

BIOINFORMATIC ANALYSIS OF DNA SEQUENCES

The DNA sequence “read” from the gel is your entry into bioinformatical analysis of the insertion site of the transposon you have analyzed so far this semester. The following outlines how you can use that sequence to identify the insertion site in the *Drosophila melanogaster* genome, identify nearby genes, and analyze them. In conjunction with the β -galactosidase expression pattern observed in your transposon line, you may be able to deduce a role(s) of the enhancer that drives expression of the transposon reporter in your line.

Read the sequence: Read your ascending sequence beyond the transposon sequences on your gel (i.e., beyond the 31 bp IR sequence of P or beyond the *Sau*3AI or *Hir*P1I site in the sequence). Do this for both sequences (representing genomic DNA at either end of the transposon)

Carry out a BLAST search, using your sequences: Go to either the NCBI BLAST server (through <http://www.ncbi.nih.gov>) or the FlyBase server (through <http://flybase.bio.indiana.edu>). Enter the first sequence and submit it to BLAST analysis (using the standard nucleic acid-nucleic acid search, BLASTn). You should get an extensive results listing, rank ordered from more perfect to less perfect matches. Repeat this for the second sequence.

Examine the matches and select: You should focus on *Drosophila* hits and, preferably, an annotated genomic scaffold sequence in Genbank (usually an “AE00xxx.x” file). This same file should appear on the output for both query sequences that you have used (because they are near each other in the genome). Examine the actual sequence alignment provided in the BLAST output of each sequence, noting the nucleotides in the subject sequence that correspond to your query sequence. Click on the highlighted address for that Genbank file.

Locate the sequence in the Genbank file: Scroll down the Genbank entry to the extensive annotations, which show the ordered locations of various genes in the scaffold. Find your sequence and note the name of the gene(s) within or between which it lies.

Use FlyBase to examine the structure and function of that region: Go to <http://flybase.bio.indiana.edu> and select the Genes or GadFly options; enter the symbol or name of the gene you found in Genbank file. You should now find an extensive page of information on that gene, including diagrams of the genomic vicinity of that gene, as well as its transcriptional organization. You should explore these sites (try both Genes and GadFly) to learn more about the region in which your transposon is inserted.