Biology 482 Spring 2008 - Final Project Lab Sequence Analysis and Report Instructions

1. **Paste your sequence into a MS Word document** (or other word processing program).
2. **Trim vector sequence using VecScreen** (http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html) at NCBI. (With small inserts, vector sequence will be the majority of your sequence.)
   a. 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.
   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 relatively short sequence (200 - 400 bp) only from the insert. Confirm location to trim by comparing to sequence of multiple cloning site of vector pCR8GW-TOPO. You may also find it useful to search for the EcoRI sites (GAATTC) surrounding the insert region, bearing in mind that the insert can have such sites.

3. **BLAST your insert sequence** against the "nr" database (default) using **blastx** (nucleotide query, translated to translated database). Examine and save the top three or four alignments to a Word file. (To show alignments, change the font to Courier, and reduce the font size to 10 point or less. Adjust the margins outward.) If there are no significant matches via this method, try a **blastn** search. For both types of searches, you may use the default parameters. Which gene does your insert sequence match best?

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 homeodomains, and the worm intron consensus sequence to determine exon vs. intron, and map them in your sequence.

5. Use the **SIM - Alignment Tool for protein sequences at EXPASY** or 'manually' align the predicted amino acid sequence of your insert with known worm homeodomain sequences from *C. elegans* or other nematode. Prior to this, remove the sequences that are derived from the primers. These may not be the true sequences of the gene amplified (although they must be close), since the primer dictates the sequence (some mismatches with the genomic priming site are tolerated).

Each student should analyze two different clone sequences (both M13 forward primer and M13 reverse primer sequences). The complementary strand should give the same results and serve as confirmation. To ease some analyses (especially where 'manual' examination is used), you may need to generate a reverse complement of a sequence using the **Reverse Complement tool** via the appropriate link.

Report for the Project Lab.

1. Introduction – about 1 page on homeotic complex genes, with references. Note that this lab (including the degenerate primers used) was based on the Pendleton et al., 1993 paper. A selection of references on the evolution and function of animal and nematode HOM-C genes and complexes is found in the project lab folder. You are encouraged to find others appropriate to your introduction.

2. Materials & Methods – a summary of what was done (NOT a detailed description of the protocols, and not necessarily just what you did)) to amplify, clone and sequence the clones you analyzed. This requires you to select the most pertinent methods and results to explain (i.e., you don't need to say

how may microliters of something you pipetted, or what the tube was labeled). Try to keep this close to 1 page in length.

3. Results & Discussion – You may estimate the size of bands on your gels by eye. Include picutures of gels associated with the sequences that you analyzed. Then focus on the sequence analysis and indicate based on your analysis the best guess for the identity of the clones you analyzed (best match to a known *C. elegans* or other nematode HOM-C or homeodomain gene). Spend little or no time discussing what didn't work. (As many pages as needed, but need not be really long.)

From your sequence analysis (be sure to include clone names):

a. Isolated (no vector) insert sequence and (as appropriate) predicted translation, with intron and exon boundaries, if applicable. Also include a 'spliced' version with the intron removed, and the resulting translation.
b. Top two matches from blastx (and/or blastn)search. (Should include blasting with the complete predicted protein derived from your 'spliced' version sequence.)
c. Alignment of your predicted translations with one other (best match) HOM-C homeodomain (yours should be on top).