

Fall 2015 Course: 11:126:485 / 16:765:585

COURSE SYLLABUS

Location

LECTURE: Foran Hall, Room 138A, Tuesdays, 7th Period, 7:15pm - 8:35pm

JOURNAL CLUB (grad section ONLY): Lipman Hall, Room 325 or 016, Tuesdays, 6pm-7pm

LAB: Foran Hall, Computer Lab, Room 124, Thursdays, 5:35pm - 8:35pm

Lecturer:

Dr. Yana Bromberg

Lipman Hall 218

Email: yanab@rci.rutgers.edu

Office Hours: Tuesdays 5-6 or By Appointment

Lab Instructors:

Yannick Mahlich

Lipman Hall 222

ymahlich@bromberglab.org

Chengsheng Zhu

Lipman Hall 222

chengzhu@scarletmail.rutgers.edu

Office Hours: By Appointment

Brief Intro:

Bioinformatics as a field attempts to build computational models of the biological systems and mechanisms. More specifically, bioinformatics involves creating algorithms, databases, systems, and web applications to solve problems in molecular biology. Here, all computational advances are “fair game”. Bioinformatics tools use artificial intelligence, rely on “cloud” computing, and borrow concepts from signal processing and circuit theory. ALL these developments are necessary to deal with the inordinate amounts of data that is being produced by modern high-throughput experimental techniques. Due to the drop in sequencing costs we are awash in DNA, RNA, and protein sequences. Massive genomics and metagenomics efforts are opening new horizons in variation analysis. The past few years of structural genomics efforts have produced a crystal structure representative of almost every protein family. Microarray technologies allow simultaneous studies of expression of thousands of genes on a single chip. The improvements keep on coming – more information, higher resolution. Yet the unintended result of improved experimental techniques is a flood data that we have yet to make sense of. What does our genome encode? What about the soil metagenome? Can we decipher the mechanisms of disease? How are we different from other organisms? How are we different from each other? Bioinformatics attempts to answer all these questions... and give the statistical significance.

What this course IS:

This course is designed to introduce experimental biologists to bioinformatics concepts, principles, and techniques within the framework of basic shell scripting and web-based databases/tools. Students are expected to know how to work in a command-line environment and have a basic understanding of programming/scripting. The course includes a brief introduction to working with UNIX/LINUX systems, writing Python scripts, and automating/using existing applications for the analysis of large datasets. All work will be done in a live development environment; *i.e.* students will have access to the same computational resources used by dedicated bioinformatics labs on campus. By the end of this course, students will possess a sufficient bioinformatics skill set, including an informed vocabulary and knowledge of basic script development, for productive collaboration within a multi-disciplined research team.

What this course IS NOT:

This is NOT an applied methods course; rather, this class is aimed at *understanding of underlying algorithms*. We will NOT attempt to list all available tools for every project or teach you how to use them. Method selection, along with the corresponding cutoffs, thresholds, and settings, is specific to each and every research project. If you keep up with the class material you will understand the method underpinnings and be able to able to optimize your project choices on your own.

COURSE OBJECTIVES

1. Introduce students to the current bioinformatics algorithms/concepts and their implementations.
2. Introduce students to the basics of working in a Linux environment, GridEngine submissions for parallel computing, and Python scripting.
3. Teach students to cast a molecular biology problem as a bioinformatic problem, provide them with the skills necessary to independently select relevant tools, optimize their settings, and build pipelines to solve the set problem.
4. Prepare students for more advanced bioinformatics courses involving method development.
5. Teach students a sufficient bioinformatics skill set, including an informed vocabulary and knowledge of basic script development, for productive collaboration within a multi-disciplined research team.

REQUIRED TEXT

There is no required text for lab or lecture.

LECTURES AND LECTURE SLIDES

Lectures will be taught as a combination of PowerPoints with blackboard and discussions. Slides will be posted, but will contain only an outline of the work done in class. They are intended to help you reconstruct the work from class, but are not intended as a substitute for taking notes. Slides will sometimes be posted before class, but this is not guaranteed. Additionally, these are meant to be drafts, provided as a “heads up” for anyone interested

to do some online reading prior to coming to class. Students are not expected to look at them before lecture, and it is possible that corrections will be made to these slides before the final version is posted.

