BIOLOGY 2250 LABORATORY 1
2011

Genetic Resources on the Web and Analysis of Interspecific Variation in DNA Sequences

Determination of the sequences of **macromolecules** (proteins and nucleic acids) is a powerful analytical tool for understanding the nature and function of genes and their products. The first protein to have its complete amino acid sequence worked out was bovine insulin, for which Fred Sanger received the Nobel Prize in 1958. Using a similar approach, Robert Holley digested **transfer RNA** with ribonucleases that cleaved the molecule into a series of short fragments (polyribonucleotides), whose composition could be determined chemically. By comparing the sequences of a series of overlapping fragments, he was able to reconstruct the complete "gene sequence" of **tRNA**^{ALA} by 1965. Holley shared the Nobel Prize in 1968 for this work. Techniques for sequencing DNA were worked out in the mid 1970s, the most common of which is Sanger"s **dideoxy terminator** method, described below and in greater detail in lecture. Sanger received his second Nobel Prize in 1980 for this invention, and is the only person to receive two science prizes.

Sanger sequencing has been adapted for use in automated **DNA** sequencers such as the Applied Biosystems machine that produced the data used in this laboratory. Briefly, a **DNA** synthesis reaction that mimics the process of **DNA** replication in the cell nucleus is carried out in a test tube. The reaction includes **DNA polymerase**, a single-stranded **DNA template**, a **primer** complementary to one end of the template, and a mixture of the four **deoxynucleotides** (**dA, dC, dG, & dT**) as well as four **dideoxynucleotide terminators** (**ddNs: ddA, ddC, ddG, & ddT**), each of which has a different fluorescent dye attached to it. **ddNs** resemble **dNs**, except that **ddNs** lack the **-OH** on the **3′-C** as well as the **1'-C** of the sugar. Incorporation of a **ddN** terminates further extension of the **DNA** strand because the **DNA polymerase** cannot add further bases into the **3′-C** end. This produces a series of single-stranded **DNA** molecules that differ in length by plus-or-minus one nucleotide. These fragments can be separated by size by **electrophoresis**. Because each **ddN** has a dye attached to it, a scanning laser and photometer can "read" the colour of the last (dideoxy) nucleotide in each fragment, and a computer places the fragments in proper order. This is displayed in a **DNA chromatogram** as a series of coloured peaks that indicate the **DNA** sequence, which can be called automatically and/or edited by eye. Automated sequencers can now analyze **DNA** molecules of up to 1500 bases at the rate of several hundred nucleotides per hour. (http://www.terranovagenomics.com/id43.html).

In this laboratory, we will analyze the **DNA** sequence data from an automated sequencer. This lab will familiarize you with the nature of the **Genetic Code**, show the effects of DNA sequence **variation** ("mutations": see note) on protein sequences, and will introduce a powerful on-line **bioinformatic** resource, **GenBank,** from the **National Center for Biotechnology Information,** (http://www.ncbi.nlm.nih.gov/BLAST), called the **BLAST** tool.

IMPORTANT: DNA differences within and among species are sometimes referred to as "*mutations*". This is incorrect for several reasons. Most naturally-occurring genetic variation observed within species has been inherited from the parental organisms, and has little or no effect on phenotypes. Such **DNA** variants are called **single-nucleotide polymorphisms** (**SNPs**). These variants originally arose by a process of genetic mutation sometime in the past, in which one type of base has been replaced by another. The observation of a nucleotide difference between two individuals does not imply that one of them has just "mutated" from the other, or is somehow "diseased" or "defective" with respect to the other. "Mutation" is best reserved for the process by which new variants are produced: avoid it when discussing existing genetic variation.

PART 1- GENETIC RESOURCES ON THE WEB

Go to:
http://www.mun.ca/biology/scarr/2250_BLAST_A.html

ataaacccattcatctctattattacatttacaacactcatcctaagcacaacaattgta

60 bp DNA sequence from the Fortune Bay "Sea Monster"

DNA testing was used to identify a 7 m "sea monster" that washed ashore on St. Bernard's, Fortune Bay, Newfoundland, in August, 2001 [Carr *et al.* 2002. Biol Bull, 202:1]. The test involves the **polymerase chain reaction**, which generates a large number of **DNA** copies from a single original gene. The sequence of the gene can then be determined on an **automated DNA sequencer**. The identification is made by means of a **BLAST** search (**B**asic **L**ocal **A**lignment and **S**earch **T**ool), which compares the degree of similarity ("match") of an unknown **DNA** sequence to the library of all known **DNA** sequences. **BLAST** is one of several **bioinformatics** tools available from the **National Center for Biotechnology Information**.

