

Laboratory Worksheet, Friday, Feb. 6.

I. Pairwise Sequence Comparison. This week I would like you to make two sets of sequence comparisons. One is a query to database comparison and the other is a collection of pairwise comparisons.

For the first, compare CFTR to the database at PAM levels 40, 120 and 250. What are the best scores and how do they depend on the level? Do the optima, matches change with level? What could be an explanation?

Next look at cdata in the 548 resources directory. This is cytochrome-c data for a few species. Go through a similar experiment as above, but test human cytochrome-c against several of the others and see how things depend upon the evolutionary distance. Is the effect visible clearly here?

II. Putative Highly Expressed Genes (PHX). a) This week, begin writing a script which calculates the Karlin-Mrazek distance function between distributions. Write the pseudocode which brings a data file from a directory into your calculator. Set up a file which contains the results of the calculation and send your calculation there.

b) Realistically, we are not going to get all the species and all the genes done which were treated by Karlin-Mrazek. Let us discuss picking out three (tops!) species we would like to look at ourselves.

c) There is a lot done about codon bias in yeast, as noted earlier. Locate the data from yeast at Stanford. This should be listed last week or the week before.

You can also look up the CodonW program online, and see what analysis this offers: <http://www.hgmp.mrc.ac.uk/Registered/Option/codonw.html>. Is this limited to yeast data? What kind of analysis does it perform?

III. Molecular Viewers. A fix for now is the iMol viewer. I have a copy available locally to save everyone pounding on the distribution site at once. Go to 548 resources directory and download iMol.dmg to a folder on your desktop. This will disappear when you log out today, but you can transfer it to your ITD space, if you can use a Mac application elsewhere.

You should be able to click twice on the disk image and mount the disk. Then use it to view the molecules discussed last time, if you have not already done so.