

BIOL-360 / BIOL-560 Bioinformatics

Fall 2014

1 Course Information

Course	BIOL 360/560 Bioinformatics, Fall 2014
Credit hours	3
Class meetings	Mon Wed Fri, 10:00–10:50, Wheatley W01-0006
Prerequisites	Genetics, Cell Biology, Population Biology, Precalculus
Professor	Todd Riley
Office hours	Tue and Wed 11:00am - Noon and by appointment
Office	Wheatley W03-019
Phone	617.287.3236
Mailbox	In the biology office, W03-021
Blackboard Learn login	https://umb.umassonline.net
Blackboard help	Info https://login.umassonline.net/boston.cfm .
Teaching Assistant	Cory Colaneri < ccolaneri1@gmail.com >
TA Office Hours	Tue and Thur 3:30 - 4:30pm and by appointment
TA Office	Wheatley W04-048

Blackboard and email are always the best and fastest way to reach me.

There are two cases:

- For **general questions** about a lecture, homework, schedule, or final project, then it's *99% certain that other students have the exact same question*. Please **post your general questions** to our **anonymous** Blackboard class-wide discussion list so that the entire class will benefit from this exchange. To post your questions, go to Blackboard and select **Course Questions and Answers**.
- For **private discussions only**, please contact me via regular email at Todd.Riley@umb.edu.

2 Semester schedule

Important course dates:

Date	Event
September 3 (Wednesday)	BIOL 360/560 First Class
September 9 (Tuesday)	Add/Drop Ends
October 13 (Monday)	Columbus Day (Holiday)
November 11 (Tuesday)	Veterans Day (Holiday)
November 19 (Wednesday)	Pass/Fail & Course Withdrawal Deadline
November 27 to 30 (Thursday to Sunday)	Thanksgiving Recess
December 1 (Monday)	Classes Resume
December 12 (Friday)	Last BIOL 360/560 class
December 15 to 19 (Monday to Friday)	Final Exam Period (no final exam)
December 19 (before end of Friday)	BIOL 360/560 Final Project due

3 Overview and objectives

Bioinformatics is a large and rapidly growing field at the intersection of biology, computer science, mathematics, and statistics. BIOL 360/560 is a bioinformatics course specifically designed for undergraduate biology majors, although computer science, math, physics, ECOS, and other majors are certainly welcome too.

In its very short history, bioinformatics has rapidly grown as biology and biomedical research have become more quantitative with the need to analyze exponentially increasing amounts of data in order to answer important biological questions. Bioinformatics has become an essential investigative tool. It's as important as the microscope. The skills you learn in this course will be highly valuable in graduate school and beyond. These skills will greatly benefit students who

wish to pursue careers in biotechnology, clinical medicine, pure research, or in any other biology-associated field. Bioinformatics itself is a strong career path and advanced bioinformatics skills are in extremely high demand.

There are three fundamental areas that are driving the bioinformatics fields today and all three will be central to this course. First, **high-throughput DNA and RNA sequencing** have become very fast and inexpensive, and the resulting data gives researchers an immensely valuable resource. Due to the immense amount of generated sequencing data, scientists have to manage it, cross-link it, and search it with large computers and databases. Second, **molecular evolution** occurs through changes in germ-line DNA. Studying these sequence changes through time provides important clues about evolution itself, as well as insights into key biological mechanisms such as enzyme catalysis, protein folding, transcriptional regulation, and cellular signaling. Third, the field of **systems biology** is rapidly growing. Its focus is the system-wide study of organisms as complex networks of chemical interactions. Successful analysis of these large and complex networks requires massive amounts of data and computations, and cannot be analyzed by hand.

Using these three fundamental areas as our framework, we will survey several central areas of bioinformatics, both **conceptually** and **practically**. The conceptual material relates biological ideas to bioinformatics algorithms. For example, you will learn about DNA sequence evolution between individuals and between species, and learn how computer algorithms can reconstruct an evolutionary tree from those DNA sequences. At the practical level, you will learn about bioinformatics tools that are available on the web, and you'll use these tools in your homework and Final Project.

