



MICROBIOLOGY 657: BIOINFORMATICS FOR MICROBIOLOGISTS
SPRING SEMESTER 2019

Course Credit: This is a 3-credit course, which meets for two 75-minute class periods each week over the spring semester. This course carries the expectation that students will work on course learning activities (readings, assignments, studying, etc) for about 2 hours out of classroom for every class period. This syllabus includes more information about meeting times and expectations for student work.

Canvas URL: <https://canvas.wisc.edu/courses/141452>

Meeting Time & Location: Tuesdays & Thursdays 1:00 - 2:15 pm, 1420 MSB

Instructional Mode: The mode of instruction for this course is face-to-face

Professor: Garret Suen, Associate Professor
5159 Microbial Sciences Building
Phone: 608.890.3971, Email: gsuen@wisc.edu
Office hours: By Appointment

Teaching Assistant: Kristopher Kieft
4545 Microbial Sciences Building
Phone: 608.265.4307, Email: kieft@wisc.edu
Office hours: By Appointment

Course Description. This course will provide a practical and fundamental introduction to sequence-based analysis focused on microbial systems. No prior knowledge of computational biology is required. Emphasis on gaining a basic understanding of the principles of both classical and newer algorithms useful for bioinformatics analysis. Students will spend a significant portion of the class on the computer learning a number of skills including, but not limited to, Unix command line, installation of bioinformatics programs (e.g. BLAST), databases, and introductory programming in Perl. Evaluation will include a midterm exam and a project geared towards solving a bioinformatics problem. Topics to be covered include: BLAST; RNA-seq analysis; transcriptional binding prediction; genome sequence assembly, analysis and annotation; and comparative genomics. Other topics may be covered, as time permits. This course requires that each student have access to a laptop that runs a Linux/Unix Operating System such as a Mac or a ChromeBook. PCs running a VM are also acceptable.

Prerequisites. This course is open to both graduate and undergraduate students with consent of the instructor. All undergraduates are expected to have taken one of MICROBIO 303, BIOCHEM 501, GENETICS 466, or GENETICS 467.

**Learning
Outcomes:**

At the conclusion of this course, both undergraduate and graduate students will gain the following skills and concepts:

- Students will gain familiarity with Unix/Linux-based platforms and how to interact with an operating system using command line
- Students will be familiar with basic sequence data formats and how to access publicly available datasets
- Students will be able to download and install sequence analysis programs locally on their own computers
- Students will conceptually understand key sequence analysis algorithms, including BLAST, genome assembly and annotation, RNA-seq analysis, and metagenomic analysis
- Students will gain familiarity with programming in the language python
- Students will be able to construct pipelines of programs that start with raw sequence data and build in more complex analysis tools.
- Students will be able to evaluate the outcomes of sequence data analysis and determine if an analysis is statistically relevant and robust.

In addition, graduate students in the course will also gain the following skills and concepts:

- Apply sequence data analysis to projects ongoing in their research labs.
- Master Unix/Linux command line tools and be able to troubleshoot problems arising from installation of tools.
- Learn to utilize advanced parameters in BLAST, genome assembly and annotation, RNA-seq analysis, and metagenomic analysis and further understand how to tailor these programs for comparative analyses.
- Code complex and advanced programs in python that not only facilitate pipeline construction, but also more in-depth parsing of tool output.
- Utilize bioinformatics tools to enable comparative genomics analysis that extend the standard use of those tools.
- Evaluate primary literature and determine what available tools are most appropriate for a specific problem in their field, as assessed by the term project.

