**Biol4250 – Lab Exercise #6: Molecular systematics of the Order Carnivora 2024 Nov 25 smc**

Living members of the order Carnivora are characterized by a particular arrangement of the cheek tooth dentition called the Carnassial Pair, abbreviated as **P4/M1**, for the **fourth upper premolar** and first lower molar. When the jaw closes, the **P4** passes outside the **M1**, so as to create a shearing surface that assists in the processing of the typical carnivore diet, flesh. The **P4/M1**form has been variously modified in various lineages, including reduced dentition in the insectivorous aardwolf (*Proteles cristata*) or crab eater seals (*Lobodon carcinophaga*), or as a massive crushing surface in hyaenas (*Crocuta*). Other taxa with less dependence on flesh have more or less flattened molariform dentition.

Living Carnivora occur in two 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), and **Odobenidae** (walrus). The latter classically comprises three forms, **Felidae** (cats), **Hyaenidae** (hyaenas), and **Viverridae** (mongooses & civets). The last was the least understood, and has undergone extensive revision in the light of 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. [Kelt & Patton (2020) 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 set of 13 **mitochondrial DNA** (**mtDNA**) protein **coding regions** (ca. 12 kbp) from multiple species of carnivores. The methods are those available through the program **MEGA** (Molecular Evolutionary Genetics Analysis, 11th ed.), **Neighbor Joining** (NJ), **Maximum Parsimony** (MP), and **Maximum Likelihood** (ML).

The data are arranged in a MEGA file of 13 protein coding regions from multiple carnivore species, aligned and spaced so as to preserve the first **5’-3’ reading frame** of the first locus (**ND1**) on the **Heavy Strand** consistent through the end of the last (**Cytb**), including *reverse complements* [Why?] of four loci 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 retrieved from the **GenBank** library of taxonomic DNA sequence data. The 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’.
	* 1. 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].
		2. 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.
		3. 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’
		4. 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.
		5. Proceed as above to any additional GenBank records as required. You can back click to the root menu, or start over above.
		6. 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. Open the MEGA datafile as directed, something like **4250\_Carnivore.mas** where **\*.mas** is the alignment format.
	1. Choose **Analyze**. Choose **Vertebrate mtDN**A as the genetic code, if not already chosen.
	2. Move the slider all the way to the left. Notice that positions 1-3 are the **ATG** start **triplet** [*not* a codon, which occur only in mRNA] for the first coding locus, **ND1**.
	3. Scan the slider forward to position **2014**. Notice that position 2011-2013 are a Stop triplet, that there is a 12-space gap corresponding to exactly 4 x 3bp blank triplets, and the second coding region ND2, which 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 all loci constant in the first reading frame. Gaps are not consistently 12 or other multiples of three, for various reasons.

* 1. Import your new files, edited as in Step 2, as follows
		1. Under the second MEGA pull-down menu Edit, choose ‘Insert sequence from file’
		2. Choose one or all the new sequences to be inserted.
		3. The new files will show up more or less at random among the existing files. Gather them at the bottom
		4. 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.
		5. 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 ‘aaaaa’s, corresponding to a terminal ‘g’ on the Sense Strand]. Indicates edits as lower-case letters.
		6. 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 your 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.