After completing this course, you will have a basic understanding of key bioinformatics concepts and practices, and you will be able to apply this knowledge to new and diverse problems. As equally important, you will have a solid foundation for learning new concepts and practices as the bioinformatics field continues to evolve and expand.

The prerequisites for this course are Genetics, Cell Biology, Population Biology, and Precalculus. If you haven't had these prerequisites, or if you're very rusty in this material, please contact me. BIOL 360/560 course does not require a background in computer science and we won't be doing any programming^{1 2}. The mathematics level will be similar to Genetics Biol-252/254, and we will cover some basic concepts of probability.

BIOL 360/560 places a **strong emphasis on collaborating with your peers** during class meetings and through the UMass Blackboard system (Section ??). You will work on group homework, and collaborate on your final projects over Blackboard. You'll find this structured collaboration to be very educational, rewarding, and great practice for your future careers.

¹The lab course associated with BIOL 360/560 does some introductory programming.

²If you have a computer science background and wish to use it on your BIOL 360/560 final project, please contact me.

4 Materials

4.1 Textbooks

The **required textbook** is *Bioinformatics and Functional Genomics*, 2nd edition by Jonathan Pevsner. Make sure to get the 2nd edition (2009); the 1st is much too obsolete. The book has an associated web site <http://www.bioinfbook.org/> with good slides and audio lectures by the author.

We'll also read several research and review papers, as well as documentation of different bioinformatics tools throughout the course.

For biological background and review information, textbooks in genetics, cell biology, biochemistry, and molecular biology are useful - particular the chapters on genomics, molecular evolution, mutations, gene regulation, and protein structure. Definitely check out the **terrific free online biology textbooks** at NCBI (<http://www.ncbi.nlm.nih.gov/books>). The NCBI Bookshelf is an amazing resource for biology students.

4.2 Course web site and computers

You will need web access throughout the entire semester for accessing Blackboard and for accessing bioinformatics tools on the web. You can use your own computers or machines in UMB labs. Macs, PCs, and Linux machines are all equally sufficient to fulfill class duties.

Note that BIOL 360/560 is using the new Blackboard Learn 9, which may be different from the Blackboard you have used in other courses in the past. Be sure to login at <https://umb.umassonline.net/>.

You will need a supported Web browser to access Blackboard, NCBI, and other web-based bioinformatics tools. Please check your browser compatibility at:

<http://www.ncbi.nlm.nih.gov/guide/browsers>

We'll start using Blackboard and iClickers on **Wednesday, Sept 10** right after the Add-Drop date, so make sure you have your iClickers ready.

4.3 Protecting your data

For your important files, *and especially for your Final Project*, you should **regularly make backup copies!!** Computers get viruses, laptops get stolen, hard drives crash, and thumb drives get lost. Losing important files is never justification for late or missing work. To perform backups people often backup files on a USB flash drive, email files to themselves, or use an online file storage and backup service. Whichever approach you take, make sure you have multiple copies of your work in multiple physical locations. **Again, perform backups often!**

5 Grading

The BIOL 360/560 grading system is designed to keep you engaged and involved throughout the entire semester, instead of cramming last-minute for a couple huge exams. If you are genuinely engaged and involved in the homework, discussions, reading, and classwork, you will receive top grades. And **you'll learn more**.

The following sections describe the components that make up the final grade.

5.1 iClicker questions & Class participation: 1/3 of final grade

To encourage attendance and lots of in-class participation, we'll heavily use the iClicker. You'll get iClicker points for:

- Participating in classroom group exercises and discussions
- Answering short questions about the reading
- Answering short questions about your homework
- Working out some more detailed quiz questions
- Answering survey questions to check whether critical concepts need further discussion

Bring your iClicker to every class and always bring spare batteries! Answers on paper will not be accepted. Read the instructions on the back of the iClicker so you'll know when the batteries need replacement.

