

Laboratory Worksheet, Tuesday, March 25.

0. Basic Lab Procedure (New) Start a new directory entitled 548_0325. Please keep in it the results of any computations which you do today which you do not keep in topic-titled files like "actin-HSP" or "Codon usage bias", for example. Please sort out whether you can securely transfer files back and forth between your space here and your home machine, if you are going to work on these computations at home, too. [This may mean depositing files at an intermediate ITD account, or uploading and downloading through the web.] The point is that I want you to have access to your current work on these projects during the Tuesday sessions.

Recall that you also have access to accounts on the math lab machines in East Hall. They are pretty well stocked, monitored and supported, and we can get more software uploaded there if people need it. I know that some of you have recently gone back to get this account opened and use those machines. Let me know what you want to do there and we can see that it is supported. The math department is very large and they are going through a system update now, but they still can get to support requests in decent time. They are Unix based machines, but if you can use the linux machines, the Unix machines are easy to adjust to.

In particular, the math labs all have Matlab mounted, though I don't know whether all your class accounts are "LSA enough" to get the Neural Nets Toolbox for you (I think yes, because LSA is responsible for your instruction). Please check to see whether you still have the initial account passwords distributed the first day of 548. I think I can get the software required to work on simple network problems mounted by next week.

I. The Actin-HSP link. Dormant this week, but there is an interesting link possibly on the horizon. See part II below.

II. tRNA Synthetases. This is to start a third project, on tRNA synthetases. This is related to the ancient history of the origins of the various amino acids, or rather, their incorporation into the standard repertory of life. This was in last week's assignment, but that seems to have gotten lost in the shuffle, so here it is again with a few modifications and (hopefully) improvements.

First, we have to collect some data. I repeat here the suggestion from last time, but you can get good data from the 548 web directory in the file `trnasynth16data`. This has the FastA data for 16 of the 20 acyl-tRNA synthetases for *E. coli*. You could try to improve this by searching for the remaining four. Glycyl-tRS is there, but there are two species.

Go to Entrez and look in the protein database with the query "tRNA synthetase, *E. coli*". Search for twenty tRNA synthetases, one for each amino acid residue. Try to get recent listings, and sequences for the whole molecule, as opposed to the active subunit. Ask for FastA display, click the ones you want and build up a text file that you will save to your home space. This should be a file which will be acceptable to aligners and phylogenetic programs.

Next, make a multiple sequence alignment for these sequences. You can refer to the old worksheets for addresses for this. Don't everybody use Pasteur. You might try the trick of making a smaller alignment and then aligning the rest to an HMM.

You could also try SAM at

<http://www.soe.ucsc.edu/compbio/research/HMM-apps/T-99-tuneup.html>, which enables you to derive a multiple sequence alignment from HMM training, something you haven't yet done. This can be time consuming, however.

If time permits, make a phylogenetic tree for these sequences, probably preferably with the Nearest Neighbor program so that there is a little more significance in the distances displayed. Again, you may have to cut back on the number of sequences you submit to avoid server overload or too long computation run-time.

The idea is that according to certain laboratory simulation experiments, there seems to be a partition of amino acid residues into those which are primitive and those which are more complex and more recent. The first group can be generated "spontaneously" from inorganic settings. I am looking to see whether you can see a cluster removed from the others which might correspond to the "ancient" residues. On the other hand, there is a more obvious breakdown of acyl-RS's according to "class": I and II. These are markedly distinct in their homologies, and they should be showing up in the alignments above. There are recent comments on this in Carter-Duax, letter to *Mol. Cell*, **10** (2002), 705-708. The classification is discussed in Eriani, et al., *Nature*, **347** (1990), 249-255, and Cusack, et al., *Nature*, **347** (1990), 203-206. [The classification (ten of each kind) can be deduced online from pdb.weizmann.ac.il/scop/data/scop.b.d.ch.b.b.html, which lists Class I, and links to structural analysis of these enzymes.] Later we will try to compare the whole family done today to subfamilies split along the lines of the Class I and II distinction, and the "inorganic/early" versus "organic/complex/later" distinction.

Reprinted from last week's edition, if needed...

Transferring Files to Home. This may not be the easiest way to do this, but here is a way to put files in your ITD personal/Public space, so that you can get them from home via the web, or an scp function if you have such. At any rate the point is simply to get them outside the BICC firewall.

You have to get yourself into the right directory on the linux command line. Then you type

```
scp <file or directory name in the current directory>
```

```
[[space]]<your unqiename>@login.itd.umich.edu:<target directory>. Here unqi-  
name is your usual ITD unqiename, your file or directory can be a directory. The target  
directory should be something like
```

```
/afs/umich.edu/user/d/b/dburns/Public/html/. [In Linux/Unix, leaving the "/" on  
the end means that the copying will be into the last directory named, and the naming of  
the directories and files to be copied will be the same as in the originals.] This enables  
you to access your file via the web. If you can scp from home, you just have to deposit  
your transfer from the BICC in any directory that is accessible from home.
```