GRADING

Coursework will be weighted as follows:

Attendance/Class Participation	10%
Lab Homework/Quizzes	30%
Midterm	20%
Final	40%
Journal Club (graduate component only)	No Credit

Attendance / Class Participation

Regular, on-time, attendance is expected of all students. If you are going to miss a class, please inform the instructor ahead of time.

Lecture: Adequate class attendance generates 5% of the final grade. Consistent minor lateness will reduce this percentage. Since there is no textbook for this course, *attendance* is necessary for understanding of the material. Attendance will be taken in the first half of the class, so excessive lateness is equivalent to absence. Please note that you are responsible for all material covered in class whether it is present in lecture PowerPoints or not. If you intend to miss classes, find a friend who takes good notes. You may miss one lecture without an explanation. Missing two lectures without a valid (WRITTEN and DOCUMENTED explanation) will result in a loss of 10% of the final grade (attendance AND participation). Note that the entire class consists of 12 inter-dependent lectures. Missing three or more lectures (a quarter of the class) suggests that the information presented in the course was not learned. Unless prior arrangements are made with the instructor, missing more than three lectures will result in a FAILING GRADE.

Class Participation: is necessary for understanding of the material – you will get up to a maximum of 5% of the final grade for asking questions and/or participating in class discussions. Please note that if you do not participate voluntarily, you will be called upon. Class participation grade has nothing to do with being correct – it will only reflect your willingness to work towards a solution for the posed problem.

Lab: Missing any number of labs without a valid (WRITTEN and DOCUMENTED explanation) will result in a FAILING GRADE. Since time is short relative to typical lab classes, late attendance is also not acceptable. Important information and quizzes will be typically given at the beginning of class without a make-up option.

Homework / Quizzes

Completed homework assignments are due at the beginning of lab or lecture class one week from the date they are assigned, unless otherwise specified. Late submissions will NOT be accepted. Assignments containing scripts (written code) must run properly in the

standard development environment. No submission, empty submissions, or “fake” submissions (i.e. scripts that are clearly not expected to do the assignment) will receive 0% grades. Properly commented scripts that produce errors/warnings and/or fail to provide the correct, formatted output will receive no more than 50% of the grade. That being said, your programs will not be expected to handle user-input errors (unless otherwise specified) and will not be tested for such.

Quizzes will be given at the discretion of the lab/lecture instructor. Quizzes **may be** announced, but **do not have to be**. Quizzes may be written, coding, or both. They may cover lab and/or lecture material, but they will always relate to current topics. We are not looking to “burn” students with Linux questions in week 10, though you should get perfect scores if such a quiz was given. A quiz may be given at any time during any class period - immediately before or after a lecture, during a class, etc. There will be no make-up quizzes.

Midterm/Final

The **midterm** will have a written portion (taken in-class) **and** a take-home programming assignment. It will be based on material covered in lecture AND lab. This should underscore the importance of keeping up with the material. You will have one week to complete the project, including the time in regularly scheduled lab.

Late projects will NOT be accepted.

Mr. Zhu will be available during lab-time to discuss assignment problems (NOT to help you solve them).

The **final** will be a multi-tool workflow/pipeline exercise (very flexible in implementation, but necessarily well explained and documented), focusing on all techniques learned throughout class. Note that this may (but not necessarily will) have both an in-class and a take-home component. The grade for the final project will be based on three components: (1) computational implementation, (2) selection of appropriate algorithms/methods, and (3) written component describing the pipeline. The written component will be in article format; *i.e.* introduction and background, results, materials and methods, and discussion. You will have at least a week to complete the final project.

Late projects will NOT be accepted.

Both Dr. Bromberg and Mr. Zhu will be available during class-times in the first week to discuss assignment problems (NOT to help you solve them).