1. Export the **\*.mas** alignment to a **\*.meg** file for analysis.
	1. Analysis can also be done within the \*.mas file, but annotation and organization of files is more flexible and annotatable in \*.meg. The drawback is that you have to go back to \*.mas to modify the file content.
	2. Under ‘Data’ / ‘Export alignment’ / ‘MEGA format’
		1. If you’re luck, this will go smoothly.
		2. If you’re not, some silly formatting error can hang you up. Consult instructors.
	3. 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 for incorporation in your lab reports.
2. **Systematic investigations:**
	1. **Relationships of Giant Panda and Lesser Panda** to other caniform taxa. The Giant Panda (*Ailuropoda melanoleuca*) differs from typical bears in several respects, including a diet that comprises exclusively bamboo, highly modified cheek teeth 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 *Ailuropoda* 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*.
	2. ***Genetic 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 measuring differences among representative species for each family. Choose the type genus where possible, ie, *Canis* from Canidae, *Felis* from Felidae, and so on.
	3. **Molecular diversity within the Felidae**. In comparison to other carnivore families, it is often stated that all felids are pretty much alike, the only notable exception being the cheetah (*Acionyx jubata*), which has non-retractile claws. Otherwise, cats have often been divided between two genera, the smaller (*Felis*) and the larger (*Panthera*). Test this hypothesis by an appropriate selection of felid species, including Acionyx and a variety of larger and smaller species.
	4. **Molecular & taxonomic diversity within mongooses, civets, & relatives**. Unlike carnivore families with Nearctic and Palearctic distributions, relationships among the Global South and African species of catlike 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. Test the hypothesis of a single monophyletic family. Suggest an alternative arrangement of viverrid taxa, and their relationship to the other two feliform families, Felidae and Hyaenidae.
	5. **Evolution of pagophilic behavior in Phocidae**. Based on a small region of the mtDNA Cytochrome b gene, Perry et al. (1995) proposed a theory of the origins of pagophilic *(“ice-loving”)* breeding behavior in earless seals of the North Atlantic. Test this hypothesis with the complete mitogenomic sequences for all species in that study.

