

Research Problem, Spring 05
Bio 4342, Research Explorations in Genomics

The scientific problem we will tackle during spring 05 is a question of comparative genomics: can we distinguish heterochromatic and euchromatic domains based on sequence organization and/or characteristics of the genes in these different environments? In *Drosophila melanogaster* (the fruit fly), the small fourth chromosome (sometimes called the “dot” chromosome) is unusual in that it appears to be essentially heterochromatic- packaged in a relatively condensed form, replicated late in S phase, exhibiting no meiotic recombination, etc. An examination of the DNA sequence indicates that the 1.2 Mb arm has a normal gene density (82 genes), but a ten-fold higher frequency of repeated sequences (generally remnants of transposable elements) than the other chromosome arms, which are euchromatic. It appears that many of the genes on the fourth chromosome are associated with Heterochromatin Protein 1 (HP1), but can be expressed in this heterochromatic environment. While the dot chromosomes of most other *Drosophila* species examined do show an association with HP1, the dot chromosome of *D. virilis*, a very distant species, does not, suggesting that the dot chromosome in this species is euchromatic. Recent studies have suggested that heterochromatin formation is targeted by the presence of repetitious sequence elements, although it is not clear whether all repetitious elements can trigger heterochromatin formation.

During the past summer, we have screened a gridded genomic library from *D. virilis* to find fosmid clones that contain one of the dot chromosome genes from *D. melanogaster*. Previous work by others has indicated that genes found on the dot chromosome of one species are likely to be present on the dot chromosome of other *Drosophila* species, and work by Bio 4342 students last year supports this. Each student in Bio 4342 will take on the challenge of sequencing and analyzing two overlapping fosmids from this set (about 60 kb). Our pooled data should tell us much more about this interesting chromosome.

In annotating your fosmid, you will want to address the following questions: What genes and/or pseudogenes are present? What is the gene density? What types of repetitious elements are present, at what density? (We will use both Repeat Masker and an internal comparison to address this question.) What can be said from a comparison of this region (defined by the genes present) among *D. melanogaster*, *D. pseudoobscura*, *D. virilis* (your data), and any other *Drosophila* species for which sequence data are available? We will then want to pool our data, and consider the answers to these questions for the pooled data, in particular looking at the relative spacing of genes and repetitious elements. If possible, we will also want to look for potential regulatory elements, in particular noting differences between the dot chromosome genes of *D. virilis* and those of other species, based on the hypothesis that genes that function within a heterochromatic environment might exhibit special characteristics. The answers to these questions should tell us more about the relationship between DNA sequence organization, chromatin packaging, and gene regulation.