

Laboratory Worksheet, Tuesday, March 18.

0. Basic Lab Procedure (New) Start a new directory entitled 548_0318. 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 intermediary 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.

I. The Actin-HSP link. First step, review worksheet from last time to see that you have completed the HMM part for the fused seed, as well as the alignment of the total fusion of the Pfam actin and HSP seeds.

Second, examine the output alignment visually for signs of significant alignment (un-gapped regions, regions of large residue conservation, etc.). Especially interesting, of course, would be regions where this similarity extends to sequences from both of the original families.

Finally, and this part should be saved for outside the lab, find an editor to give some added visual clues about your alignment, especially a logo or other method program which would quantitatively evaluate the significance of the alignment details.

II. Phylogeny Exercise: Outliers for "Rootings". As mentioned in class, often phylogeny programs produce unrooted trees, as in the example of the Pfam cytochrome-c seed alignment we tested in class Monday. (The FastA form of that seed is at <http://www.math.lsa.umich.edu/dburns/548/ccpfam>; this is in *ASCII* form – i.e., no HTML tags.) Here I would like to go back to the earlier cytochrome-c mini-dataset at <http://www.math.lsa.umich.edu/dburns/548/ccdata>. Download this file. It has several species (with informal names) and their cytochrome-c amino acid sequence data. Submit this family to a multiple sequence aligner. This can be done as earlier, using CLUSTALW, for example. Then run this alignment through a PHYLIP phylogeny program. Make sure that you put it through at least the **nearest neighbor method**, and use **drawtree** to print out a phylogenetic tree. Do you see an obvious outlier? Which would you expect from the original data set, where you recognize the species? Can you make sense yet of where (i.e., in terms of the distances marked on a nearest neighbor tree) the root should be? If this doesn't give at least an obvious tree topology, while not specifying the distance

to the root node, try setting a true outlier e.g., by picking another cytochrome-c from the Pfam seed, of a species as far removed as possible from the seven we have here (e.g., *E. coli*), or a completely different protein, of roughly the same length. How does the rest of the tree get effected by the addition of these extra sequences?

Write this part up for next week, please.

III. 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.

First, we have to collect some data. 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.

If time permits, make a phylogentic 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 primitve 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.

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.
```