**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

AACGTAAAGTACTAGGTTATATACAGCTTCGAAAAGGTCCTAACATTGTAGGACCTTACGGACTACTTCA

ACCCATCGCAGACGCTGTAAAACTATTCACAAAAGAACCACTACGACCCCTCACATCCTCTATATCCATA

TTTATTTTAGCCCCCATTCTTGCCCTAACACTAGCCCTAACAATATGAATTCCACTGCCAATACCATACC

CACTTACCAACATAAACCTGAGTATTCTCTTCATACTAGCCATATCAAGCCTCGCTGTTTACTCAATTCT

ATGATCAGGTTGATCTTCAAACTCAAAATATGCTCTAATTGGAGCCTTACGAGCTGTAGCCCAAACAATC

TCCTACGAAGTAACACTAGCCATTATCCTATTATCAGTATTATTAATAAACGGATCATTCACTCTATCTA

CTTTAATCACTACACAAGAACACTTATGATTAATCTTCCCCGCATGACCACTGGCCATAATATGATTCAT

CTCAACCCTAGCAGAAACTAATCGAGCACCATTTGATTTAACAGAAGGAGAATCAGAACTAGTATCAGGA

TTTAATGTAGAATACGCAGCTGGCCCCTTCGCACTATTCTTTTTAGCAGAATATGCTAACATCATTATAA

TAAATATCCTTACTACCATTCTATTCTTCGGCGCATATCATAATCCATTAATGCCGGAACTATATTCAAT

TAACTTCACTACTAAAACACTACTCCTAACAATTTCTTTCCTGTGAATCCGAGCATCATACCCCCGATTC

CGCTATGATCAACTTATACACTTATTATGAAAAAATTTTTTACCCCTGACACTAGCCCTTTGCATATGAC

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

CCCACTGATTATTAATCTGAATTGGATTTGAAATAAATATACTAACAATTATCCCCATTCTAATAAAAAA

TTTTAACCCACGAGCCATAGAAGCATCAACAAAATATTTTCTAACACAGGCTACAGCATCAATAATCCTA

ATACTAGGCATTGCCACAAATCTACTACTAACAGGACAATGAACAATGCTAAACATGCTAAACCCAATTG

CATCAAACATCACTACCATTGCCCTAGCAATGAAACTCGGACTCTCCCCTTTCCACTTCTGAGTACCCGA

AGTAACCCAAGGAGTCTCTATATTTTCAGGAACAGTACTACTAACCTGACAAAAAATCGCACCATTATCA

GTCTTATACCAAATTTCACCATCCATTAACCCTAGCCTACTAACCATTCTAGCCATTATATCAGTACTAG

TAGGAGGATGAGGAGGACTAAACCAAACTCAGCTCCGAAAAATCCTGGCTTACTCATCCATCTCCCATAT

AGGATGAATAACAGCAGTATTAACATATAATCCAACCTTAACAACCCTTAACCTACTAATTTATATTATA

ATAACAACAGGAACCTTTTTATTACTTATTCTAAATTCTTCCACCACCACACTATCACTATCTCACACAT

GAAACAAGTTACCCCTAAGCACACTCCTAGTACTGATACTCATACTATCCATCGGAGGCCTACCACCCCT

ATCAGGTTTTATACCAAAATGAATAATCATTCAAGAACTCACAAAAAACGACCTAATTATGATACCAACA

CTCATAGCAATAGCAGCCCTACTAAATCTATATTTTTACATACGACTTATATACACAACAACACTCACAA

TATTCCCCTCATCAAACAACATAAAAATAAAATGACAGTTCGAAAACTCAAAAATAATACCCCTTCTACC

ACCACTAATTATTCTATCAACCATATCACTACCACTCTCCCCAATACTAATAATTCTACACT

>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]

