February 13 2008

Lab exercise 4 and Writing exercise 1

Annotation

Annotate one region of your *D. erecta* fosmid that contains one Genscan-predicted gene (e.g.: 26.1 from fosmid 26). You can choose any one of the Genscan predicted genes as long as it has at least 2 exons.

Write a report in which it should include the following information:

Analysis of the gene you chose:

You should present here a more detailed analysis of this gene, that whether it is a pseudo-genes or partial gene you find in your fosmid sequence. Remember that in flies (based on results from *D. melanogaster*), pseudogenes are rare, so it is unlikely (but not impossible) that you will find pseudogenes in your fosmid. Remember also that since the ends of your clones are randomly generated that it would not be surprising if the ends land in the middle of a gene and thus you will only find a partial fragment of the gene from *D. erecta.* Your best source for this information will be a combination of Genscan output and the *D. melanogaster* genome. It is quite possible, given the evolutionary distance from *melanogaster* to *erecta*, that this will not be a trivial exercise.

Include the accession number and name of any *melanogaster* gene that you consider likely to be homologous.

Be sure to investigate and discuss the Genescan predicted gene as well as any interesting blast hits.

For each gene you should determine as closely as you can the exact location of each exon. Construct a model for each putative isoform. You may or may not be able to infer the positions of the untranslated 5' and 3' ends of mRNA's. Comparisons between *melanogaster* and *erecta* may be most helpful in this regard. You will need to report the exact position of each exon in every gene as a table in an Appendix, and as a separate electronic file in the format provided. Remember when trying to precisely place intron/exon boundaries that introns (almost) always start with the two bases: 'GT' and end with 'AG'.

Function of the gene

Discuss the function of the gene and what they found in their final paper on annotation.

Do a Clustal analysis, find homologous genes from a variety of species, run a Clustal analysis on the protein sequence from at least four different species, and report on the results.

Appendix

Include various sequence files that will be needed in subsequent analysis of the compiled results. For the predicted gene you should append three files:
1. A fasta formatted file of the protein sequence.
2. A fasta formatted file with the nucleic acid sequence which codes for the protein (make sure it translates!)
For your Clustal analysis within Drosophila include any fasta sequence you used in your analysis.