Lab IV: RNA and Intro to Proteins

Project 1: Finding Protein Coding Regions

In Lab III, we checked our DNA sequence for contamination, computed various composition statistics, and identified its repeats. Now we’re ready to move on to the next step. We can search if and where a protein is encoded within the nucleotide sequence. In an earlier lab, we learned that a DNA coding region (made up of exons in eukaryotic genomes) can be translated into a protein product. However, we also learned that eukaryotic DNA contains many non-coding regions, like introns and repetitive elements such as Alu. Given a DNA sequence, how can we know whether it contains relevant information for the study of proteins? At this point, you’ve been briefly introduced to DNA, RNA, and protein, but you may not completely understand the relationship between these key players in molecular biology. Perhaps a short overview of the “central dogma” will be helpful (Figure 1).

![Diagram of the central dogma of molecular biology](http://www.accessexcellence.org/AB/GG/)

**Figure 1.** The central dogma of biology. In the nucleus, DNA is transcribed into messenger RNA, which is then translated into a protein product at ribosomal sites in the cytoplasm.

• **Replication** - DNA replicates its information in the nucleus of the cell.

• **Transcription** - DNA acts as a template for the synthesis of a complementary RNA copy called nuclear RNA (nRNA).

• **RNA splicing** - Introns are removed from the nRNA copy to form messenger RNA (mRNA).

• mRNA leaves the nucleus and travels to ribosomes in the cytoplasm of the cell.

• **Translation** - mRNA is decoded as proteins are synthesized from amino acids using the genetic code (**Figure 2**).

---

<table>
<thead>
<tr>
<th>T</th>
<th>C</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT Phe (F)</td>
<td>TCT Ser (S)</td>
<td>TAT Ty (Y)</td>
<td>TGT Cys (C)</td>
</tr>
<tr>
<td>TTC &quot;</td>
<td>TCC &quot;</td>
<td>TAC &quot;</td>
<td>TGC</td>
</tr>
<tr>
<td>TTA Leu (L)</td>
<td>TCA &quot;</td>
<td>TAA Ter</td>
<td>TGA Ter</td>
</tr>
<tr>
<td>TTG &quot;</td>
<td>TCG &quot;</td>
<td>TAG Ter</td>
<td>TGG Trp (W)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTT Leu (L)</td>
<td>CAT His (H)</td>
<td>CGT Arg (R)</td>
</tr>
<tr>
<td>CTC &quot;</td>
<td>CCC &quot;</td>
<td>CAC &quot;</td>
</tr>
<tr>
<td>CTA &quot;</td>
<td>CCA &quot;</td>
<td>CAA Gln (Q)</td>
</tr>
<tr>
<td>CTG &quot;</td>
<td>CCG &quot;</td>
<td>CAG &quot;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATT Ile (I)</td>
<td>AAT Asn (N)</td>
</tr>
<tr>
<td>ATC &quot;</td>
<td>ACC &quot;</td>
</tr>
<tr>
<td>ATA &quot;</td>
<td>ACA &quot;</td>
</tr>
<tr>
<td>ATG Met (M)</td>
<td>ACG &quot;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>G</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>GTT Val (V)</td>
<td>GAT Asp (D)</td>
</tr>
<tr>
<td>GTC &quot;</td>
<td>GCC &quot;</td>
</tr>
<tr>
<td>GTA &quot;</td>
<td>GCA &quot;</td>
</tr>
<tr>
<td>GTG &quot;</td>
<td>GCG &quot;</td>
</tr>
</tbody>
</table>

---

**Figure 2. Table of Standard Genetic Code**

Each triplet of nucleotides (codon) codes for one amino acid. Only the amino acids abbreviations are listed in this table. ATG is the Start codon. Ter stands for termination and indicates Stop codons.

http://molbio.info.nih.gov/molbio/gcode.html

---

• **A quick note:** When the instructions say to type something into a box, the actual letters and numbers that you should type will be in regular font (NOT bolded), in contrast to the procedural steps which are bolded.

---

**Procedure 1A: ORFing Your DNA Sequence**

Adapted from pages 160-161 from Bioinformatics for Dummies (2003).

2. Type NM_000616 in the input box for the Accession number.
   Alternatively, you can paste the sequence into the larger input box.