ATGTTCATCAACCGATGGCTATTTTCTACAAATCACAAAGACATCGGCACTCTTTACCTCCTATTCGGGG

CATGAGCCGGAATAGCTGGAACAGCCCTTAGTCTTCTAATTCGAGCTGAATTAGGACAACCTGGAACCCT

ATTAGGAGATGACCAGATTTACAATGTAGTAGTCACAGCCCATGCATTCATCATAATCTTCTTCATAGTA

ATACCAATCATAATTGGAGGCTTTGGAAACTGACTAGTGCCATTAATGATTGGAGCCCCAGATATAGCAT

TTCCACGAATAAATAACATAAGCTTCTGATTATTACCCCCATCCCTCCTTCTACTTATAGCTTCCTCCAT

GGTAGAAGCAGGTGCAGGAACTGGGTGAACAGTATATCCTCCACTAGCAGGTAACCTAGCCCACGCAGGA

GCATCCGTAGATCTAACAATCTTCTCCTTACATCTAGCAGGTGTATCCTCAATTTTAGGGGCTATTAACT

TCATTACTACAATTATTAATATAAAACCCCCTGCAATATCCCAATATCAAACTCCCCTGTTCGTATGATC

TGTACTAATTACAGCAGTCCTCTTACTCTTATCCCTACCAGTATTGGCAGCCGGCATTACTATATTACTC

ACAGATCGAAACCTAAATACAACCTTCTTTGACCCAGCTGGAGGAGGAGATCCAATCCTATATCAACACT

TGTTCTGATTCTTTGGACATCCAGAAGTCTATATTCTAATTTTACCAGGATTTGGTATAATTTCACATAT

CGTAACTTATTACTCAGGCAAAAAAGAACCATTTGGATATATGGGTATAGTATGAGCAATAATGTCTATT

GGCTTCTTAGGCTTTATTGTATGAGCCCACCATATATTTACTGTAGGTATAGACGTAGATACACGAGCCT

ACTTTACCTCAGCTACCATAATCATTGCAATTCCAACTGGAGTAAAAGTATTTAGTTGACTAGCTACTTT

GCACGGAGGCAATATCAAATGATCCCCTGCCATACTATGAGCACTAGGCTTCATTTTTTTATTCACAGTT

GGTGGCCTTACAGGTATTGTATTATCCAATTCTTCACTAGACATTGTACTTCATGATACGTACTATGTAG

TAGCCCACTTCCACTATGTGTTATCAATGGGAGCAGTATTTGCAATCATAGGAGGATTTGTTCATTGATT

CCCCTTATTTTCAGGTTATACACTCAATGATGCATGAGCAAAAATCCACTTCACAATTATATTTGTAGGA

GTAAACATAACATTTTTCCCACAACATTTTCTAGGTCTATCAGGAATACCCCGACGTTATTCAGATTACC

CAGACGCTTATACAACATGAAACACAGTATCTTCTATAGGCTCTTTTATCTCGCTAACAGCAGTAATACT

AATAATCTTTATAATCTGAGAAGCCTTCGCATCCAAACGAGAAGTTCTAGTAGTAAATTATACCACCACT AATATTGAATGACTACATGGATGTCCTCCCCCATATCACACATTTGAAGAACCCACTTATGTAATACTAA

AATAA

**And so on through thirteen blocks**

**2) Paste a** **twelve** (**12**) **hyphen string** **‘------------**’ over every **lcl** text block ***except*** the *first*;

