

## Laboratory Worksheet, Friday, April 2.

**I. Putative Highly Expressed Genes (PHX).** There is an alarming lack of new data being entered into the workspace resources directory. Are people able to deposit things there, or are there permissions problems? According to the report I get from the site, everybody should now have permission to enter new material in these directories.

**II. Training an HMM.** Please be prepared to submit next time your results from training. This can be a simple text file printing out the parameters you obtained. Be sure to record also the options you used in the training program and the initial conditions for each run. There is help in Matlab for printing out to a separate file on your desktop or wherever you like.

### III. Genetic code evolution, early amino acids.

<http://www.math.lsa.umich.edu/dburns/548/rnasyn/>. There are a collection of materials there now concerning the temporal hierarchy and the evolution of the genetic code, especially by E. Trifonov. Roughly speaking, the synthesis version of this project would be to take the amino acid synthase of a protein “recent” on the consensus dating scheme (cf., p.9 of Trifonov’s lecture slides [.../548/rnasyn/trifITP.pdf](http://www.math.lsa.umich.edu/dburns/548/rnasyn/trifITP.pdf) in the subdirectory), and find that the consensus a.a.-synthase for that residue only involved “older” residues. (The Miller-Urey “abiotic” residues are highlighted by a red dot in the reference above.) There is data there which last year’s class helped gather on the TrpB-synthase across species (TrpB = the beta subunit of tryptophan synthase, the component directly responsible for the synthesis of tryptophan from serine and indole). It is found under [.../rnasyn/TrpBsynthasedata](http://www.math.lsa.umich.edu/dburns/548/rnasyn/TrpBsynthasedata) in the 548 resources directory.

There is still a valuable reference for novices on this subject in the paper of Weber and Miller from 1981, “Reasons for the twenty coded protein amino acids”. I have not yet gotten this old paper scanned and mounted in the 548 directory. Check early next week.

We will look at Trifonov’s version of the ordering of amino acid residues, and consider two questions: (1) do we have any methods we might use to confirm or deny the succession, or at least provide evidence? (2) Can we analyze the consensus sequence for such a synthase? Is it sensitive to species successions? For reading for next time, read the Trifonov paper enough to see where his evidence for the hierarchy is coming from.

The file [.../rnasyn/TrpBsynthasedata](http://www.math.lsa.umich.edu/dburns/548/rnasyn/TrpBsynthasedata) is incomplete in that it contains only prokaryotic species data. We will discuss collecting more from ExPasy (SwissProt) and from the Pfam collections.

### Appendix: Software Fixes.

**A. The Matlab Fix.** Go to the *548 Resources* link from the course home page. Download the file *KPM.tar.gz* onto your desktop. In the Mac’s these should mount themselves on the desktop. Now call Matlab from the dock (it is the brown cone shaped icon in the

bar at the bottom of the desktop). At the command line prompt (which looks a bit like >>) type `addpath ~/Desktop/KPM/HMM` [return], and so on, according to the names of the four downloads (.../KPM/KPMtools, KPMstats,netlab). This sets the paths to these functions for the permanently mounted Matlab in the lab. These mounted toolboxes will *disappear* when you log off. You can save them in your personal space, if you have a way to use them within Matlab there. I have not yet double checked the new `addpaths` since incorporating all the components for the HMM download into one local download.

*Please note that there is a Licence agreement which you are accepting for the Nabney netlab package.*

**B. The SeqAln Fix.** The local systems people have stored the Waterman program suite *SeqAln* locally at `http://www.math.lsa.umich.edu/courses/seqaln-2.0.dmg` as a disk image. Download this file to your desktop, where it becomes an icon. From the systems person here: “This is a MacOSX Jaguar binary, and the executable files can be found in that tree. So...students have to download and mount the dmg file and then type: `/Volumes/seqaln-2.0/seqaln/bin/SOMEPROGRAM` to run a program. Hope this helps.” This fix has been tested and works.