3. Leave the Genetic Code as the default Standard and click the OrfFind button.

4. Click the longest rectangle in the graphical display (or the first in the list to the side).

5. Print a copy of the results.

6. Repeat this procedure using the GenBank entry for the complete genome of HIV-1 (NC_001802).

Procedure 1A Formal Response Questions

a) What is the location (nucleotide position) of the longest ORF for NM_000616? To which reading frame does this correspond?

b) How does the location of the longest ORF for NM_000616 predicted by the ORF Finder compare to the location of the CDS listed under the FEATURES section of the GenBank entry? What does this tell you about your ORF Finder results?

c) On which reading frame are the longest ORFs for NC_001802? How does the location (nucleotide position) of the 8 longest ORFs compare to the locations of the CDS listed under the FEATURES section of the GenBank entry?

d) How many of the six possible reading frames contain a verifiable ORF (meaning that it matches a CDS segment in the GenBank entry) for NM_000616? How many for NC_001802? What does this tell you about the HIV-1 genome?

Procedure 1B: Analyzing Your DNA Sequence with GeneMark

Adapted from pages 162-163 from Bioinformatics for Dummies (2003).


2. Click the here link under Eukaryotic Genomic Sequence Analysis.

3. Copy the NM_000616 sequence from the GenBank entry.
4. Paste the sequence in the input box.

5. Click the Start GeneMark.hmm button to run the search.

6. Repeat this procedure for the HIV-1 genome (NC_001802) using the Heuristic models link under the Prokaryotic Genomic Sequence Analysis header at http://opal.biology.gatech.edu/GeneMark/.

Procedure 1B Formal Response Questions

a) Compare the GeneMark output for NC_001802 with the one from ORF Finder. How many of the genes predicted by GeneMark match exactly with the locations predicted by ORF Finder? How many match either the beginning or end of the ORF but not both at the same time?

b) Which GeneMark predicted gene is not found with the ORF Finder? What is the main difference between this gene and the other six found in GeneMark?

Procedure 1C: Finding Internal Exons in Vertebrate Genomic Sequences

Adapted from pages 163-164 from Bioinformatics for Dummies (2003).

1. Access the exon finding program at EBI at http://corba.ebi.ac.uk/cgi-bin/sp/wrapper.cgi.

2. In FASTA format, paste the X87579 sequence in the input box.

3. Click the Submit sequence button.

Procedure 1C Formal Response Questions

a) What is the location of the exon predicted by the exon finding program?

b) How does this compare with the location of this exon according to GenBank entry X87579? How many exons are specified in the GenBank entry?

c) How accurate is this exon finding program judging from the results?

Procedure 1D: Translating DNA Coding Sequences
After determining the coding regions of a nucleotide sequence, one may want to translate the DNA sequence into the corresponding protein sequence for further analysis. This technique is most useful after you have sequenced a previously unknown nucleotide sequence and want to conduct further analysis on the amino acid sequence to investigate the function of the protein.


2. In the box, paste the CDS segment ONLY of the nucleotide sequence for GenBank entry NM_000616.

3. Click the Translate Sequence button.

   Your browser returns six possible protein sequences translated from the six possible reading frames of the CDS nucleotide sequence.

**Procedure 1D Formal Response Question**

a) Which protein sequence is the correct one? How do you know?

Refer to Chapter 5 in *Bioinformatics for Dummies* (2003) for additional information on finding protein-coding regions.

**Project 2: Working with RNA**

There are three major types of RNA, each of which performs a different function within the cell.

- **Messenger RNA (mRNA)** – serves as a temporary copy of genes that is used as a template for protein synthesis.
- **Transfer RNA (tRNA)** – functions as adaptor molecules that decode the genetic code.
- **Ribosomal RNA (rRNA)** – catalyzes the synthesis of proteins.

RNA is usually thought of as a single stranded linear molecule, but this is not the case in a biological system. Frequently, different regions of the same RNA strand will fold together via base pair interactions to make intricate *secondary* and *tertiary structures* that are essential for correct biological function. Pairing of long stretches of complementary bases forms a *stem*. The unpaired bases between the stem strands make up a *loop* ([Figure 3](figure3.png)). Stems do not have to be perfect; they can also contain unpaired residues called *bulges*. 

![Figure 3](figure3.png)
Procedure 2A: Using Mfold to Predict Secondary Structure

Adapted from pages 367-371 from Bioinformatics for Dummies (2003).


