Computational Biology
Spring 2019

Class meetings: Enzi STEM Building 155
3:10–5:00 p.m., Tuesday & Thursday

Websites: www.uwyo.edu/buerkle/compbio
uwyo.instructure.com (WyoCourses)

Course numbers: Botany 4550, 5550 and Comp. Sci. 5010

Professor: Alex Buerkle
Email: buerkle@uwyo.edu
Office: Aven Nelson 202
Office Hours: Wednesday and Friday 1–2:30 p.m.
(and by appointment)  

T.A.: Jason Mercer
Email: jmercer4@uwyo.edu
Office: Agriculture 316
Office Hours: Thursdays 10–11 a.m.
(and by appointment)

Overview

Most subdisciplines in modern biology involve the analysis of large amounts of data, inference based on probabilistic models, or both. This course exists to help students gain skills in data analysis and combines elements of applied computational science (data wrangling, computer systems, etc.) with probability and statistics. These are challenging and rewarding aspects of modern research. You will be able to learn and do new useful things very quickly.

Computational biology can open a world of possibilities for research and employment. The computational and analysis skills we will practice are increasingly imperative for modern biologists (e.g., see the mention of computation, statistics and quantitative analysis in many job advertisements). Without them the scope and ambition for biological research are unnecessarily constrained. In addition, the problems and computational approaches we will study are good application domains for students from outside biology who are interested in pursuing computational science.

This course will be motivated by practical applications of probability, simple mathematics, and computational tools to biology. In each section of the course we will begin with biological questions and then investigate computational methods for graphical and statistical analysis of real data sets.

I have a few main goals in this class: 1) to discover the importance of probability, mathematics and computational methods in biology, 2) to understand philosophies and conceptual frameworks underlying commonly-used statistics and analytical methods, and 3) to become proficient with analytical and computational tools that can be applied to biological problems. I hope that by the end of the course you will have an appreciation of the diversity of applications of these analytical tools and the role they play in modern biology, and have begun to develop proficiency in these areas.

Course Materials

Please acquire a paper copy or gain access to the free e-book copy through the UW Library.

2. Additional materials will be distributed in class or linked from the course website.

3. Optional supplemental and potentially useful materials:

   a. G. Grolemund and H. Wickham. 2016. R for Data Science. (see course webpage for URL to free online version)
   b. R. Peng. 2015. R Programming for Data Science. Leanpub. (see course webpage for URL)
   c. B. M. Bolker. 2008. Ecological Models and Data in R. Princeton University Press. (we will read and use several chapters from this book; I will provide pdf copies of chapters).

ASSESSMENT AND GRADING

Assessment of your work in this course will be done regularly by you. You will do independent work (e.g., reading, exercises, problem solving) and determine through interactions with the class what you did correctly, where you need further study or have questions, and so forth. Several small assignments will be graded for completion (not accuracy) to offer an incentive to keep up with classwork. Given that this is an upper-level undergraduate or graduate course, it will be your responsibility to gauge your competency and understanding of these exercises. I will help motivate you by offering challenging and engaging course material, but your success will depend heavily on your personal motivation.

Late submission of project write-ups or the graduate assignment will be penalized, unless prior arrangements have been made with the instructor to accommodate an extraordinary circumstance. The automatic penalties are: 1) 1–2 hours after the deadline on the due date, -5%, 2) 2 < x < 24 hours after the deadline, -10%, and 3) 1 < x < 3 days after the deadline, -20%.

Likewise, any absence during a scheduled exam time must be coordinated with the instructor in advance. Small graded assignments will not be accepted late; instead the lowest two scores will be dropped. Please plan accordingly.

Your performance will be evaluated formally based on three write-ups of lab projects, two exams, and several small assignments (graded for completion). In addition, graduate students will write and present short papers on specific topics in computational biology. Grading will be on a standard scale (i.e., 90–100=A, 80–89=B, etc.).