To make this identification from the DNA data from the "sea monster":

Copy the **DNA** sequence from the previous page:

Go to the **NCBI** site (http://www.ncbi.nlm.nih.gov/); click on **BLAST**

Under **Basic BLAST**, choose a **nucleotide blast**.

Paste the **DNA** sequence into the "**Enter …**" box.

Under **Database**, choose the "**others (nr etc)**" button.

This should bring up "**nucleotide collection (nr/nt)**".

Hit the **BLAST** button and wait for the results (< 30 sec).

What is the "Sea Monster" pictured on the previous page?

____

Identification of the Fortune Bay sea monster has led to molecular investigation of a number of other historical carcasses. Using the same procedure described above, you will now be given a DNA sequence from another "sea monster" and will investigate its identity.

Record the letter of the unknown animal given to you by the Lab Instructor here:

____

1. What is the genus and species name (properly written) of the Sea Monster **you** identified? (0.25)

2. Click on the Max Score # of the Organism. Record the **Bit score** and the **Expect value**.

Bit score: ____________ Expect: _______________

Based on these numbers, what is the likelihood that your sequence was a reasonable match with the sequence in the database? Explain. (0.25)

Go to the Animal Diversity Web Site at: http://animaldiversity.ummz.umich.edu/site/index.html

3. Based on web information, give the common name, the geographic range and the size (length) of the organism you identified. (0.25)

4. Based on information provided about the "sea monster" on page 2 as well as the information you found in Question 3 about the animal, is the identification from BLAST *reasonable* for a North Atlantic "sea monster"? Why or why not? (0.25)

PART 2–ANALYSIS OF INTERSPECIFIC VARIATION IN DNA SEQUENCES

Vertebrate mitochondrial DNA genetic code

First Letter	Second Letter U	C	A	G	Third Letter
U	F	S	Y	C	U
	F	S	Y	C	C
	L	S	*	W	A
	L	S	*	W	G
C	L	P	H	R	U
	L	P	H	R	C
	L	P	Q	R	A
	L	P	Q	R	G
A	I	T	N	S	U
	I	T	N	S	C
	M	T	K	*	A
	M	T	K	*	G
G	V	A	D	G	U
	V	A	D	G	C
	V	A	E	G	A
	V	A	E	G	G

In this laboratory, we use sequence data from the **mitochondrial DNA (mtDNA)** genomes of various animals. The mitochondrial code differs from the **Universal Genetic Code** in several respects. Note in particular that **"UGA"** is a "stop" codon in the Universal code, but codes for **"W"** (tryptophan) in the **mtDNA** code. The code table can be used to read the amino acid sequence corresponding to an **mtDNA** sequence directly from an ESEE layout:

Identify the **5'→3'** direction on the **DNA** "sense" strand.
Substitute "**T**" for "**U**" in the above table.

IMPORTANT NOTE: Because the sense strand of **DNA** and the **mRNA** are co-linear (they line up one-to-one), a computer program like **ESEE** can predict the expected amino acid sequence directly from the sense strand of the **DNA**. In the cell, the "antisense" template strand of the **DNA** must first be *transcribed* into **mRNA**, and it is **mRNA** that is then *translated* into the protein sequence. The computer does not "translate" the DNA strand. Likewise, although "start" and "stop" **codons** in the **mRNA** are recognizable as particular triplets in the sense strand of the **DNA**, these triplets are **not** "codons". Be sure you understand clearly the difference between the *logical* process by which the computer program works and the *biochemical* process by which the cell works.

1. Collecting the DNA data

Chromatographic data from the Applied Biosystems automated DNA sequencer can be analyzed with the program **CHROMAS** by Conor McCarthy.

