Biology 4505 - Biogeography & Systematics
Dr. Carr

Molecular Evolution and DNA systematics

Ultimately, the source of all organismal variation that we have examined in this course is the genome, written in the four letters of the DNA genetic code. In recent years, DNA has become a powerful systematic tool to investigate evolutionary relationships within and among species. In this lab, we will investigate three basic aspects of molecular systematics, (1) **data collection**, in this case the analysis of DNA sequence data from an automated DNA sequencer, (2) **analysis of molecular evolution**: patterns of mutational change between more and more distantly related creatures, and (3) **DNA systematics**, the use of DNA sequence variation to ascertain phylogenetic relationships among species.

1. Collecting the DNA data

Chromatographic data from the Applied Biosystems automated DNA sequencer can be analyzed in PC-DOS format with the program **CHROMAS (v1.4)** by Conor McCarthy (Griffith Univ., Australia). [available at http://trishul.sci.gu.edu.au/~conor/chromas.html]

In the POPULATION BIOLOGY group, click on the **CHROMAS 1.4** icon.
Open any file with the extension *.ABI: **OPEN / FILE / wtdeer.ABI**
The available files includes whales, fish, sharks, deer, and frogs.

Under **OPTIONS**, choose **CONTINUOUS EDIT** and **EDITED SEQUENCE**, and adjust the X-scale for clarity: **X_ZOOM / 4** (these may already be selected for you). You will see a chromatogram with a series of coloured peaks marked **N**. To read the sequence, change **N** to the correct base call, using lower-case letters: **green** = **a**, **blue** = **c**, **black** = **g**, and **red** = **t**. The correct spacing has been done for you: you should not have to insert or delete bases. Note that sometimes two colored peaks will have similar intensity: leave these as **N**, or use the ambiguity code: **c** or **t** = **Y** (p**Y**rimidine), and **a** or **g** = **R** (pu**R**ine). Try to read at least 300 nucleotides.

Within the first 20-30 nucleotides, you should see an **ATG** sequence (corresponding to an **AUG** start codon in the mRNA): mark this with capital letters (this will make it easier to align).

SAVE your file to a new ***.ABI** file: **FILE / SAVE AS / mybeast.ABI** {for example}
DO NOT save your base calls to the original *.ABI files

EXPORT your sequence to a ***.SEQ** file: **FILE / EXPORT / PLAIN TEXT / mybeast.SEQ** {for example}

2. Analyzing your sequence data

DNA sequences can be manipulated with the **Eyeball Sequence Editor** (**ESEE 3.0S**) by Eric Cabot.
In the POPULATION BIOLOGY program group, click on the **ESEE 3.0** icon.
Restore a previously-saved set of 14 reference species: **FILE / RESTORE SAVE FILE / 3900C.SAV**

Note the blinking cursor: this indicates the sequence and nucleotide being edited.
Choose sequences by moving the cursor with the up/down/left/right arrows.

Insert your sequence: **FILE / INPUT ASCII / C:\CHROMAS\mybeast.SEQ**
Move your sequence above the most closely related sequence: **LAYOUT / MOVE SEQ / number**

ALIGN your sequence with a closely related reference sequence.
In your sequence, delete all bases before the **ATG** start codon;
insert a '---' triplet to align with non-*Xenopus* sequences
Any uncalled nucleotides will show up as a string of **NNNNN's** at the end of your sequence,
move to the first **N** and delete the rest of the string with the **ctrl-Z** key
Place the cursor on the reference sequence: hit **F6** , or **COMMAND / MATCH**
Check the alignment: vertical lines ("|") indicate matches, gaps (" ") indicate mismatches
The **% Similarity** will be indicated in a box at the top of the screen

EDIT your sequence.
If your sequence suddenly goes out of alignment,
you may have added or deleted nucleotides during editing.
Errors occur towards the 3' end: a single broad peak is read as two bases,
or two peaks run together are read as a single base.
Edit by hand; if time permits, go back to **CHROMAS** and correct.

TRANSLATE your edited sequence into amino acids.
Change the genetic code to the **mammalian mitochondrial code**:
FORMAT / GENETIC CODE / MAMMALIAN MITO
Move the cursor to the **first letter** of your sequence:
translate: hit **F4** or **COMMAND / TRANSLATE**
The amino acid sequence is given in the IUPAC single-letter code.
Re-name this line: **FORMAT/ NAME / MybeastAA**
Translate the reference sequence & re-name it;
MOVE your translation next to it in the same way you did the nucleotide sequence
COMPARE the **% Similarity** of the two amino acid sequences
in the same way you did the nucleotide sequences.

3. Patterns of Molecular Evolution

A. Nucleotide and amino acid similarity

Delete all but the following sequences with **Shift-F2**
Xenopus, *Gadus*, *Alces*, *Rangifer*, *Lynx*, and *Canis*

Compare nucleotide and amino acid differences for the following pairs of sequences:

moose (*Alces*) vs. caribou (*Rangifer*)	(different genera)
lynx (*Lynx*) vs. coyote (*Canis*)	(different suborders)
frog (*Xenopus*) vs. codfish (*Gadus*)	(different classes)

Move each pair of sequences adjacent to each other and MATCH:
A box will indicate the **% (sequence) similarity**: RECORD this number in the tables below.