Grading: Your grade will be based on the following elements totaling 100% based on if you are an undergraduate or graduate student, as described below.

| Element | Undergraduate | Graduate |
|------------------|----------------|-----------------|
| Midterm Exam | 20% | 30% |
| Assignments (3) | 60% (20% each) | 30% (10 % each) |
| Final Project | - | 40% |
| Final Assignment | 20% | - |
| Total | 100% | 100% |

Grading Scale:

| Grade | Undergraduate | Graduate |
|-------|---------------|-----------|
| A | 91 – 100% | 95 – 100% |
| A/B | 85 – 90% | 90 – 94% |
| B | 80 – 84% | 85 – 89% |
| B/C | 75 – 79% | 80 – 84% |
| C | 70 – 74% | 75 – 79% |
| D | 65 – 69% | 70 – 74% |
| F | < 65% | < 70% |

Textbook: There are **NO** required textbooks for this course. All material discussed in class and tested on exams will be provided to you on Canvas. Readings will include primary literature on relevant topics. A sample reading list is attached to the end of this syllabus.

Midterm Exam: The midterm exam will be administered halfway through the course and will be worth either 20% or 30% of the final grade, based on if you are an undergraduate or graduate student, respectively. The exam itself will be a practicum that will be administered in class and will cover all material up to the midterm exam date. The exam will be open book/computer.

Make up exams with NOT be given except for extenuating circumstances as described by UW-Madison policy.

Assignments: Three assignments will be provided throughout the semester and will be used to assess mastery of the material. Each assignment will be 20% or 10% of your final grade, depending on if you are an undergraduate or graduate student, respectively. For undergraduates a final assignment will also be provided at the end of the course, in order assess material covered in the second half of the class.

Final Project: A final project will be expected of each graduate student for the course, worth 40% of the final grade. The final project will be due on the last day of classes. It is expected that students will work on an existing problem that is part of their research. If no such projects exist for a given student, contingencies will be provided. The final project is designed to demonstrate mastery of the material presented in class and further apply the material to a research question relevant to the student's current lab project.

Sample Reading List:

- Altschul, S.F., Madden, T.L., Shaffer, A.A., Zhang, J., Zhang, Z., Miller, W. & D.J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25(17): 3389 – 3402.
- Baker, M. *De novo* genome assembly: what every biologist should know. *Nature Methods* 9(4): 333 – 337.
- Croucher, N.J. & N.R. Thompson. 2010. Studying bacterial transcriptomes using RNA-seq. *Current Opinion in Microbiology* 13(5): 619 – 624.
- Denisov, G., Walenz, B., Halpern, A.L., Miller, J., Axelrod, N., Levy, S. & G. Sutton. 2008. Consensus generation and variant detection by Celera Assembly. *Bioinformatics* 24(8): 1035 – 1040.
- Edwards, D.J. & K.E. Holt. Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. 2013. *Microbial Informatics and Experimentation* 3:2.
- Kunin, V., Copeland, A. Lapidus, A., Konstantinos, M. & P. Hugenholtz. 2008. A Bioinformatician's guide to Metagenomics. *Microbiology and Molecular Biology Reviews* 72(4): 557 – 578.
- Magoc, T., Wood, D. & S.L. Salzberg. 2013. EDGE-pro: estimating degree of Gene Expression in Prokaryotic Genomes. *Evolutionary Bioinformatics* 9: 127 – 136.
- Metzker, M.L. 2010. Sequencing Technologies – the next generation. *Nature Reviews Microbiology* 11(1): 31 – 46.
- Zerbino, D.R. & E. Birney. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research* 18(5): 821 – 829.

RULES, RIGHTS & RESPONSIBILITIES

See the Guide to Rules, Rights and Responsibilities for undergraduates (<http://guide.wisc.edu/undergraduate/>) or graduates (<http://guide.wisc.edu/graduate/>).

ACADEMIC INTEGRITY

By enrolling in this course, each student assumes the responsibilities of an active participant in UW-Madison's community of scholars in which everyone's academic work and behavior are held to the highest academic integrity standards. Academic misconduct compromises the integrity of the university. Cheating, fabrication, plagiarism, unauthorized collaboration, and helping others commit these acts are examples of academic misconduct, which can result in disciplinary action. This includes but is not limited to failure on the assignment/course, disciplinary probation, or suspension. Substantial or repeated cases of misconduct will be forwarded to the Office of Student Conduct & Community Standards for additional review. For more information, refer to studentconduct.wiscweb.wisc.edu/academic-integrity/.

