**35%: Final project and presentation – Read Both Pages of This Document**

10%: Your project program performs the stated function(s).

Due Mon June 5th 11:59pm.

You will turn in your class project by having your code pipeline in your home directory on crick.oregonstate.edu in a folder called YourLastName\_Project (e.g. Megraw\_Project). Inside your project folder, you will place a file called README.txt, which tells me the steps to run your scripts (you can just do this for your Class\_1 sequences for example, then I will know how to repeat the process to produce your Class\_2 feature table from those sequences as well). If you use a classifier script other than quickSVM as provided, point this out to me as well.

For example, you might have your fasta sequences in a subdirectory called OrigSeqs, your splitter in a subdirectory called SplitFasta, your SGE files for feature table generation in a subdirectory called GetFeatures, and your feature table concatenator in a subdirectory called MakeFinalFeatureTables. Whatever organization works for you is fine-- then in README.txt, you would tell me which scripts (in which folders) to run, and in what order, to produce your Class\_1 (or alternatively Class\_2) feature tables and run the classifier to compute accuracy (quick\_svm.R if this is what you chose to use).

**IMPORTANT:**

**Put the commands you want me to run into shell scripts** rather than telling me to type something at the command line in the README. For example, make a script set 1\_SplitFasta.sh, 2\_Generate\_Features.sh, 3\_CollectFeatureTables.sh, 4\_Run\_SVM.sh (or however you wish to organize) with the commands to perform these tasks. I will call these scripts in the prescribed order.

Please make sure that **all files** in your Project turn-in directory are **group-readable** (check file permission settings before turn-in) so that I can access them.

Please make sure that you have used only **relative path references** in all of your scripts, and all files necessary to run your project pipeline are in your Project Folder, so that your Project Folder is **self-contained** (i.e. I can copy this folder to a new location, run your scripts, and your pipeline will still work).

15%: 8-minute presentation BOT 499, 12-minute presentation BOT 599.

1. (~1 minute) State the Class\_1 and Class\_2 datasets chosen (organism, genomic entities), and why this choice was interesting to you.

2. (~3 minutes) Describe each feature that you computed over the course of the project (in all feature sets tried), and the reason why you felt that this particular feature might help distinguish the two different genomic entity types in your Class\_1 and Class\_2.

3. (~2 minutes) State the results that the best-performing classifier achieved, and why you think that the feature set used produced this outcome.

4. (~4 minutes—Grad-required, Undergrad-optional) Describe performance differences in different feature sets that you tried, and state biologically why you think that the best-performing feature set was more successful than other attempts.

Finally, state or re-cap the aspects of the project that challenged you the most. What did you learn from this project? Briefly, place the project into the context of a problem or task related to your lab/rotation research that you could attack with a similar approach.

5. (~2 minutes) Answer questions from the class and the instructor about your project.

10%: Attendance and participation at the project presentations of classmates during the last week of class (Week 10).

Attendance/participation is recorded and graded during presentation periods—full points are awarded for 100% on-time full-time attendance at all presentation periods with full attention to the presenter, in accordance with giving the highest respect for colleagues and receiving the benefit of maximal learning from others’ presentations.