Fall 2012 BIOL B216: Introduction to Genomics and Bioinformatics

Joshua Shapiro
office: 112 Park Science
phone: 201-777-0217
email: jshapiro01@brynmawr.edu

Course Objectives

The recent proliferation of genomic data has transformed biology, making previously laborious and expensive experiments easier and cheaper, enabling new avenues of inquiry, and fundamentally altering our understanding of biology and medicine. This course will introduce you to the questions that can be asked and answered with genomic data, and to the computational tools available to analyze that data. The primary goals are:

- to learn how genomic data are being used to provide new insights throughout biology and medicine.
- to become familiar with the tools and databases available for bioinformatic analysis, with an appreciation of the quantitative concepts that underly those tools.
- to develop the ability to formulate and investigate genomic research questions, and to effectively communicate your questions, methods, and results.

Course Description

Lectures will introduce some of the common techniques and algorithms used in genomic analysis, including sequence alignment, BLAST, gene expression profiling, and prediction of protein structure and gene function. Throughout the course, we will explore how these techniques, and genomic data in general have been used to explore topics such as evolutionary history, genetic causes of disease, the cancer biology, and microbial ecology (metagenomics). Since genomics is a new and rapidly-changing field, we will emphasize topics chosen from recent literature, discussing both the scientific and cultural implications of the work.

The computational lab is an essential component of this course. That said, there is no assumption of previous experience beyond knowing how to move files around and use the web, word processors and spreadsheets. In the first set of labs, you will become comfortable with online databases, sequence alignment, gene expression analysis, and genome-scale data. The second half of the lab will be devoted to developing and completing a final project; you and a partner will work with a recent *Streptomyces coelicolor* gene expression data generated by Dr. Chander to formulate and answer a research question of your choice.

Lecture

Tuesday/Thursday 11:15am – 12:30 pm
Park 10

Lab

Thursday 1 – 4 pm
Park 10

Office Hours and Contact info

Tuesday 2–3:30pm, Friday 10-11:30am, and by appointment. I am quite flexible, but I live pretty far away, so appointments are helpful. My office door is almost always open when I am there, so you should feel free to drop by other times if you want to take a chance.

We will be using Piazza.com for out-of class questions and discussion, so any questions that might be of general interest should be asked there, rather than by email. This will give you the opportunity to see the questions that other students have already asked, and to answer each other's questions. Active participation on Piazza (questions and/or answers) will count toward class participation.

I will answer emails (and Piazza queries) fairly quickly during normal business hours, but generally less quickly in the evenings and weekends (possibly not until the next day).

Textbook and Readings

The textbook for the class is *Introduction to Genomics* by Arthur M. Lesk, but it is by no means a definitive source for the class. There will be additional readings from the primary literature and other sources on topics that the textbook does not cover in sufficient detail or for which it presents outdated information (all too common in a rapidly moving field.) All readings will be posted on Moodle.

Evaluation and Grading

Problem sets (4)	20%	
Exams (2)	20%	
Labs	30%	(10% from final project)
Final Paper	20%	
Class participation & presentations	10%	

Late assignments will incur a 10% penalty per day unless arrangements are made for an extension at least 24 hours prior to the due date.

Problem Sets

Problem sets will include both computational exercises and short-answer questions. You may discuss topics and technical questions with other students (and on Piazza), but all submitted work should be your own, with any assistance from other students clearly acknowledged.

Exams

Because this is a class that emphasizes the integration of information from diverse sources, all exams will be open-book, open-notes, and open-internet. In fact, many will require the use of internet resources. Exams will be take-home, and should take 2-4 hours, including time for use of online tools and databases. Exams must be completed individually.

Any scheduling conflicts must be addressed well in advance of the exam. Make-up exams will not be given, except in cases of illness or unforeseen circumstances, or with prior arrangement.