Go to: www.mun.ca/biology/valeriep and choose and right click on any one of the **Rt Files** with an **NNN.SCF** extension. The available files include a variety of protein-coding sequences from the mitochondrial DNA genome of a variety of mammal, bird and fish species. The particular species set available will be discussed with you by the instructors.

Select "**Save link as...**" and save your selection to the "**desktop**". After it's completed its downloading, close the website.
On the desktop, click on the **CHROMAS** icon.
Under **File**, choose **OPEN** from the dropdown menu and look in "**desktop"**. The Rt file should be there. Open that file.

Under **OPTIONS**, choose **CONTINUOUS EDIT** 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, all called as **N**. To read the sequence, change **N** to the correct lower-case letter for each coloured peak:

green = **a**, **blue** = **c**, **black** = **g** **red** = **t**.

Do not insert or delete bases: the correct spacing has been done for you. Note that sometimes two peaks will have similar intensity: leave these as **N**, or use the ambiguity codes:
c or **t** = **Y** (p**Y**rimidine) **a** or **g** = **R** (pu**R**ine).
Do not close the CHROMAS program until you complete Exercise 2 below. **However, when you do exit out, DO NOT Save the changes!!**

2. BLAST your data

Review the on-line exercise for BLASTing a Sea Monster.

Copy your **DNA** sequence from the **CHROMAS** screen by:
From the pull-down menu, choose **EDIT/Copy Sequence/Plain Text** (you won't see anything happening. That's okay, just minimize CHROMAS).
Go to the **NCBI BLAST** site (http://www.ncbi.nlm.nih.gov/BLAST). Click on **BLAST.**
Under "**Basic Blast**", choose a **nucleotide blast**.
Paste the **DNA** sequence into the "**Enter Query Sequence**" box; **DELETE all Ns at the end of the sequence.**
Supply a **Job Title** for your sequence.
Under **Database,** choose the **Search Set** called "**Others (nr etc)**".
This should bring up "**nucleotide collection (nr/nt)**"
Hit the **BLAST** button and wait for the results (≤ 30 sec).
Once the results are given from BLAST, check it with the Lab staff.

3. Genetic Comparisons: Patterns of Base and Amino Acid Variation Between Species

Use the tables on Pages 8-10 for recording your raw data.

In this exercise, you will learn the two types of nucleotides and **compare the nucleotide and amino acid differences** of three pairs of Vertebrate species at increasing degrees of evolutionary divergence: different **genera**, different **families** and different **classes**, to determine what types and positions of nucleotide changes cause amino acid changes and how this is seen in increasingly distant taxa.

STRATEGY: The 3 Photos you will use are of the 3 pairs of taxa. They were initially prepared using a program called ESEE which compares taxa. It is able to convert the nucleotide sequences seen (symbolized by the letters a, t, g and c), using the **mammalian mitochondrial code,** into triplets and then they were translated into the corresponding amino acids. The **capital letters** that you see are the **single-letter code** for the amino acids corresponding to the triplet above it.

****NOTE:** When you compare the last pair of animals (the pair belonging to different classes) you will notice that one of them has an extra triplet after the start **atg** codon. All the other species have a space (–) to accommodate for this.

In order to determine how similar (or dissimilar) the 2 taxa are *in* each pair and to compare the differences *between* the different pairs, record the number and their positions (1st, 2nd or 3rd) of transitions and transversions there are in each of the triplets between each pair of animals, **as seen in the example shown below and on the next page**.
Working with each pair of animals separately, start with the pair that are in different Genera, and then do the pair that are in different Families and finally, record the changes in the pair that are in different Classes. Use the worksheet tables on pp. 8-10 to record your results. **DO NOTE WRITE ON THE PICTURES OR THE PROTECTIVE COVERS THEY ARE IN!**

Also take note as to whether a change results in a change in the amino acid between the two animals in each pair. Make sure to also record the **% nucleotide similarity** and the **% amino acid similarity** between each pair of taxa which is shown on each photo.

For each pair of sequences, go through the complete nucleotide sequences and do not stop until you get to the happy faces☺!!