Translate each sequence as indicated above,
re-name the amino acid translation, move adjacent, and MATCH:
Again note the **% (amino acid) similarity**: RECORD this number in the tables below.

B. Patterns of mutational substitution

For each of the three pairs of species listed above:
In each sequence, move the cursor back to the first **A** of the start codon:
space the DNA and amino acid sequences in codon triplets
hit **F1** ; re-do the **MATCH** command.
spacing can be removed by hitting **F2**

For each comparison, **RECORD** the **Number** of **1st, 2nd, and 3rd codon position changes** in the tables on the data sheet..

4. Evolutionary analysis: Evolutionary relationships of the Giant Panda

A long-standing controversy in mammalogy concerns the evolutionary relationships and systematic position of the Giant Panda (*Ailuropoda*). General morphology seems to link it to bears (*Ursus*) (Ursidae); some details of the skull and its peculiar diet (bamboo) and biogeographic distribution (China) seem to link it to lesser pandas (*Ailurus*) and raccoons (*Procyon*) (both Procyonidae). DNA sequence data can be used to resolve this question.

A. Phenetic (UPGMA distance) analysis

Restore the original **3900C.SAV** file;
Remove all sequences except *Canis*, *Ailuropoda*, *Ursus*, *Procyon*, and marten(*Martes*)
Translate all five sequences; re-label the translations.

Determine the nucleotide and amino acid similarities for all 10 pairwise comparisons
[Complete the Tables 1 & 2 on the next page:
put similarities above the diagonal]

Convert the similarities to differences (**%D** = 100% - **%S**)
[Put difference values below the diagonal]
CALCULATE a UPGMA dendrogram from the data in Table 1.

B. Cladistic (Maximum Parsimony) analysis [See **Notes on Parsimony Analysis** on the webpage].

Remove the *Canis* sequence (leaving *Ailuropoda*, *Ursus*, *Procyon*, and *Martes*).
Move *Ailuropoda* to the top line.
Make a 'dot comparison' matrix: COMMAND / GLOBAL COMPARE / Yes
the latter three sequence are identical to *Ailuropoda*, except where indicated.

Inspect each nucleotide position to see if it is **phylogenetically informative**:
An informative character is one for which two species are homologous
(they share a character state, in this case a nucleotide),
and the other two species share a different character state
(they share a different nucleotide).
Such a character suggests that the two species have a common ancestor,
from whom they inherited the particular character state.

In your data set, COUNT the number of positions where:

```
panda = bear    & raccoon = marten     : positions of type 5
panda = raccoon & bear = marten       : "          "   "    6
panda = marten  & bear = raccoon      : "          "   "    7
```

```
     Ai   aat tcg ctt cta gga atc tgc cta atc ctg
     Ur   ... ..a ..g ..a .t. ... ... t.. ... ..a
     Pr   ... ..a ..c ..c ... ..t ... ... ... t.a
     Ma   ... ..a ..a ..g ..g ..t ... t.t ..t t..
          1     2   3   4       5     6         7
```

COMPLETE Table 3: the cladistic "maximum parsimony" answer is the one with the highest score.

Biology 4505 - Biogeography & Systematics NAME _____
Dr. Carr _____

1. Is the Giant Panda *more similar* to the Bear or to the Raccoon? Explain.

2. Is the Giant Panda *more closely related* to the Bear or to the Raccoon? Explain.

3. FOR THE FINAL EXAM: From a four-taxon DNA data set, you should be able to

1. From a matrix of differences, calculate a **UPGMA** tree.
2. Recognize all *phylogenetically informative* sites.
3. Identify: the most *similar* pair of species (by the phenetic [distance] criterion), and
   the most *closely related* pair of species (by the cladistic [parsimony] criterion).

Biology 4505 - Biogeography & Systematics
Dr. Carr

Molecular Systematics Data

	%nuc	%aa	#1st	#2nd	#3rd
Rangifer* vs. *Alces					
Lynx* vs. *Canis					
Gadus* vs. *Xenopus					

Tables 1 & 2. Nucleotide & amino acid comparison among caniform carnivores

Record % similarities above diagonal, % differences below diagonal

Table 1: Nucleotides

	Ca	Ai	Ur	Pr	Ma
Canis	-				
Ailuropoda		-			
Ursus			-		
Procyon				-	
Martes					-

Table 2. Amino acids

	Ca	Ai	Ur	Pr	Ma
Canis	-				
Ailuropoda		-			
Ursus			-		
Procyon				-	
Martes					-

Table 3. Parsimony sites favouring three hypotheses of *Ailuropoda* relationships:

	Ai Pr / **Ur** Ma	**Ai** Ur / **Pr** Ma	**Ai** Ur / **Ma** Pr
Ailuropoda	**X**	**X**	**X**
Ursus	**X**	Y	Y
Procyon	Y	**X**	Y
Martes	Y	Y	**X**