Labs

Lab activities will be entirely computational. If you have your own laptop, please bring it to every lab; loaner laptops are available for those who do not have their own. Lab writeups will be short (1-3 page) reports of results obtained in lab, including responses to questions presented in the lab instructions. You will work with a partner or group on lab activities, but you should be sure that you know how to complete all tasks on your own, as some may appear in problem sets and exams.

The final lab project will be conducted in teams of two, with the two lab members collaborating on all aspects of the project. There will be three phases: A proposal, preliminary results (on which you will receive feedback and suggestions for further work), and a final full lab report, in the form of a scientific paper (Introduction, Methods, Results, Discussion). The report should include clear statements of which aspects of the analysis and writing were performed by each team member.

Class Presentations

Roughly every other week, a group of students will lead a discussion of a current topic in genomics (broadly defined). The topic can be chosen from a recent scientific publication, popular science article, newspaper report, or blog post of the groups choice, which will be

assigned for the class to read. The group will give a brief (15 minute) presentation about the work and relevant background, which will be followed by a class discussion. Each student will present twice during the course of the semester.

Final Paper

Rather than a final exam, the final evaluation for the course will be a 5-7 page discussion paper on a topic of your choice. While the topic must be genomic or bioinformatic in nature, it does not need to be a topic that we have explicitly covered in class, and may range into ethical or policy questions related to advances in the field. No matter what the topic, your paper must include significant scientific content; a large portion of the grade will be determined by your ability to discuss your chosen topic in the context of relevant genomic or bioinformatic results. I will meet with each of you near the middle of the semester to discuss your choice of topic and planned approach to the paper.

Accommodations

Students who think they may need accommodations in this course due to the impact of a learning, physical, or psychological disability are encouraged to meet with me privately early in the semester to discuss their concerns. Students should also contact Stephanie Bell, Coordinator of Access Services (610-526-7351 or sbell@brynmawr.edu), as soon as possible, to verify their eligibility for reasonable academic accommodations. Early contact will help to avoid unnecessary inconvenience and delays.

Schedule (tentative)

subject to change through the semester, depending on progress and interests

Week 1	*Topics*	Why study genomes? Principles and background
Sep 4 – 6	*Lab*	Movie: *Gattaca*
Week 2	*Topics*	Sequence analysis of DNA and proteins; pairwise alignment
Sep 11 – 13	*Lab*	Online tools: websites and databases
Week 3	*Topics*	Pairwise alignment, BLAST & relatives
Sep 18 – 20	*Lab*	Identifying gene families with BLAST
Week 4	*Topics*	Multiple sequence alignment
Sep 25 – 27	*Lab*	Multiple sequence alignment & phylogenies
Week 5	*Topics*	Sequencing, building a genome, and genotyping
Oct 2 – 4	*Lab*	Getting started with Galaxy
Week 6	*Topics*	Comparative Genomics
Oct 9 – 11	*Lab*	Using the UCSC genome browser
		Exam I
	Fall Break	
Week 7	*Topics*	Genome-Wide Association Studies
Oct 23 – 25	*Lab*	Introduction to R
Week 8	*Topics*	Global analysis of gene expression
Oct 30 – Nov 1	*Lab*	RNAseq in *Streptomyces coelicolor*, intro
Week 9	*Topics*	Gene regulation & the ENCODE project
Nov 6 – Nov 8	*Lab*	RNAseq in *Streptomyces coelicolor*, begin individual projects
Week 10	*Topics*	Proteomics
Nov 13 – 15	*Lab*	Individual projects (*S. coelicolor* analysis)
Week 11	*Topics*	Systems Biology
Nov 20		
Week 12	*Topics*	Metagenomics
Nov 27 –29	*Lab*	Individual projects
		Exam II
Week 13	*Topics*	TBA (catchup)
Dec 4 – 6	*Lab*	Individual projects: final analysis, revision
Week 14	*Topics*	Final presentations
Dec 11 – Dec 13	*Lab*	No Lab, Final lab report due.