>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

AACGTAAAGTACTAGGTTATATACAGCTTCGAAAAGGTCCTAACATTGTAGGACCTTACGGACTACTTCA

ACCCATCGCAGACGCTGTAAAACTATTCACAAAAGAACCACTACGACCCCTCACATCCTCTATATCCATA

TTTATTTTAGCCCCCATTCTTGCCCTAACACTAGCCCTAACAATATGAATTCCACTGCCAATACCATACC

CACTTACCAACATAAACCTGAGTATTCTCTTCATACTAGCCATATCAAGCCTCGCTGTTTACTCAATTCT

ATGATCAGGTTGATCTTCAAACTCAAAATATGCTCTAATTGGAGCCTTACGAGCTGTAGCCCAAACAATC

TCCTACGAAGTAACACTAGCCATTATCCTATTATCAGTATTATTAATAAACGGATCATTCACTCTATCTA

CTTTAATCACTACACAAGAACACTTATGATTAATCTTCCCCGCATGACCACTGGCCATAATATGATTCAT

CTCAACCCTAGCAGAAACTAATCGAGCACCATTTGATTTAACAGAAGGAGAATCAGAACTAGTATCAGGA

TTTAATGTAGAATACGCAGCTGGCCCCTTCGCACTATTCTTTTTAGCAGAATATGCTAACATCATTATAA

TAAATATCCTTACTACCATTCTATTCTTCGGCGCATATCATAATCCATTAATGCCGGAACTATATTCAAT

TAACTTCACTACTAAAACACTACTCCTAACAATTTCTTTCCTGTGAATCCGAGCATCATACCCCCGATTC

CGCTATGATCAACTTATACACTTATTATGAAAAAATTTTTTACCCCTGACACTAGCCCTTTGCATATGAC

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

CCCACTGATTATTAATCTGAATTGGATTTGAAATAAATATACTAACAATTATCCCCATTCTAATAAAAAA

TTTTAACCCACGAGCCATAGAAGCATCAACAAAATATTTTCTAACACAGGCTACAGCATCAATAATCCTA

ATACTAGGCATTGCCACAAATCTACTACTAACAGGACAATGAACAATGCTAAACATGCTAAACCCAATTG

CATCAAACATCACTACCATTGCCCTAGCAATGAAACTCGGACTCTCCCCTTTCCACTTCTGAGTACCCGA

AGTAACCCAAGGAGTCTCTATATTTTCAGGAACAGTACTACTAACCTGACAAAAAATCGCACCATTATCA

GTCTTATACCAAATTTCACCATCCATTAACCCTAGCCTACTAACCATTCTAGCCATTATATCAGTACTAG

TAGGAGGATGAGGAGGACTAAACCAAACTCAGCTCCGAAAAATCCTGGCTTACTCATCCATCTCCCATAT

AGGATGAATAACAGCAGTATTAACATATAATCCAACCTTAACAACCCTTAACCTACTAATTTATATTATA

ATAACAACAGGAACCTTTTTATTACTTATTCTAAATTCTTCCACCACCACACTATCACTATCTCACACAT

GAAACAAGTTACCCCTAAGCACACTCCTAGTACTGATACTCATACTATCCATCGGAGGCCTACCACCCCT

ATCAGGTTTTATACCAAAATGAATAATCATTCAAGAACTCACAAAAAACGACCTAATTATGATACCAACA

CTCATAGCAATAGCAGCCCTACTAAATCTATATTTTTACATACGACTTATATACACAACAACACTCACAA

TATTCCCCTCATCAAACAACATAAAAATAAAATGACAGTTCGAAAACTCAAAAATAATACCCCTTCTACC

ACCACTAATTATTCTATCAACCATATCACTACCACTCTCCCCAATACTAATAATTCTACACT**------------**

~~>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]~~

ATGTTCATCAACCGATGGCTATTTTCTACAAATCACAAAGACATCGGCACTCTTTACCTCCTATTCGGGG

CATGAGCCGGAATAGCTGGAACAGCCCTTAGTCTTCTAATTCGAGCTGAATTAGGACAACCTGGAACCCT

ATTAGGAGATGACCAGATTTACAATGTAGTAGTCACAGCCCATGCATTCATCATAATCTTCTTCATAGTA

ATACCAATCATAATTGGAGGCTTTGGAAACTGACTAGTGCCATTAATGATTGGAGCCCCAGATATAGCAT

TTCCACGAATAAATAACATAAGCTTCTGATTATTACCCCCATCCCTCCTTCTACTTATAGCTTCCTCCAT

GGTAGAAGCAGGTGCAGGAACTGGGTGAACAGTATATCCTCCACTAGCAGGTAACCTAGCCCACGCAGGA

GCATCCGTAGATCTAACAATCTTCTCCTTACATCTAGCAGGTGTATCCTCAATTTTAGGGGCTATTAACT

TCATTACTACAATTATTAATATAAAACCCCCTGCAATATCCCAATATCAAACTCCCCTGTTCGTATGATC

TGTACTAATTACAGCAGTCCTCTTACTCTTATCCCTACCAGTATTGGCAGCCGGCATTACTATATTACTC

ACAGATCGAAACCTAAATACAACCTTCTTTGACCCAGCTGGAGGAGGAGATCCAATCCTATATCAACACT

TGTTCTGATTCTTTGGACATCCAGAAGTCTATATTCTAATTTTACCAGGATTTGGTATAATTTCACATAT

CGTAACTTATTACTCAGGCAAAAAAGAACCATTTGGATATATGGGTATAGTATGAGCAATAATGTCTATT

GGCTTCTTAGGCTTTATTGTATGAGCCCACCATATATTTACTGTAGGTATAGACGTAGATACACGAGCCT

ACTTTACCTCAGCTACCATAATCATTGCAATTCCAACTGGAGTAAAAGTATTTAGTTGACTAGCTACTTT

GCACGGAGGCAATATCAAATGATCCCCTGCCATACTATGAGCACTAGGCTTCATTTTTTTATTCACAGTT

GGTGGCCTTACAGGTATTGTATTATCCAATTCTTCACTAGACATTGTACTTCATGATACGTACTATGTAG

TAGCCCACTTCCACTATGTGTTATCAATGGGAGCAGTATTTGCAATCATAGGAGGATTTGTTCATTGATT

CCCCTTATTTTCAGGTTATACACTCAATGATGCATGAGCAAAAATCCACTTCACAATTATATTTGTAGGA

GTAAACATAACATTTTTCCCACAACATTTTCTAGGTCTATCAGGAATACCCCGACGTTATTCAGATTACC

CAGACGCTTATACAACATGAAACACAGTATCTTCTATAGGCTCTTTTATCTCGCTAACAGCAGTAATACT

AATAATCTTTATAATCTGAGAAGCCTTCGCATCCAAACGAGAAGTTCTAGTAGTAAATTATACCACCACT AATATTGAATGACTACATGGATGTCCTCCCCCATATCACACATTTGAAGAACCCACTTATGTAATACTAA

