

## Laboratory Worksheet, Tuesday, March 11.

**I. The Actin-HSP link.** Let us complete what we started last week in this direction. Hopefully you have saved the multiple sequence alignment for your fused actin-HSP seed, and saved the HMM file you got from the PASTEUR Institut (or wherever you got it from). This week I would like to complete the alignment of the full Pfam seeds for the two families to the HMM model. We can do this with the version of HMMer which is installed in the lab.

We use the program *hmmalign* in the HMMer suite. The command has options which we would like to use, most especially the “save to file” option. This is simply the addition of `-o <filename>` to the linecommand. The basic command is simply `hmmalign [options] <hmmfile> <seqfile>`. This will align the file `<seqfile>` using the model specified by `<hmmfile>`, which will have the `.hmm` extension for HMMer. The file that is produced will be designated by a `.ali` extension.

If you have trouble with the command line local version, you may try next the Pasteur Institut site <http://bioweb.pasteur.fr/seqanal/motif/hmmer-uk.html>. Be sure to save your completed alignment.

**II. First Phylogeny Exercises.** We will just make some first uses of the range of programs available in Felsenstein’s PHYLIP package. Again, the friendliest server site seems to be at Pasteur: <http://bioweb.pasteur.fr/seqanal/phylogeny/phylip-uk.html>. I would like you to cut out a small core, the best core, of highly aligned sequence from your fused model for actin-HSP above, and let us try to perform protein sequence phylogeny analysis on it. This should mean today that we just run it through the sequence of programs: `protdist` will create a matrix of distances between the sequences. (Be sure to save the outputs in these steps, since they become inputs for later steps.) Then use this data in the program `neighbor`. Also run the sequence data through `protpars`, which will run the parsimony method on the data. Finally, you can get a visual presentation of the results from `drawtree`. There is good documentation available online for each of these steps at <http://www.hgmp.mrc.ac.uk/Registered/Help/phylip/phylip.html>, which is from the UK Human Genome Mapping Project Resource Centre, Hinxton, Cambridge. Felsenstein’s home for PHYLIP is at <http://evolution.genetics.washington.edu/phylip.html>.

We will later use the more intensive ML methods and also try the T-HMM technique which Bin Qian described this week.

**III. Codon Usage Bias Project.** I want to ask this week about where these programs stand at the moment, that is, check whether they are running up to the point of being able to perform the functions we have to calculate from the Karlin-Mrázek paper, at least on a single file which you have saved locally. This might be best to do towards the end of the hour, so that if your program is running up to this point, you won’t need to stay. The next step will be to write scripts to get the programs to upload data automatically, so that you can go through larger collections of data.

**IV. Alignment Editing.** I repeat below the text of one of the points in a previous

lab. If there is any time today, we could try to edit small blocks of the seed or HMM-generated alignments for the actin-HSP test case, to see conservations and information contents highlighted. Take care of the tasks above first, however.

There are some interesting alignment “editing” tools available on the web. The two I have in mind at the moment are: motif logos which produce visual summaries of the entropic profile of a piece of multiple sequence alignment. One is available at <http://www-lmmb.ncifcrf.gov/toms/sequencelogo.html>. Actually, examples and documentation are available there, and the logo maker is available via a link. CINEMA helps display information in a multiple sequence alignment visually. It can be accessed through <http://www.biochem.ucl.ac.uk/bsm/dbbrowser/CINEMA2.02/kit.html>. This requires installation, however, and I will try to get it installed by the week after the break. [This has NOT gotten installed!]