The lowest iClicker grade will be dropped when computing your final iClicker score.

5.2 Online homework and discussion questions: 1/3 of final grade

You will work in small groups using Blackboard to discuss concepts, work through homework questions, and peer-review Final Projects. You will need to think carefully, post answers to the questions, and post follow-up messages to your peers' responses. Importantly, the questions are open-ended, just like in the real-world of bioinformatics.

Here is our system:

- Homework runs in 8-day cycles. Assignments will be posted by end of Friday, and the discussion will be closed by 9 PM the next Sunday.

- Your writing serves multiple purposes: It helps other students understand the material; it helps you reinforce your own learning; and it brings new ideas into the discussion. Therefore, **it's very important to write clearly**. Writing informally is perfectly fine. However, you won't receive any points for writing IM style like *omg*, *imo tps cya*, or for answers like *I agree with her*.
- Your discussion will benefit by including **outside materials** such as textbooks, bioinformatics web sites, and scientific journal papers. When you mention outside materials, be sure to cite them so others can follow your ideas. Informal citation format is fine here. It's also great to integrate ideas from your other courses and labs.
- When posting comments about previous messages, always be **constructive and professional**.
- You need to **post your messages within your Blackboard group**. Regular email messages won't receive any points.
- You may discuss the questions with anyone outside your Blackboard group, but you cannot take their ideas and simply reword them. If you get outside ideas from others, you must **acknowledge their contribution** in your written answer, and you must add your own new ideas to the discussion.

Here is how your homework discussions are graded:

- **“A” Discussion (90-100): Distinguished/Outstanding**

Students earning an “A” for discussion activities have participated two or more times during the week and have posted outstanding information. “A” grade postings:

- are made in time for others to read and respond
- deliver information that is full of thought, insight, and analysis
- make connections to previous or current content or to real-life situations
- contain rich and fully developed new ideas, connections, or applications.

- **“B” Discussion (80-89): Proficient**

Students earning a “B” for discussion activities have participated at least two times during the week and have posted proficient information. “B” grade postings:

- are made in time for others to read and respond
- deliver information that shows that thought, insight, and analysis have taken place
- make connections to previous or current content or to real-life situations, but the connections are not really clear or are too obvious
- contain new ideas, connections, or applications, but they may lack depth and/or detail.

- **“C” Discussion (70-79): Basic**

Students earning a “C” for discussion activities have participated at least one time during the week and have posted basic information. “C” grade postings:

- may not all be made in time for others to read and respond
- are generally competent, but the actual information they deliver seems thin and commonplace
- make limited, if any, connections which are often cast in the form of vague generalities
- contain few, if any, new ideas or applications, but instead are often a rehashing or summary of other comments.

- **“D” and “F” Discussion (10-69): Below Expectations**

Students earning a “D” and “F” for discussion activities have participated at least one time during the week and have posted information that was below expectations. “D” and “F” grade postings:

- may not all be made in time for others to read and respond
- are rudimentary and superficial; there is no evidence of insight or analysis
- are impossible to understand, or worse, rude to the other team members
- contribute no new ideas, connections, or applications
- may be completely off topic.

Here’s an example of a homework question:

Q: Your lab studies how gasoline pollutes the environment, in particular, how soil bacteria respond to these toxic hydrocarbons. You ran a traditional 2D-gel proteomic experiment measuring *E. coli* response to gasoline treatments, found a protein that differed when treated with gasoline, and performed a mass spec on the gel spot to identify the protein. Next, you used a software tool to search the mass spec data in an *E. coli* protein database, and it reported several possible matches but with extremely low confidence scores. What might be causing these low scores?

Here’s an example of a well-written reply post that would get a perfect score.

One possible problem is contamination. If that occurred, the mass spec data would have unexpected peaks, and this could confuse the database searching algorithm. The contamination could have come from other types of bacteria or fungi in the sample. The homework question doesn’t state whether the experiment was done under careful sterile conditions or with regular soil bacteria from the ground.