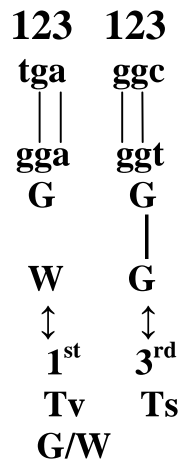
Example of a species pair:

```
SpeciesX      tga tga aat ttt ggc tct cta cta gga atc tgc tta att
              |||  || ||| ||| ||  ||  |||  || |||  |  ||| ||| ||
SpeciesY      tga gga aat ttt ggt tcc cta tta gga gtt tgc tta atc
SpeciesX.aa    W   G   N   F   G   S   L   L   G   I   C   L   I
               |       |   |   |   |   |   |   |       |   |   |
SpeciesY.aa    W   W   N   F   G   S   L   L   G   V   C   L   I
```

Transitions [**Ts**] are interchanges of pyrimidines, **or** of purines; (eg. **c**↔ **t**, or **a** ↔ **g**)

Transversions [**Tv**] are exchanges of pyrimidines for purines, or vice versa; (eg. **c**↔**g**,or **c**↔**a**, or **t**↔**a**, or **t**↔**g**)

The first two changes in the example on the previous page are:



Use the worksheets on the following 3 pages to record your raw data and include in your write up. Analyse and transfer this data to the Results table on Page 12 and include with your scientific write up.

Raw Data- WORKSHEET I. *Different Genera* [record final results in Results section]

Taxa compared:_______________________ **vs**__________________________

% Nucleotide similarity: ____________________

% Amino Acid similarity: ____________________

Position and type of substitution	**1st**	**2nd**	**3rd**
Transitions (**c↔t, a↔g**)			
Transversions (all others)			

1. a) What is the kind, position and nucleotides involved in the majority of the nucleotide substitutions?

b) Did this result in a change of amino acids?

2) How many amino acid changes did you find in the whole sequence?
Was this expected? Why?

/ (0.25)

Raw Data- WORKSHEET II. ***Different Families*** [record final results in Results section]

Taxa compared:____________________ vs_____________________________
% Nucleotide similarity: ____________________
% Amino Acid similarity: ____________________

Position and type of substitution	**1st**	**2nd**	**3rd**
Transitions (c↔t, a↔g)			
Transversions (all others)			

1. How do these patterns differ from those of the two intergeneric species?

2. Were more amino acids changed as compared to the intergeneric species?
   How many amino acid changes were there?
   Was this expected? Why?

/ (0.25)

Raw Data- WORKSHEET III. ***Different Classes*** [record final results in Results section]

Taxa compared:____________________ vs_____________________________
% Nucleotide similarity: ____________________
% Amino Acid similarity: ____________________

Position and type of substitution	**1st**	**2nd**	**3rd**
Transitions (c↔t, a↔g)			
Transversions (all others)			

1. How do these patterns differ from those of the two interfamilial species?

2. Were more amino acids changed as compared to the interfamilial species?
   How many amino acid changes were there*?*
   Was this expected? Why?

/ (0.25)

Name ________________
MUN # ________________
Lab Slot___________

Genetic Comparisons: Patterns of Base and Amino Acid Variation Between Species will be written up in the Scientific Method. Follow the outline as seen in the Appendix on the Scientific Method). You will have until your next lab to pass this report in. **Also, pass in the BLAST questions (staple them to the scientific write up), however, do NOT include them in the Scientific writeup!!**

Background: (1.5)

____
____
____
____
____
____
____
____
____

Purpose: (0.25)

____
____
____
____

Prediction: (0.25)

____
____
____
____
____

Materials and Methods: (1.0)

____
____
____
____
____
____
____
____
____
____

Results: (5.25)

Table 1.TITLE: ____

Taxa pairs:	Different Genera	Different Families	Different Classes
Taxa compared			
% nucleotide Similarity			
% amino acid Similarity			
# transitions			
# transversions			
# 1st position Changes			
#2nd position Changes			
# 3rd position Changes			
# Amino Acid Changes			

Results Summary: (1.0)

____

____

____

____

____

____

____

____

____

____

____

Discussion: (3.5)

(0.5) for organization, grammar, sentence structure, etc.