Biology 382 Spring 2009 - Sequence Cleanup and Analysis of PCR clones.

1. **Paste your sequence into a MS Word document** (or other word processing program).
2. **Trim vector sequence from your clone sequence.**
   a. You can get a good first indication of which part of the sequence is likely vector by using **VecScreen** (http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html) at NCBI. Paste your sequence into the VecScreen window and hit return. Results will show suspected vector sequence graphically, and with alignments. VecScreen will give you a good general idea of where to start trimming. (With small inserts, vector sequence will be the majority of your sequence.)
   b. You can also determine the vector sequence by doing a 2 sequence alignment of your sequence with the L44440gtwy-postLR sequence. (If you are still working with a pCR8/GW/TOPO clone, you need to compare to that vector sequence.) Note that you may need to reverse complement your sequence (or the vector sequence) – use the **Reverse Complement tool** via the appropriate link.
   b. Make a second MS Word file using "Save As" and with name including "-trimmed."
   c. Select and delete vector sequences from this file. Remove sequence from both the front and back. This should leave you with a much shorter sequence only from the insert. Confirm the location to trim by comparing to sequence of multiple cloning site of vector L44440gtwy-postLR (or pCR8GW-TOPO, depending on your clone). You may also find it useful to search for the EcoRI sites (GAATTC) surrounding the insert region, bearing in mind that the insert may also have such sites (you should know by now whether it does or not).
3. **BLAST your trimmed (insert) sequence** against the "nr" database (default) using **blastn** (megablast – the default). Select C. elegans for organism. (To show alignments after cutting and pasitng into MS Word, change the font to Courier, and reduce the font size to 10 point or less. Adjust the margins outward.) Confirm the identity of your insert sequence.
4. **Translate your insert sequence** in 6 frames using the **EXPASY DNA to Protein translation tool**. Select the "With nucleotide sequence" output format. Choose and copy the translation that matches the translation from the Blast search that yielded significant matches. Paste this sequence into your Word file with the insert sequence. Note that your sequence may have an intron. Use the known sequences of your gene of interest, and the worm intron consensus sequence to determine exon vs. intron, and map them in your sequence. To ease some analyses (especially where 'manual' examination is used), you may want to generate a reverse complement of a sequence using the **Reverse Complement tool** via the appropriate link. Realize that your cloned sequence can be in either orientation in the vector.