Conventions for displaying gene sequences:

- Note in the databases that you visited for Part 3 of this lab, only the mRNA-like strand is displayed (complementary strand not shown).
- The sequence is always displayed 5’ to 3’, which is same direction that it will be translated.
- In euks, a cDNA sequence will reflect the sequence of the spliced mRNA and will therefore not include intron sequence.
- In euks, a genomic sequence will include introns and exons and adjacent regulatory regions – sometimes the introns will be indicated in lower case and the EXONS in uppercase.

Transcription of RNA from DNA:

<table>
<thead>
<tr>
<th>DNA</th>
<th>RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>T C C A A T G G C T T A T T T G C A</td>
<td>U G A G U A G G C A A T A A A A C G T</td>
</tr>
</tbody>
</table>

- The bottom strand of the DNA molecule above is the template for RNA synthesis.
- RNA polymerase makes a copy of the DNA sequence but substitutes uridine (U) in place of thymine (T).
TRANSLATION AND OPEN READING FRAMES

In principle, any given stretch of DNA has six potential translational reading frames: 3 on each strand. The diagram below shows the three possible reading frames –for the sequence that is displayed. Note that only one frame gives a long stretch of amino acids as the other two frames are interrupted with stop codons.

ORF = open reading frame: stretch of DNA that does not contain a stop codon in a particular reading frame

READ MORE HERE:
http://bioweb.uwlax.edu/GenWeb/Molecular/Seq_Anal/Translation/translation.html

SEE IMAGE NEXT PAGE: There are many sites where you can translate nucleotide sequence. Here’s one (next page) that is easy to use. Note the settings. Since we are translating sequence in the middle of the gene, we don’t need a translational start codon (ie AUG methionine), so any codon will do for the beginning of the reading frame. Since you know that you are working with the mRNA-like strand (see diagram on pg 1), then you can choose to have the program translate only the direct strand which means that it will translate the input
sequence but not look at the complement to the input sequence. Also, tell the program to translate all three reading frames since the first base of the sequence could be the first, second or third base of a codon in the correct reading frame. Finally, since we know that your entire sequence is coding sequence, tell the program to look for ORFs that are at least 100 codons long (we actually expect much longer ORFs -- you can do the math and figure out how long the ORF could be for each sequence file). This will eliminate the nuisance of sorting through short ORFs that are in the wrong reading frame.

http://www.bioinformatics.org/sms2/orf_find.html
The translation of sequence #1 from your set is shown below.

- Copy each translation and rename it so its identity is not lost. Be sure to keep the Fasta format with the > preceding the name.
- Note that translation terminates before the end of the sequence (see UGA stop codon at base 552). This will be important to think about when you address Part 3 in the sequence workup instructions.

**ORF Finder results**

Results for 644 residue sequence "1-rpoB-S 843 47 644 0.05" starting "CCGATTTCCG"

>ORF number 1 in reading frame 1 on the direct strand extends from base 1 to base 552
CCGATTTCCGACGACTGAAAGAGTCTTCTCGGTCCAGCCAGCTGTCCTACGTATTGACAC
CAGACACACCCGCTGCTGTGAGATTACGCACAAACGGTGTATCTCCGCACTCGGCCGACG
GGTCTGAGCCCGTGAAACGGTCAAGCTGCTGAAAGACGTCACACCAGCTCACTACGGT
CCGATATCGTCCATTATGAAACCCCTTGAGTGTCGCACAGTACATTCGTCAGTCAGAC
GTGTACGCACAGACTAAAGCAATACGGCTCTATCGACTCGTGATACGTAAAGTGAACGAC
GGGTGGTGTAACTGACGAAATTTCACTAACCTGCTGTGGTATCGGAAGAGCACAATAGATGTC
GCCAGCGGAACCTCAAGCCCTGATGAAGAGCACTACGTGAACGTGTAACCTTGCTG
CTGAGAAAGCCGAATCAGCATTGTTTCTGACGTGAGATGGACATACATGGACCTATCC
ACCGACAGGTGTATCCGTCGCTGCTCTATCGGATACCGTACGAGTACAGCC
ACCGTGCAATTTG

>Translation of ORF number 1 in reading frame 1 on the direct strand.
PISAAVEKEPGSSQLSQFMQNNPLSEITHKKRISALPGGLTREAGPFEVDHVHPTHYG
RVCPLETPPEGPQINLINSLSVYAYQTEYGFLETPYRKVTDGVVTDEIHYLSIAEBGNVVI
AQANSNLDEEEGFVEDVTCRSGKESSLFSRDQVDYMDVSTQQVSVGASLIPFLEHDDA
TVH*

No ORFs were found in reading frame 2.

No ORFs were found in reading frame 3.