2. On the black menu bar at the top of the form, click Genome.

3. Click the Microbial link, near the top of the Related Resources section (right column).

4. Click the Escherichia coli O157:H7 link.

5. Click the Structural RNAs link under Feature table.

6. Click on the gene link for the third ribosomal RNA on the list, rrFH (5S ribosomal RNA).

   The nucleotide sequence for this RNA product is displayed in FASTA format.

7. Copy the sequence including the definition line.

   Now that we have our rRNA sequence, we’re ready to use Mfold.

8. Access the Applications page of the Center for Bioinformatics (Rensselaer Polytechnic Institute) at http://www.bioinfo.rpi.edu/applications/mfold/.

9. Scroll down the page and click the RNA Folding link.

10. Enter a title for your sequence in the Name field.

11. Paste your rRNA sequence in the Sequence window.

12. Keep all of the default parameters and click the Fold RNA button near the bottom of the form.

13. Scroll down the Output page until you find the section entitled View Individual Structures.

14. Look for the listing of different file formats and then click the jpg that corresponds to Structure 1 link.

   The secondary structure prediction for this rRNA product shows that RNA is not a linear molecule even though it is single-stranded.
15. Try clicking on different base pairs to view the structural details in the Loop Free-Energy Decomposition window.

16. Save your secondary structure prediction by right-clicking it and then choosing the Save As option.

Procedure 2B: Interpreting Mfold Results

Adapted from pages 371-374 from Bioinformatics for Dummies (2003).

- The Dot Plot Folding Comparisons section: This section makes it possible to compare the fold of several suboptimal solutions.

1. Choose jpg from the Image format drop-down menu.

2. Select structures 1 and 2 from the Compare selected foldings line.

3. Click the Do the Comparison button.

A graph similar to the energy dot plot appears. It indicates the elements of predicted secondary structures that are common to the selected suboptimal solutions. Colored elements correspond to stems that occur only in suboptimal secondary structure predictions. Finding a high consistency between the optimal solution (black dots) and the suboptimal ones (colored dots) makes it more likely that the optimal solution is in fact biologically correct.

Procedure 2B Formal Response Questions

a) What is the energy at position (65,73)? What does the energy tell you about the strength of the base pairs at this location? Which nucleotides are involved in this position?

b) What is the energy at position (59,65)? What does the energy tell you about the strength of the base pairs at this location? Which nucleotides are involved in this position?

c) How do Structures 1 and 2 differ concerning the position of U65 (use the View Individual Structures section)? Based only on this base pair, which is more likely according to the energy dot plot?

d) On the comparison dot plot, which pair (65/73 or 59/65) corresponds to the optimal structure? How do you know?
e) Compare the comparison dot plot to the energy dot plot. Which of the colored diagonals on the comparison plot correspond to optimal folds on the energy dot plot? What does this tell you about the optimal structure?

Procedure 2C: Comparing Secondary and Tertiary Structure

You can use Mfold to predict the secondary structure of any RNA strand, including viral genomes. Retroviruses like HIV-1 encode all of their genetic information in the form of RNA with a genome size of about 10,000 nucleotides. The secondary and tertiary folding of RNA is important for processing and packaging of the genetic material during propagation of the virus. The primary function of the nucleocapsid (NC) domain of the Gag polyprotein is to bind specifically to the packaging signal and deliver full-length viral RNAs into the assembling virion.\(^2\) The packaging signal (called \(Ψ\) in HIV and E or L in some other retroviruses) is composed of four RNA hairpins, SL1 through SL4, each made up of a stem and a loop. The tertiary folding and positioning of the nucleotides in the hairpins is crucial to the interaction between the packaging signal and NC during virus assembly. In this lab, we will examine the secondary and tertiary structure of SL3, one of the hairpins with a flexible loop that enables it to interact with NC.\(^3\)

