Practical Bioinformatics -- BIO 490S – 02

Duke University, Fall 2012

Instructor	Carrie Olson-Manning
Email	cfo3@duke.edu
Office	FFSC 3335
Lecture	3:05-5:35 Mondays, FFSC 2237
Office hours	1pm-4pm Thursday, FFSC 3319

Requirements

Book	Practical Computing for Biologists, Haddock and Dunn
Access to a laptop	Mac or PC, it doesn't matter. You'll need to bring your laptop to class every day.

Course Syllabus

Bioinformatics is the application of computer science to biology and medicine. The magnitude of biological data, at all levels from environmental to genomic, is growing exponentially. For example, in 2001 the first draft of the human genome was completed; five years from now there will be hundreds, if not thousands, of human genomes publically available (not to mention thousands of bacterial, fungal, animal, viral and plant genomes). Current and future biologists now require skills once only used by computer scientists to make sense of these vast amounts of data.

This course will introduce you to a varied sampling of publically available biological data and the basic skills to organize, manage and analyze that data. You will conduct an independent project based on your research interests and future goals.

Skills you will gain in this course

Software development
Reading and comprehending scientific literature
Generating and testing hypotheses with real data
Understanding of how to manipulate biological data
Simple statistics implementation
Use of a computing cluster

What I expect of you

You are learning a language in this course. Like learning a language, or a musical instrument, you won't absorb the material it if you only practice once a week! I expect you to practice nearly every day.

Turn in homework on time! This is tied to the previous expectation. If you aren't keeping up with the homework, you probably aren't practicing enough.

Completion of an independent project. We will discuss this more in class, but the best way to learn to deal with biological data is to have a dataset you want to analyze!

Help each other! If you can't find the problem with your code, sometimes you just need another pair of eyes. I will help you troubleshoot, but only after your classmates have had a look.

Use all available resources! There is no need to reinvent the wheel. First we are biologists, then we are computer programmers. Go online, ask your friends, find useful books. Chances are someone has already solved the problem you have (even if it doesn't seem that way at first).

What you can expect from me

I will provide example problems to work so you can get practice every day.

Useful feedback and timely grading of your homework.

I will hold a troubleshooting session once a week. You are likely to run into problems you can't solve on your own or with the help of your study group. Once a week I will hold office hours in a seminar room to help you troubleshoot.

Helpful discussion about your project. As you will quickly learn, my focus is evolutionary biology, but I can hopefully offer guidance for any type of project.

Grading

	Points
Homework (10 pts each)	120
Bioinformatics seminar	20
Dryad project	50
Independent project	110
Total points	**300**

* 14 or 15 homework problem sets will be assigned. I will drop the lowest 2 or 3 scores.

Course Policies

Attendance: You are expected to attend all lectures. If you have to miss a class, you are required to meet with me to get caught up on missed material. We have a lot of material to cover in a short amount of time. We only meet once a week, so each lecture is crucial.

Making-up missed assignment: I will drop the lowest 2 or 3 homework scores.

Missed deadlines: If you miss many lectures due to a long-term illness, authorized representation of the University, or extraordinary personal reason, you must substantiate this absence by submitting an official Dean's excuse. An absence due to a **short-term illness** requires that you complete an on-line short-term illness notification form. You can find the procedure for completing this form and a link to it at:

http://trinity.duke.edu/academic-requirements?p=policy-short-term-illness-notification.

Students with learning or physical disabilities: Students requiring special accommodations should identify themselves as soon as possible so I can accommodate your needs. If this may influence your performance in this course, please provide me with documentation from Learning Disabilities Services, as soon as possible.

Practical Bioinformatics -- BIO 490S – 02
Schedule Fall 2012

(Key: *CS* – Computer Science Topic, *ICE*- In-class exercise, *IP*- Independent project, *PCFB*- Haddock and Dunn, Practical Computing for Biologists)

Date	Lecture Topic	Readings	Exercises
Aug 27	Introduction to Bioinformatics	*PCFB* Chapter 1 Unix handout	Introduction to *IP* Reading scientific papers Paper assignment
Sept 3	Regular expressions (RE)	*PCFB* Chapters 2-3	Organize data from a public database *IP* ideas Paper discussion
Sept 10	Command-line Beginning Python Programming	*PCFB* Chapter 4-5 *PCFB* Chapters 7	RE in Python
Sept 17	Python Programming	*PCFB* Chapters 8-10	Public databases
Sept 24	Sequence Analysis	*PCFB* Chapters 11-13	Manipulating sequence data with PyCogent
Oct 1	Gene Expression	*PCFB* Chapter 16	Next-generation sequencing Microarray data
Oct 8	Data manipulation and sequence analysis	R handout: Guest lecture by Matt Johnson	R programming *IP* proposals first draft DUE!
Oct 15	--*Fall break*--		
Oct 22	Statistical Methods: Cameron et al. PNAS 2011		R programming *IP* proposals and feedback
Oct 29	Dryad repository and publically available data		
Nov 5	Group presentation on Dryad project		
Nov 12	IP first draft due		
Nov 19	IP working session		
Nov 26	IP working session		
Dec 3	IP in-class summaries		
Dec 10	Turn in final projects		