Math/Stats 548, Winter 2004:
Computations in Biological Sequence Analysis

Laboratory Worksheet, Friday, April 16.

I. Putative Highly Expressed Genes (PHX). I would like to schedule a demonstration of the tools you are putting together. The obvious time would be the period of the exam for 548. This is possible if we can schedule presentations around that time. We can set a time today. This would require working together to solder the parts of the project together that are being gathered in the workspace.

For further litertature on this topic, there is a survey by Karlin, Campbell and Mrázek on comparative DNA analysis, including PHX. This is a 1998 paper, and there are later references. The review by Karlin et al. is "Comparative DNA analysis across diverse genomes", *Ann. Rev. Genet.* **32** (1998), 185-225.

Two of the specialty areas (model organism) where the HX genes idea has been followed most closely are *Drosophila* and yeast. To begin searching out information in this direction, google `yeast ''codon useage bias''`, and *mutatis mutandis.*

II. Genetic code evolution, early amino acids.

There is a large body of work around the evolution of the genetic code. A very friendly introduction, and not too false!, may be found in the current issue of *Scientific American*, written by Steve Freeland and Laurence Hurst. (April 1, 2004: "Evolution Encoded"). Further references are included there. The UM library has this online: try `http://www.sciamarchive.org` and search for "freeland" from a UM IP.

A quote from the last edition of *548 Lab Worksheets*: "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." [My omission; forgot to do this!! Coming soon! Note, however, that it will be mounted in the 548 workspace, not in the 548 resources directory, under resources, library reserves.]

What did people do on this project? What are the data looking like?

III. End of term procedures.

There are two points here.

(a) We will tentatively meet, either in the lab room or in the lecture room, at 6 PM on Thursday, April 22, to check together whether we have a working pipeline for running the PHX experiment and variants. You might want to touch base with each other beforehand about this.

(b) Please turn in whatever you have worked up about the tryptophan synthase concensus sequence in the **Dropbox** associated to your name in the class workspace online. Also, if you are missing the dice data write-up, you can submit it directly online that way.