Energy minimization calculations in programs such as Mfold are used to predict the secondary structure of biological molecules, but tertiary structure predictions are slightly more complicated. Two methods, Nuclear Magnetic Resonance (NMR) and X-ray crystallography, can be used to experimentally determine the 3-D structure of molecules like DNA, RNA, and proteins at the atomic level. Unlike X-ray crystallography, which requires that the molecule be crystallized for analysis, NMR spectroscopy is used to determine the 3-D structure of vibrating molecules in solution. Certain atoms, like hydrogen, carbon, nitrogen, fluorine, and phosphorous (the building blocks of nucleotides and proteins) have a nuclear spin that generates a magnetic dipole. This magnetic dipole is analogous to a bar magnet, with both a north and a south pole. When a strong magnetic field is applied to a solution containing a single type of macromolecule, the magnetic dipoles are aligned in the field in one of two orientations, parallel (low energy) or antiparallel (high energy). When the NMR machine emits a pulse of magnetic energy with the proper frequency, the nuclei in solution will absorb some of this energy and “spin-flip” to a new, high-energy state. When this spin-flip occurs, the nuclei are said to be in resonance with the applied magnetic field – hence the name nuclear magnetic resonance. Because different nuclei in solution absorb energy at different frequencies, we observe a whole spectrum of absorbance. This spectrum describes both the identities of the nuclei and their near-molecular neighbors. The data from many such experiments performed on a sample are averaged to generate a two-dimensional NMR spectrum, which is then translated into a three-dimensional structure using additional measurements (for example, bond lengths, bond angles, and atomic radii).\(^4,5\)

Let’s take a look at the secondary and tertiary structures of the SL3 RNA hairpin from HIV-1 using available bioinformatics techniques. **NOTE:** The formal response questions
for this procedure are interspersed throughout the instructions. Make sure you don’t overlook them!


2. Type 1BN0 in the search box and click the Find a structure button.
   
   Two possible matches appear.

3. Click the 1BN0 link to access the Structure Explorer page for the SL3 hairpin in the RNA genome of HIV-1.

4. Click the Sequence Details link in the blue box on the left.
   
   The Tabular Overview tells you that there are 20 residues in this RNA segment.

5. Click the Download all chains in FASTA format link.

6. Copy the sequence without the definition line.

7. Paste the sequence into the Mfold sequence box at [http://www.bioinfo.rpi.edu/applications/mfold/old/rna/](http://www.bioinfo.rpi.edu/applications/mfold/old/rna/).

8. Scroll down and click the Fold RNA button.

   
   The hairpin structure that you observe has an external loop called a tetraloop with the sequence GGAG and a stem that consists of eight base pairs. Mfold’s prediction of the secondary structure of viral RNA looks reasonable, but it doesn’t tell us anything about the tertiary structure. We need to use a 3-D viewer to examine the RNA structure at this level. Protein Explorer is one such viewer that displays 3-D molecular structure through easy-to-use buttons and menus. Although we will only look at the structure of the RNA hairpin loop in this lab, we will also use Protein Explorer to view more complicated 3-D structures, such as proteins, in a future lab.

10. Return to the Structure Explorer page for 1BN0 and click the View Structure link on the left.

11. Click the Protein Explorer link under Interactive 3D Display.

12. In the pop-up window, click the Protein Explorer 2 Beta link.
Protein Explorer will load the 3-D image of the 1BN0 molecule. This may take several seconds. Once the molecule appears, click the OK button in the yellow pop-up window.

13. **Click the Toggle Spinning button to stop the rotation of the molecule.**

14. **Click the Explore More with QuickViews! link.**

   A box pops up alerting you to the fact that this molecule has multiple models. Click OK. We will look at these structures in a moment.

   The QuickViews menus (SELECT, DISPLAY, and COLOR) allow you to manipulate the appearance of your structure in several ways. When you change the menu options, additional information on each option is displayed in the middle box on the left. Spend a few minutes exploring the different options in each menu. You can use your mouse to rotate the structure by holding down the left-click button and moving the mouse.

15. **Click on the ReSet View link in the top box to restore the original structure.**

16. **In the top left box, select Nucleic from the SELECT pull-down menu and ACGTU from the COLOR pull-down menu.**

   Your RNA structure is now color-coded according to the nucleotide present at each position on the chain. The color scheme is listed in the middle box on the left. Clicking on a residue will display its identity in the Command box (the bottom box on the left).

**Procedure 2C Formal Response Question**

   a) Compare the order of the nucleotides in this 3-D structure to those predicted by Mfold. How accurate was Mfold’s prediction?