Journal Club (Graduate Component Only)

Graduate students in the class will be **required** to attend journal club meetings. The number of sessions will be adjusted depending on the number of people in the class. In the span of the semester you will be required to read, analyze, and present a bioinformatics paper of your choice. The presentations will not be graded, but without a presenting you will be assigned a FAILING grade in the course. Undergraduates are encouraged to attend the journal club, read the papers, and potentially present. Note, however, that this will NOT count as extra credit.

ACADEMIC HONESTY:

Academic honesty is an absolute requirement for students taking Bioinformatics. Dishonesty, in any form, will NOT be tolerated. This includes cheating on homework, quizzes, projects, as well as any form of plagiarism. Please note that working together on homework assignments and submitting identical work is NOT THE SAME; same goes for searching the web for solutions to problems, text for your project written components, and ready-made code. ***VERY IMPORTANT:*** We read papers and Wikipedia entries too and know where certain texts come from. It's easy to tell when you've copied a sentence or two. It is even easier to tell if script code was copied – please keep this in mind. ALL CHEATING WILL BE REPORTED. The Rutgers University Academic Integrity Policy can be found at:

<http://academicintegrity.rutgers.edu/integrity.shtml>.

Students contemplating cheating should consider the severe repercussions of getting caught.

Group work policy: In order to facilitate learning, students are encouraged to discuss homework problems amongst themselves. Copying a solution is not, however, the same as “discussing”. According to one colleague, Dr. Iddo Friedberg, a good rule of thumb is the “cup of coffee” rule. After discussing a problem, you should not take away any written record or notes of the discussion. Go have a cup of coffee, and read the front page of the newspaper. If you can still re-create the problem solution afterward from memory, then you have learned something, and are not simply copying.

GROUP WORK ON MIDTERM/FINAL PROJECTS IS NOT ALLOWED.

CLASSROOM RULES OF CONDUCT

1. No checking email, IM'ing, Texting, or talking to friends during class time. Phones must be off, or set to quiet. Repeat offenders will be asked to leave the classroom and noted as absent.
2. Food and beverages are not permitted in the classroom, whether opened or not.
3. Class lab time is to be spent on lab work. Lab time is not free time. If you finish early, you may start on the associated homework assignment.
4. Again, please be on time.

YOUR IDEAS, EVALUATIONS, ETC.

In general, your ideas, comments, suggestions, questions, grade challenges, etc. are welcome. Your discretion in these matters is expected, however. No part of your grade will be based on anything other than your coursework and attendance.

SUGGESTIONS FOR SUCCESS

Make sure you stay on top of your homework assignments. Waiting until the last minute to complete an assignment will not work in this course.

TENTATIVE SCHEDULE (subject to change)

1-Sep	Intro to Bioinformatics	3-Sep	No Lab: First week
8-Sep	No class: Monday Schedule	10-Sep	Intro to Linux
15-Sep	Gene Finding	17-Sep	Intro to Python
22-Sep	Pairwise sequence alignment, deriving BLOSUM	24-Sep	Python II / EMBOSS / BLAST
29-Sep	BLAST, affine gap costs, database searches	1-Oct	Python III / PSI BLAST
6-Oct	MSAs and domain families	8-Oct	AmiGO
13-Oct	Sequence signatures and motifs	15-Oct	InterPro
20-Oct	Structural Bioinformatics	22-Oct	Python IV, Chimera
27-Oct	<i>Midterm in-class and take-home portion assigned (collected Oct 30th)</i>		
3-Nov	Phylogenomics	5-Nov	MAFFT
10-Nov	Metagenomics	12-Nov	MG-RAST lab
17-Nov	Gene expression and Microarrays	19-Nov	LAB (TBA)
24-Nov	Variation and molecular level natural selection	26-Nov	Thanksgiving Break, no class
1-Dec	Disease gene prioritization	3-Dec	SNAP, Weka I
	<i>Last Lab: STRING, Weka II</i>		
	Take-home portion of final assigned (if any)	10-Dec	In-class Final (if any)
8-Dec			
15-Dec	<i>Take home portion of finals collected</i>		