# AATAA------------

**And so on through thirteen blocks**

**3) Copy species name from original file. Text file now look like this**. Save edited filed as ‘**NC042596 Conepatus chinga.txt**’

>lcl|NC\_042596 Conepatus chinga

ATGCTAATTATTAACATTATTTCTCTAATCGTACCTATTCTCCTGGCTGTAGCTTTCTTGACCCTAGTAG

AACGTAAAGTACTAGGTTATATACAGCTTCGAAAAGGTCCTAACATTGTAGGACCTTACGGACTACTTCA

ACCCATCGCAGACGCTGTAAAACTATTCACAAAAGAACCACTACGACCCCTCACATCCTCTATATCCATA

TTTATTTTAGCCCCCATTCTTGCCCTAACACTAGCCCTAACAATATGAATTCCACTGCCAATACCATACC

CACTTACCAACATAAACCTGAGTATTCTCTTCATACTAGCCATATCAAGCCTCGCTGTTTACTCAATTCT

ATGATCAGGTTGATCTTCAAACTCAAAATATGCTCTAATTGGAGCCTTACGAGCTGTAGCCCAAACAATC

TCCTACGAAGTAACACTAGCCATTATCCTATTATCAGTATTATTAATAAACGGATCATTCACTCTATCTA

CTTTAATCACTACACAAGAACACTTATGATTAATCTTCCCCGCATGACCACTGGCCATAATATGATTCAT

CTCAACCCTAGCAGAAACTAATCGAGCACCATTTGATTTAACAGAAGGAGAATCAGAACTAGTATCAGGA

TTTAATGTAGAATACGCAGCTGGCCCCTTCGCACTATTCTTTTTAGCAGAATATGCTAACATCATTATAA

TAAATATCCTTACTACCATTCTATTCTTCGGCGCATATCATAATCCATTAATGCCGGAACTATATTCAAT

TAACTTCACTACTAAAACACTACTCCTAACAATTTCTTTCCTGTGAATCCGAGCATCATACCCCCGATTC

CGCTATGATCAACTTATACACTTATTATGAAAAAATTTTTTACCCCTGACACTAGCCCTTTGCATATGAC

ACATAGCCCTACCTATTATCACAGCAAGCATCCCTCCACAAACAT

**------------**

ATAAAACCCACTATCATAATCGTTATCTCATCAACCGTTATCATAGGAACCATAATAGTAATAATAAGCT

CCCACTGATTATTAATCTGAATTGGATTTGAAATAAATATACTAACAATTATCCCCATTCTAATAAAAAA

TTTTAACCCACGAGCCATAGAAGCATCAACAAAATATTTTCTAACACAGGCTACAGCATCAATAATCCTA

ATACTAGGCATTGCCACAAATCTACTACTAACAGGACAATGAACAATGCTAAACATGCTAAACCCAATTG

CATCAAACATCACTACCATTGCCCTAGCAATGAAACTCGGACTCTCCCCTTTCCACTTCTGAGTACCCGA

AGTAACCCAAGGAGTCTCTATATTTTCAGGAACAGTACTACTAACCTGACAAAAAATCGCACCATTATCA

GTCTTATACCAAATTTCACCATCCATTAACCCTAGCCTACTAACCATTCTAGCCATTATATCAGTACTAG

TAGGAGGATGAGGAGGACTAAACCAAACTCAGCTCCGAAAAATCCTGGCTTACTCATCCATCTCCCATAT

AGGATGAATAACAGCAGTATTAACATATAATCCAACCTTAACAACCCTTAACCTACTAATTTATATTATA

ATAACAACAGGAACCTTTTTATTACTTATTCTAAATTCTTCCACCACCACACTATCACTATCTCACACAT

GAAACAAGTTACCCCTAAGCACACTCCTAGTACTGATACTCATACTATCCATCGGAGGCCTACCACCCCT