**Procedure 2C continued:**

17. **Click on the How to tell DNA from RNA link in the middle box.**

18. **Read the information and click on the Click here to distinguish DNA from RNA link.**

   What is the distinguishing factor between DNA and RNA?

19. **Click the Back link at the bottom of the middle box.**
20. Examine the alternative structures of 1BN0 by clicking on the Advanced Explorer link and then NMR Models/Animation.

21. Click the OK button in the pop-up box that asks you to apply rainbow colors.

22. Select the Wire radio button.

This will enable you to see the interaction of the RNA bases in 3-D.

23. Click the Model+ button to switch from one model to the next.

Procedure 2C Formal Response Question

b) What do you notice about these structures? Why are there 11 possible structures in Protein Explorer but only 1 structure according to Mfold? What is a possible explanation for the alternative structures?

Procedure 2C continued:

24. Click the Mol Info link in the top left box.

The Molecule Information Window appears.

25. Click the Seq3D link.

26. Click the Show All as Thin Backbones button.

27. Choose Stick from the first pull-down menu under Show clicked residues in:

28. Choose Chain Colors from the second pull-down menu.

29. In the bottom box on the left, click once on each of the blue A residues.

30. Add all of the other residues using the same procedure. Follow this color scheme:

Brown – U
Green – G
Magenta - C

Rotate the structure using your mouse and look at the residue interactions in the hairpin loop.
Procedure 2C: Formal Response Question

c) What do you notice about the orientation of each of the individual bases involved in the tetraloop (G9, G10, A11, and G12)? How could you explain their orientation in the RNA hairpin? How is this view different from the one provided by Mfold?

Procedure 2D: Studying Evolution with Ribosomal RNA

Adapted from pages 379, 381-385, 390-391, 404-406 from Bioinformatics for Dummies (2003).

Protein synthesis (translation) occurs at ribosomal sites throughout the cytoplasm of a cell. A ribosome is made up of a small subunit and a large subunit, each made of ribosomal RNA (rRNA) and protein (Figure 4). In prokaryotes, the rRNA strands are called 5s, 16s, and 23s, which refer to their sedimentation behavior in a standard assay procedure. The eukaryotic versions are a little bit larger, at 7.5s, 18s, and 26s, although they are functionally equivalent otherwise. Generally, the RNA found in the small subunit of the ribosome is called SSU rRNA, while the RNA found in the large subunit is called LSU rRNA.

Figure 4. A ribosome has both a small subunit and a large subunit. Ribosomal RNA and protein compose the subunits.
http://www.brooklyn.cuny.edu/bc/ahp/BioInfo/TT/Tlatr.html
During protein synthesis a ribosome moves along an mRNA molecule, reading the codon and adding the correct amino acid to the growing protein (Figure 5). When a stop codon is reached, translation ceases, and the mRNA and protein are released.

RNA is present in all living species and organelles and is conserved in structure and function over a wide range of organisms. This makes RNA the perfect molecule to use in studying evolutionary relationships. The purpose of phylogeny is to reconstruct the history of life and explain the present diversity of living creatures, most often represented as a genealogic tree (Figure 6). Phylogenetics is a special kind of phylogeny that relies on the comparison of equivalent genes coming from several species for reconstructing the genealogic tree of these species and finding out who is the closest relative of whom in the family, discovering the function of a gene, or retracing the origin of a gene.

Figure 5. In translation, the ribosome moves along the mRNA and directs protein synthesis. http://www.emc.maricopa.edu/faculty/farabee/BIOBK/BioBookPROTSYn.html

Figure 6. A representation of the universal phylogenetic tree illustrating possible web-like connections resulting from lateral transfer of genetic material. http://www.nsf.gov/od/lpa/forum/colwell/rc02_swedish/tsld022.htm
Phylogenies are usually presented in the form of a tree, which is a graph composed of nodes and branches (Figure 7). Terminal nodes represent different taxonomic units, which could be species, populations, individuals, or genes. They are sometimes referred to as operational taxonomic units (OTUs). Internal nodes denote the divergence point or ancestral unit; genetic divergence occurs with a buildup of differences between populations, eventually resulting in speciation, or the formation of two distinct species from a common ancestor. A root defines a true biological phylogeny. It indicates the common ancestor of all the OTUs in the tree. Branches (also sometimes called edges) describe the descent and ancestry relationship among different nodes. In scaled trees, the branch length corresponds to evolutionary time. Each edge of the tree has a certain amount of evolutionary divergence associated to it, defined by some measure of distance between sequences.

