PROGRAMMING FOR BIOINFORMATICS – BIOL 8803 B – Fall 2015

Instructors: ***Lu Wang, Lavanya Rishishwar*** *and* ***King Jordan***
School of Biology, Engineered Biosystems Building, 2200
lu.wang@gatech.edu *lavanya.rishishwar@gatech.edu* *king.jordan@biology.gatech.edu*

Course summary: The fields of Bioinformatics and Computational Biology occupy the intersection of the life sciences and information technology. Over the last decade, there has been an explosion of data in the life sciences and the proliferation of raw information promises to continue at an even more rapid pace. Computers are needed to handle and assimilate this massive amount of information. More importantly, the role of bioinformatics is to convert information, in the form of data, into biological knowledge. In order to do this, bioinformaticists and/or computational biologists must be adept at the use of computers, *i.e.* **YOU MUST KNOW HOW TO CODE**.

This project-based/lab course will provide an introduction to programming for bioinformatics. We will begin by introducing you to the command line environment in the Unix / Linux operating system – this is where real scientific computing gets done. This will include a fairly broad coverage of Unix / Linux utilities as well as shell scripting. The course will then go on to use the Perl programming language to illustrate the fundamentals of bioinformatics programming.

This class meets for lecture sessions on **Mondays from 4:05-5:55 pm in ES&T L1105** and for lab sessions on Thursdays from **4:35-6:25pm in Cherry Emerson 206**. All required and recommended readings, lectures and exercises will be made available on the course T-square site. This is an exclusively practical and active learning class. Students will complete exercises in order to learn how to code and how to do bioinformatics. The only way to learn the course material is by doing. Accordingly, attendance and participation are mandatory and critical. Students are required to attend lecture sessions on Mondays; open laboratory sessions on Thursdays are optional and intended to provide additional support for code development. Students who show up late or miss lecture sessions will be severely penalized. Participation in lecture sessions will be judged by the degree to which each student participates in class discussions and exercise sessions. Students will also have the opportunity to demonstrate and explain their code to the class. Students will also be required to post their code to the course T-square site for evaluation.

Please see *www.honor.gatech.edu* for Georgia Tech's Academic Honor Code, which you are required to uphold.

Course Evaluation:

Class participation (attendance)	20 %
Class demonstrations	20 %
Code evaluation	60 %

Schedule of lecture / lab sessions

Date	Topic	Room
Mon Aug 17	Introduction to *nix environment	ES&T L1105
Thu Aug 20	Open lab session	CE 206
Mon Aug 24	Basic system administration in *nix	ES&T L1105
Thu Aug 27	Open lab session	CE 206
Mon Aug 31	Web access & file handling	ES&T L1105
Thu Sep 3	Open lab session	CE 206
Thu Sep 10	Open lab session	CE 206
Mon Sep 14	Utility compilation and installation	ES&T L1105
Thu Sep 17	Open lab session	CE 206
Mon Sep 21	Regex and advanced file handling	ES&T L1105
Thu Sep 24	Open lab session	CE 206
Mon Sep 28	Pipelining basic and shell scripting	ES&T L1105
Thu Oct 1	Open lab session	CE 206
Mon Oct 5	SNP calling pipeline	ES&T L1105
Thu Oct 8	Open lab session	CE 206
Thu Oct 15	Open lab session	CE 206
Mon Oct 19	Introduction to Programming Concepts and Perl	ES&T L1105
Thu Oct 22	Open lab session	CE 206
Mon Oct 26	Perl for Bioinformatics	ES&T L1105
Thu Oct 29	Open lab session	CE 206
Mon Nov 2	Regular Expressions in Perl	ES&T L1105
Thu Nov 5	Open lab session	CE 206
Mon Nov 9	Syntactic sugars and modules	ES&T L1105
Thu Nov 12	Open lab session	CE 206
Mon Nov 16	Object Oriented Programming in Perl	ES&T L1105
Thu Nov 19	Open lab session	CE 206
Mon Nov 23	High-Throughput Analysis	ES&T L1105
Mon Nov 30	Last Class & Review	ES&T L1105
Thu Dec 3	Open lab session	CE 206

Note that the syllabus is subject to change depending on the speed at which the class progresses and the performance of the students.