<table>
<thead>
<tr>
<th></th>
<th>Undergraduate</th>
<th>Graduate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project papers</td>
<td>3 × 60 =</td>
<td>180</td>
</tr>
<tr>
<td>Exams</td>
<td>2 × 30 =</td>
<td>60</td>
</tr>
<tr>
<td>Small assignments</td>
<td>10 × 4 =</td>
<td>40</td>
</tr>
<tr>
<td>Graduate student project</td>
<td>—</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>280</td>
<td>310</td>
</tr>
</tbody>
</table>

2
Project papers

Three of the four sections of the course will culminate in a project paper. The task is to report findings of a computational analysis and to interpret them in the context of the biological questions that motivated the study. The format and length of these will vary slightly from topic to topic. Project reports will be written individually, but they will be based on collaboration among students and the instructor. Due dates are in the schedule of topics in this document, and will be announced in class and posted on the course’s WyoCourse website. Papers will be submitted electronically, via WyoCourses.

Short exams

Two short exams (scheduled for 2 April and 9 May) will be used to assess your understanding of major concepts. In practice people do statistics, mathematics and computation with the assistance of books, computers and colleagues. Thus, for the exams, I will ask you to address concepts and “big-picture” items, things that you would expect to be able to do without books, etc.

Questions on the exam will call for short answers of a few sentences or short essays. We will discuss sample questions before the first exam to illustrate the types of questions I will ask.

Small graded assignments

Exercises and problem sets will be associated with several subjects and will be assigned regularly. Twelve assignments will be given and graded for completion, of which ten will count toward your grade (meaning only ten need be completed for full credit and that late assignments will be accepted). These will be listed on the course’s website and will be submitted electronically through WyoCourses. You will have opportunities to ask questions about assignments during class, but you will require time outside of class to complete the exercises. A subset of these graded assignments will be done as a group and we will compare these critically as one would for a bake- or cook-off at a fair.

Note-taking and questions

You will want to take notes on reading assignments, make outlines of important concepts, and make notes on computer code that is of use to you. In class meetings I will present lectures, answer questions, offer further explanations of material and facilitate discussions. The course’s website is an additional resource for asking questions and discussing topics outside of class, and we will use it in various ways to interact outside of class.

Graduate students

Beyond the above requirements, each graduate student will write a short paper that introduces a specific topic in computational biology (the equivalent of two substantial \( \frac{11}{2} \times 11 \)
pages, plus any images, references and URLs to key resources, on an individual web page associated with the course). Each student will write on a different topic. The purpose of this assignment is for the individual student to dig deeper into a particular subject, and to share the findings with the class and enrich our literacy in computational science.

In addition to the short paper, graduate students will give short presentations on their findings to the class. We will do this as a type of electronic poster session, in which 4–6 students will present on a given day, and the remainder of the class will visit the presentations. The presentation and paper will account for 30 points and the course grades of graduate students will be calculated out of 310 possible points. The due date for the written assignment will be April 25th and students will choose a time for their electronic poster presentations in the last two weeks of classes. Below is a list of suggested topics, each of which should be considered in the context of science. Students are welcome to suggest and discuss additional topics with me:

1. Data archiving at Dryad and other public repositories
2. The National Science Foundation’s requirement for data management plans
3. Network file systems for scientific computing
4. Types of parallel computing for science
5. High performance computing in a cluster environment (e.g., the Teton computer at UW)
6. Tools and methods for “reproducible research”
7. Version control systems for documents, software development, and other collaborative work (git, Google docs, etc.)
8. Hardware and software tools for secure and reliable data sharing and storage
9. Modern database technologies
10. Programming libraries for scientific computing (HDF5, GSL, etc.)
11. A comparison of python and R
12. What is Cloud computing and what is its relevance for academic computing?