![Figure 7. Terminology in a phylogenetic tree.](http://www-hto.usc.edu/~cbmp/2002/PhylogeneticAnalysis/tree%20terminology.htm)

Phylogenetic analysis depends on the comparison of sequences from related organisms. We can compare sequences from several species using a technique called multiple sequence alignment. Comparing nucleotide or protein sequences allows us to distinguish
whether residues have changed or remained the same from species to species. Less important residues change more easily, while important residues are not allowed to mutate. When you look at a multiple alignment, you can make the hypothesis that conserved positions (columns where all the sequences contain the same nucleotide or amino acid) are more important for the function than non conserved positions (columns where the sequences contain different nucleotides or amino acids). Of course, you could tell this from the alignment of just two sequences, but using more sequences makes it easier to discriminate between important and less important positions.

The sequences in a multiple alignment must be similar to each other (for example, they must share a common ancestor), but they should not be so similar that you can’t deduce any useful information from the alignment. You can use nucleotide sequences if they are more than 70 percent identical; anything less than 70 percent and you should use protein sequences. The best method for obtaining sequences for a multiple alignment is to BLAST one sequence and use the top hits from your results (make sure to check the percent identity from the top matches). ClustalW is the most popular program for making multiple alignments and similarity clusters. It uses a progressive algorithm in which the sequences are added little by little until the complete multiple sequence alignment is finished. In this lab, we will use a server that does a BLAST search on a ribosomal database, makes a multiple alignment of the top BLAST result sequences, and builds a phylogenetic tree from the results. All we have to do is provide an rRNA sequence – it couldn’t be any easier! Of course, you’ll do all these steps for yourself in Lab IX, but this will give you an introduction to what phylogenetic analyses can do for you.

1. Access and copy the nucleotide sequence (in FASTA format) of GenBank entry X03205, the entry for the human 18s rRNA.


This resource contains a database with all known ribosomal RNA sequences for both the large and small subunits.

3. Click on the Quick Phylogeny link on the left.

4. Read the Quick Phylogeny Search information.

This explains the basic idea behind your search query. It is important to note that the results from this database are not always accurate! We will use more sophisticated software to analyze protein sequences in the next unit.

5. Paste the sequence into the box and type in a name for your sequence (X03205).
6. **Click the Search button and wait.**

You may have to wait for several minutes before the analysis is complete. Just be patient!

**Procedure 2E: Understanding the Quick Phylogeny Results**

There are four options for viewing the results:

- **View alignment:** Your query sequence is displayed, along with several other similar sequences obtained from the BLAST search. The name of the organism is listed, along with the GenBank accession number. The numbers along the top of the alignment indicate the position in the nucleotide sequence. The sequence alignment is color coded to show conserved nucleotides:
  
  White – Not conserved
  Gray – Conserved for some of the sequences
  Light blue – Conserved for most of the sequences
  Dark blue – Conserved for every sequence in the alignment

- **View tree:** A phylogenetic tree generated from the multiple sequence alignment is displayed, showing the relationship between the organisms based only on the comparison between the nucleotide sequence for the 18s rRNA gene. Selecting the branch length values radio button displays the branch lengths on the tree.

- **Download the alignment in Clustal format:** The ClustalW format of the multiple sequence alignment displays the same information as in the first alignment, but without the color-coding. Completely conserved nucleotides are indicated by an asterisk (*).

- **Download the BLAST output:** The results of your nucleotide BLAST search are shown in the usual format, but without the graphic display. The bit score and E value are given for each hit, allowing you to verify the accuracy of this program. The actual alignments are also shown, with your query sequence aligned to each match.

**Procedure 2E Formal Response Questions**

a) Judging from the phylogenetic tree, what do the numbers in front of the organism names signify?

b) What other organisms share similar sequences to the human 18s rRNA gene? Which organisms are not as similar? How did you make these conclusions?
Refer to Chapters 12 and 13 in Bioinformatics for Dummies (2003) for additional information on RNA and phylogenetic analysis, respectively.

* A large portion of this laboratory was adapted from and text citations throughout this lab are for: Claverie, J. and C. Notredame. Bioinformatics for Dummies. Wiley Publishing, Inc., 2003.