ATCAGGTTTTATACCAAAATGAATAATCATTCAAGAACTCACAAAAAACGACCTAATTATGATACCAACA

CTCATAGCAATAGCAGCCCTACTAAATCTATATTTTTACATACGACTTATATACACAACAACACTCACAA

TATTCCCCTCATCAAACAACATAAAAATAAAATGACAGTTCGAAAACTCAAAAATAATACCCCTTCTACC

ACCACTAATTATTCTATCAACCATATCACTACCACTCTCCCCAATACTAATAATTCTACACT

**------------**

ATGTTCATCAACCGATGGCTATTTTCTACAAATCACAAAGACATCGGCACTCTTTACCTCCTATTCGGGG

CATGAGCCGGAATAGCTGGAACAGCCCTTAGTCTTCTAATTCGAGCTGAATTAGGACAACCTGGAACCCT

ATTAGGAGATGACCAGATTTACAATGTAGTAGTCACAGCCCATGCATTCATCATAATCTTCTTCATAGTA

ATACCAATCATAATTGGAGGCTTTGGAAACTGACTAGTGCCATTAATGATTGGAGCCCCAGATATAGCAT

TTCCACGAATAAATAACATAAGCTTCTGATTATTACCCCCATCCCTCCTTCTACTTATAGCTTCCTCCAT

GGTAGAAGCAGGTGCAGGAACTGGGTGAACAGTATATCCTCCACTAGCAGGTAACCTAGCCCACGCAGGA

GCATCCGTAGATCTAACAATCTTCTCCTTACATCTAGCAGGTGTATCCTCAATTTTAGGGGCTATTAACT

TCATTACTACAATTATTAATATAAAACCCCCTGCAATATCCCAATATCAAACTCCCCTGTTCGTATGATC

TGTACTAATTACAGCAGTCCTCTTACTCTTATCCCTACCAGTATTGGCAGCCGGCATTACTATATTACTC

ACAGATCGAAACCTAAATACAACCTTCTTTGACCCAGCTGGAGGAGGAGATCCAATCCTATATCAACACT

TGTTCTGATTCTTTGGACATCCAGAAGTCTATATTCTAATTTTACCAGGATTTGGTATAATTTCACATAT

CGTAACTTATTACTCAGGCAAAAAAGAACCATTTGGATATATGGGTATAGTATGAGCAATAATGTCTATT

GGCTTCTTAGGCTTTATTGTATGAGCCCACCATATATTTACTGTAGGTATAGACGTAGATACACGAGCCT

ACTTTACCTCAGCTACCATAATCATTGCAATTCCAACTGGAGTAAAAGTATTTAGTTGACTAGCTACTTT

GCACGGAGGCAATATCAAATGATCCCCTGCCATACTATGAGCACTAGGCTTCATTTTTTTATTCACAGTT

GGTGGCCTTACAGGTATTGTATTATCCAATTCTTCACTAGACATTGTACTTCATGATACGTACTATGTAG

TAGCCCACTTCCACTATGTGTTATCAATGGGAGCAGTATTTGCAATCATAGGAGGATTTGTTCATTGATT

CCCCTTATTTTCAGGTTATACACTCAATGATGCATGAGCAAAAATCCACTTCACAATTATATTTGTAGGA

GTAAACATAACATTTTTCCCACAACATTTTCTAGGTCTATCAGGAATACCCCGACGTTATTCAGATTACC

CAGACGCTTATACAACATGAAACACAGTATCTTCTATAGGCTCTTTTATCTCGCTAACAGCAGTAATACT

AATAATCTTTATAATCTGAGAAGCCTTCGCATCCAAACGAGAAGTTCTAGTAGTAAATTATACCACCACT AATATTGAATGACTACATGGATGTCCTCCCCCATATCACACATTTGAAGAACCCACTTATGTAATACTAA

AATAA

**-----------**

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

Save as a **\*.txt** file with the **accession number and taxon name**