGATCTAACAACTTCTCCTTACATCTAGCAGGTGTATCCTCAATTTTAGGGCTATTAACCT
TCATTACTACAATTATTAATAAAAAACCCCTGCAATATCCCAATATCAAACCTCCCTGTTTCGTATGATC
TGTACTAATTACAGCAGTCTCTTACTTATCCCTACCAGTATTGGCAGCCGGCATTACTATATTTACTC
ACAGATCGAAAACCTAAATACAACCTTCTTTGACCCAGCTGGAGGAGGATCCAATCCTATATCAACACT
TGTCTGATTCTTTGGACATCCAGAAGTCTATATTTCTAATTTTACCAGGATTTGGTATAATTTACATAT
CGTAACTTATTACTCAGGCAAAAAAGAACCATTTGGATATATGGGTATAGTATGAGCAATAATGTCTATT
GGCTTCTTAGGCTTTATTGTATGAGCCACCATATATTTACTGTAGGTATAGACGTAGATACACGAGCCT
ACTTTACCTCAGCTACCATAATCATGCAATTCCAACTGGAGTAAAAGTATTTAGTTGACTAGCTACTTT
GCACGGAGGCAATATCAAATGATCCCTGCCATACTATGAGCACTAGGCTTCATTTTTTTTATTACAGTT
GGTGGCCTTACAGGTATTGTATTATCCAATCTTCACTAGACATTGACTTTCATGATACGTACTATGTAG
TAGCCCACTTCCACTATGTGTTATCAATGGGAGCAGTATTTGCAATCATAGGAGGATTTGTTTCATTGATT
CCCCTTATTTTCAGGTTATACACTCAATGATGCATGAGCAAAAAATCCACTTCACAATTATATTTGTAGGA
GTAACATAACATTTTCCACAACATTTCTAGGTCTATCAGGAATACCCCGAGTATTTCAGATTACC
CAGACGCTTATACAACATGAAACACAGTATCTTCTATAGGCTCTTTATCTCGTAAACAGTAATACT
AATAATCTTTATAATCTGAGAAGCTTCGCATCCAAACGAGAAGTCTAGTAGTAAATTATACCACCCT
AATATTGAATGACTACATGGATGCTCCTCCCCCATATCACACATTTGAAGAACCCTTATGTAATACTAA
AATA-----
```

And so on for all thirteen blocks, as a continuous file.