Another type of contamination came up in the book Bioinformatics by Baxevanis, 3rd Ed. On p. 467 they mention that contamination from trypsin fragmentation could confuse the analysis of peptide fragments from the sample. Foreign protein contamination could apply to our homework question, because the software isn't expecting mammalian trypsin fragments in a bacterial sample!

In the case of trypsin, the enzyme can cleave itself in a process called autolysis. If this occurs, the mass spec signals from those cleaved trypsin fragments could confuse the search algorithm. To get around this problem, I found that mass spec experiments can use specially-engineered trypsin that's resistant to autolysis (see <http://www.stratagene.com/products/displayProduct.aspx?pid=874>). However, it's not clear whether this engineered trypsin completely eliminates all autolysis, or just reduces autolysis. It might reduce autolysis contamination enough so the database searching software finds high confidence matches.

Alternatively, I wonder if the fragment-searching software could look for common contaminants like trypsin fragments, remove that data from the input, and then search the *E. coli* protein database? Anyone seen any references to this idea?

5.3 Final Project: 1/3 of final grade

Your Final Project will research a gene and its protein using bioinformatics tools and then write a paper on your findings. The project will start early in the semester when you find an interesting gene to research. Some of the homework and in-class questions will involve your final project.

Everyone will do their own unique Final Project, but you'll also use your Blackboard group to get several cycles of feedback as your project progresses. Your peers will make suggestions and help bring in new ideas.

Further details on the Final Project will be provided in October.

5.4 Biol-560 Additional Requirements

Biol-560 graduate students need to contact me before the end of September to discuss the additional course requirements.

6 Frequently Asked Questions

- **On group homework, if somebody in my group is late or does a bad job, will that affect my grade?**

Never. No other person's work will hurt your grade.

- **Bio-360 isn't in Blackboard!! What should I do?**

You probably logged into the old Blackboard. We're using the new Blackboard Learn 9, which has a different login page. Login to the new Blackboard at <https://umb.umassonline.net/>.

- **I'm having trouble with Blackboard!! What should I do?**

Please call 1-888-300-6920 (24 hours) for help. You can also try emailing bostonsupport@umassonline.net, but I **strongly** recommend the telephone over email. Also try accessing Blackboard from another computer.

- **I forgot my iClicker. Can I email you the answers?**

No. The best solution to avoid this problem is to ALWAYS keep your iClicker in your school bag. We will use the iClickers heavily, so you need to come up with a system that works for you.

- **My iClicker batteries just died! Can I email you the answers?**

No. Keep a spare set of batteries in your bag. We'll accommodate if you need a moment to swap batteries.

- **A bioinformatics tool seems broken. Can I have more time for my homework?**

We strongly encourage you to start homework early in the weekly cycle, so that you'll have plenty of time to get past unexpected problems. **If there's a problem, definitely post to Blackboard and do it early in the homework cycle.**

7 Academic policies

The iClicker and in-class participation form a large portion of your final grade, so you should plan to attend all class meetings. **You should arrive on time because some iClicker questions happen at the beginning of each class.** You must use only your own iClicker; answering for your absent friend is not permitted.

Please turn off your mobile phone and anything else that beeps or rings before the class starts. Texting in class distracts other students, so please don't.

Academic dishonesty is grounds for automatic failure in this course. While this course strongly encourages collaboration, **all writing must be your own.** Copying/pasting unattributed text from the web into your assignments is absolutely plagiarism, and is unacceptable in this course. BIOL 360/560 enforces the UMB plagiarism policy described at http://www.umb.edu/life_on_campus/policies/code/.

Blackboard SafeAssign automatically checks text in your homework and Final Project against the web, journal papers, and other students' writing. **Please know that if you copy material from these sources, you will probably get caught.**

8 Resources

If you have a disability and feel you will need accommodations in order to complete course requirements, please contact the Ross Center for Disability Services (Campus Center 2nd Fl., Room 2010) at 617.287.7430.