

# BIOL 498/CMSI 698

## SPECIAL STUDIES: BIOINFORMATICS Spring 2006

### **Assignment 0119**

#### ***Nucleic Acids Research* Database Issue Assignment**

Each year, the journal *Nucleic Acids Research* (NAR) devotes the first issue in January to biological databases. The goals of this assignment are the following:

- to evaluate the usefulness of two biological databases, from the standpoint of a biologist (as well as you can imagine) and a computer scientist
- to report your evaluation to the class with a demonstration of the database
- to become familiar with the range of databases available through hearing all of the presentations

The issue can be found at: [http://nar.oxfordjournals.org/content/vol34/suppl\\_1/index.dtl](http://nar.oxfordjournals.org/content/vol34/suppl_1/index.dtl)

1. Read the editorials by Alex Bateman, and the first article (a commentary) by Michael Y. Galperin.
2. You will be assigned one major database to evaluate and will choose a second database to evaluate from among the others described in the NAR issue. You will need to sign up for your databases by the end of class today.
3. Prepare a 10-minute PowerPoint presentation for class next week that includes the following for each database:
  - What is the purpose of the database? What biological information does it contain? What species are covered in the database? What biological questions can it be used to answer? What type (or types) of database is it (sequence, structure model organism, or specialty [what?]; primary or “meta”; curated electronically, manually [in-house], manually [community])
  - What individual or organization maintains the database? What is their funding source(s)? Is there a license agreement or any restrictions on access to the database?
  - How often is the database updated? Are there links to other databases? Can the information be downloaded? In what file formats?
  - Evaluate the “user-friendliness” of the database. Is the Web site well-organized? Does it have a help section or tutorial? Run a sample query. Do the results make sense?

Include in your presentation a quick tour of the database Web site demonstrating its features.

4. Provide relational database schemas for your assigned databases. This can occur one of two ways: either the database designers have already specified this, or you will have to infer this based on the available information. Either way, submit what you can find or produce. Identify any missing, redundant, or conflicting data.

#### **Additional Readings for January**

1. Chapters 1, 3, and 4 in Baxevanis/Ouellette (one or more of these may actually help for this particular assignment, so look them over at the very least)
2. “Bioinformatics — An Introduction for Computer Scientists” by Jacques Cohen, *ACM Computing Surveys*, June 2004