

Drosophila Annotation: Final Presentation and Written Report, Spring 2006

The final will be made up of two parts, a 10 minute oral presentation and a written report.

The oral presentation:

Give a presentation in which you briefly discuss each of the features of your sequence(s), the distribution of repetitious elements, and synteny with *D. melanogaster*. You should discuss your best determination of the exact nature of each feature (possible gene) and your reasoning that led you to this conclusion. Remember there is no right answer; this is the first time that much of this sequence has ever been looked at by anyone. The goal is to come to the most reasonable interpretation of the data and be able to articulate the reasoning behind your hypothesis. While the discussion of genes is important, remember the long-term goals of the research are to help elucidate features of genome organization, especially as it relates to euchromatin vs. heterochromatin. Thus you will also want to discuss the analysis of repeats, and issues of synteny with *D. melanogaster* and other *Drosophila* species available. Note that *D. mojavensis* is the most closely related species among those for which sequence data is available.

You may use Powerpoint (recommended), overheads or simply draw on the blackboard, but figures of some kind are strongly recommended as they always play an important role in getting your point across to your audience.

The written report:

The second part will be your written report. Five main sections (subdivided as you see fit) are recommended: Overview, Genes, Clustal Analysis, Repeats, and Synteny. We also ask that you include an appendix with appropriate data files (see below). Your report should be submitted both as hard copy and in electronic form. The report should integrate text with figures, making a single, coherent document. This document will be the basis for future research on this topic by following Bio 4342 classes, will be the source of information for our joint publication, and will be used (with your permission) to demonstrate what undergraduates can accomplish in a challenging research endeavor.

Overview

The overview should provide a brief summary of all your conclusions. This should be similar to and about the same length as a typical abstract found at the beginning of a scientific paper. However, unlike an abstract, this section should include a figure of the entire sequence you are assigned (indicate overlapping sequence from last year's work, if appropriate) that indicates the size, position and identity of each feature. Attempt to keep this section concise and informative.

Genes

This section will be larger, depending on your fosmid. You should present here a more detailed analysis of any genes, pseudo-genes or partial genes 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.*

virilis. Your best source for this information will be a combination of Genescan output and the *D. melanogaster* genome. It is quite possible, given the evolutionary distance from *melanogaster* to *virilis*, that this will not be a trivial exercise. For each gene in your fosmid include the accession number and name of any *melanogaster* gene that you consider likely to be homologous. Be sure to investigate and discuss each 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; since it is unlikely that you will have information on untranslated 5' and 3' ends of mRNA's, you will probably only be able to report the position of the coding sequences. However, if you can generate any evidence as to the location of 5' or 3' UTR's, please feel free to include this. (The other *Drosophila* species sequences may be helpful.) 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 always start with the two bases: 'GT' and end with 'AG'.

Clustal analysis

As we will discuss in class, we would like you to do a Clustal analysis on at least one gene, and to attempt this analysis on a promoter region found within your fosmid. In analyzing a gene, 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. In analyzing a promoter, take 1-2 kb of sequence upstream of the start of translation from a series of homologous genes and use Clustal to look for any conserved upstream elements. You may find this page from J. Kadonaga's lab helpful (<http://www-biology.ucsd.edu/labs/Kadonaga/DCPD.html>).

Repeats

The analysis of repeats is most interesting with respect to the questions of heterochromatin and euchromatin. Prepare a table that lists all the repeats that you found within your fosmid. Include both those repeats picked out by Repeatmasker (generally remnants of repetitious elements; use the reference set tailored for *D. virilis*) as well as any sequences that you identified by pair-wise BLAST as possible repeats. You should also calculate and report the percentage of repetitive DNA.

Synteny

Generate a map of your fosmid including genes and repetitious elements, and then compare this with a map of the *D. melanogaster* genome over the same number of kb centered around one or more of the genes in your fosmid. Note which chromosome the gene is on in *D. melanogaster*. If synteny has been preserved, the genes on your fosmid will all be from the same region of the *D. melanogaster* genome, and your comparison will be of one map to the other (two lines). If synteny has not been preserved, please include a comparison based on each gene in your *D. virilis* fosmid to the same gene and flanking regions in *D. melanogaster* (several lines, stacked up). Determine the frequency of genes and of repetitious elements (number per kb) in these regions of *D. melanogaster*, as well as looking for evidence of events (such as recombination) that occurred during evolution, as possible.

Appendix

For your appendix you should include various sequence files that will be needed in subsequent analysis of the compiled results.

For any 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!)
3. A fasta formatted file of the genomic region around the gene, include about 500 bases upstream and downstream of the coding sequence as well as the coding sequence.

For the last file be sure to include in the header the exact coordinates of the coding sequence found within the genomic sequence. For any new repeat include a fasta formatted nucleic acid file of the sequence of the repeat you found. For your Clustal analysis include any fasta sequence you used in your analysis. A electronic tab-delimited file of every gene found in your fosmid should also be submitted.

Note

Your ability to use this format may vary. Please consult the instructors concerning any questions about your particular sequence.

Good luck and good hunting!