COMPUTER USE AND ETIQUETTE

We are using a room full of computers for this class. Occasionally, a student has difficulty focusing on the class material and not being distracted by the many distractions that a networked computer offers. These distractions are typically detrimental to the attention of the individual user, but also to the students who sit nearby. Therefore I ask that you find a solution to manage and minimize these distractions for yourself. For example, I ask that you do not consult the web while I or anyone else is presenting material to the class, because it will interfere with your ability to follow the presenter. Likewise, there is no reason to use personal email or other messaging software during class. Of course, legitimate uses of these resources during class exist, particularly during those times when I ask you to use the computers. These include transferring a copy of a file to yourself or another student via email, consulting the web for computational science resources related to class work, taking
notes, and trying out code as I present it on screen. Please use your best judgment and minimize distractions to yourself and others.

**ADDITIONAL ITEMS**

- The schedule of topics, assignments, and all other details in this syllabus are subject to change with fair warning, including announcements in class or via university email addresses and the course website.

- If you have a physical, learning, sensory or psychological disability and require accommodations, please let me know as soon as possible. In some cases you will want to contact and seek further assistance from the University Disability Support Services (Room 128 Knight Hall, (307) 766-3073, udss@uwyo.edu, http://www.uwyo.edu/udss).

- Students whose religious activities conflict with the class schedule should contact the instructor at the beginning of the semester to make alternative arrangements.

- The University has policies on excused absences in UW Regulation 2-108 (Student Attendance Policy). Please advise the instructors as soon as possible about absences, so that reasonable accommodations can be made. In the absence of extraordinary circumstances, students are expected to attend all class meetings.

- Cheating and other forms of academic dishonesty are listed in University Regulation 2-114. If you are found to be engaged in academic misconduct, at a minimum you will receive no credit for that exam or assignment. Repeat or serious offenders can expect more serious consequences.

- The University of Wyoming values an educational environment that is diverse, equitable, and inclusive. The diversity that students and faculty bring to class, including age, country of origin, culture, disability, economic class, ethnicity, gender identity, immigration status, linguistic, political affiliation, race, religion, sexual orientation, veteran status, worldview, and other social and cultural diversity is valued, respected, and considered a resource for learning.

- UW faculty are committed to supporting students and upholding the University’s non-discrimination policy. Under Title IX, discrimination based upon sex and gender is prohibited. If you experience an incident of sex- or gender-based discrimination, we encourage you to report it. While you may talk to a faculty member, understand that as a “Responsible Employee” of the University, the faculty member must report information you share about the incident to the university’s Title IX Coordinator (you may choose whether you or anyone involved is identified by name). If you would like to speak with someone other than a faculty member, who may be able to afford you privacy or confidentiality, there are people who can meet with you. Faculty can help direct you or you may find info about UW policy and resources at http://www.uwyo.edu/reportit

You do not have to go through the experience alone. Assistance and resources are available, and you are not required to make a formal complaint or participate in an investigation to access them.
**Overview of Topics**

1. **Project I:** Analysis of diet restriction and lifespan of laboratory mice  
   - Class meetings 2–8, 3.5 weeks  
   - Introduction to computational science  
   - Exploratory data analysis and simple tests of hypotheses  
   - Probability Theory I  
   - Introduction to R

2. **Project II:** Poisson processes and the distribution of palindromes in cytomegalovirus DNA  
   - Class meetings 9–14, 3 weeks  
   - Probability Theory II – Poisson processes and associated distributions  
   - Statistical analysis of categorical data  
   - Paper on Project II, due 15 March (Friday)

3. **Project III:** Population genetic differences between warblers  
   - Class meetings 15–22, 4 weeks  
   - Likelihood and Bayesian parameter estimation  
   - Numerical analysis  
   - Paper on Project III, due 18 April

4. **Project IV:** The temperature record and climate change in the U.S.  
   - Class meetings 21–28, 4 weeks  
   - Hierarchical Bayesian modeling  
   - Data wrangling  
   - Paper on Project IV, due by 16 May (Thursday of final exam week)