ACCOMMODATIONS FOR STUDENTS WITH DISABILITIES

McBurney Disability Resource Center syllabus statement: “The University of Wisconsin-Madison supports the right of all enrolled students to a full and equal educational opportunity. The Americans with Disabilities Act (ADA), Wisconsin State Statute (36.12), and UW-Madison policy (Faculty Document 1071) require that students with disabilities be reasonably accommodated in instruction and campus life. Reasonable accommodations for students with disabilities is a shared faculty and student responsibility. Students are expected to inform faculty [me] of their need for instructional accommodations by the end of the third week of the semester, or as soon as possible after a disability has been incurred or recognized. Faculty [I], will work either directly with the student [you] or in coordination with the McBurney Center to identify and provide reasonable instructional accommodations. Disability information, including instructional accommodations as part of a student's educational record, is confidential and protected under FERPA.” <http://mcburney.wisc.edu/facstaffother/faculty/syllabus.php>

DIVERSITY & INCLUSION

Institutional statement on diversity: “Diversity is a source of strength, creativity, and innovation for UW-Madison. We value the contributions of each person and respect the profound ways their identity, culture, background, experience, status, abilities, and opinion enrich the university community. We commit ourselves to the pursuit of excellence in teaching, research, outreach, and diversity as inextricably linked goals.

The University of Wisconsin-Madison fulfills its public mission by creating a welcoming and inclusive community for people from every background – people who as students, faculty, and staff serve Wisconsin and the world.” <https://diversity.wisc.edu/>

Microbiology 657: Spring 2019 Schedule (Subject to change)

| <i>Date</i> | <i>Day</i> | <i>Lecture</i> | <i>Topic</i> |
|---------------|------------|----------------|----------------------------------------------------------------------------|
| Jan 22 | T | 1 | Course overview – An Introduction to Computers |
| Jan 24 | R | 2 | Introduction to the Command Line |
| Jan 29 | T | 3 | Next-generation Sequencing and File Formats |
| Jan 31 | R | 4 | Introduction to the BLAST algorithm |
| Feb 5 | T | 5 | Working with BLAST (Assignment #1 available) |
| Feb 7 | R | 6 | Introduction to Genomes and Assembly |
| Feb 12 | T | 7 | Installing the Celera Assembler I (Assignment #1 due) |
| Feb 14 | R | 8 | Installing the Celera Assembler II |
| Feb 19 | T | 9 | Installing the Celera Assembler III |
| Feb 21 | R | 10 | Working with the Celera Assembler – An Example I |
| Feb 26 | T | 11 | Working with the Celera Assembler – An Example II |
| Feb 28 | R | 12 | Installing and Working with the Velvet Assembler (Assignment #2 available) |
| Mar 5 | T | 13 | Introduction to Genome Annotation |
| Mar 7 | R | 14 | Interacting with a Genome Annotation (Assignment #2 due) |
| Mar 12 | T | 15 | Genome Annotation I |
| Mar 14 | R | 16 | Genome Annotation II |
| Mar 26 | T | - | Midterm |
| Mar 28 | R | 17 | Comparative Genomics |
| Apr 2 | T | 18 | Python I – The Basics (variables, loops, conditionals) |
| Apr 4 | R | 19 | Python II – Basic Parsing |
| Apr 9 | T | 20 | Python III – Data Structures I |
| Apr 11 | R | 21 | Python IV – File I/O and command line arguments |
| Apr 16 | T | 22 | Python V – Putting it all together (Assignment #3 available) |
| Apr 18 | R | 23 | Python VI – Data Structures II |
| Apr 23 | T | 24 | RNA-seq I (Assignment #3 due) |
| Apr 25 | R | 25 | RNA-seq II |
| Apr 30 | T | 26 | Shotgun Metagenomics I |
| May 2 | R | 27 | Shotgun Metagenomics II |