BIOL 4106: Bioinformatics lab, Spring 2015

Department of Biology, The University of Texas at Tyler

Scheduled class time: Thursday, 5-7:50 pm, BEP 139

Instructor

Dr. Kate L. Hertweck, Assistant Professor, Department of Biology

Office: BEP 103; **Phone:** 903.565.5882; **Email:** khertweck@uttyler.edu

The best way to contact me is by email. I try to respond to student emails within 12 hours during the week and 48 hours over the weekend.

Office hours: Tuesday 11 am to 12 pm, Wednesday 9 am to 11 am. If you are unable to meet during these times, please email me with alternative suggestions for times on Monday through Friday, 8 am to 6 pm.

Course description

Analysis of large biological datasets with emphasis on genomes and transcriptomes. This is a practical course in Bioinformatics which will emphasize how to use the computer as a tool for genomic data analysis.

This course is offered as a part of an undergraduate minor in Bioinformatics and Genomics, although students who have not declared the minor are welcome to take it as well.

Course pre-requisites

Pre-requisites: BIOL 1307/1106 (General Biology 1 with lab) and BIOL 1307/1107 (General Biology 2 with lab).

Co-requisite: BIOL 4306 (Bioinformatics lecture). The material covered in these two classes are complementary and will be integrated throughout the semester. However, the focus for lecture vs. lab is slightly different, so your performance for each class will be assessed independently.

Course objectives

By the end of the course, students should be able to:

- Design and implement bioinformatics pipelines to answer pre-defined questions from a variety of biological disciplines
- Validate results from bioinformatics algorithms using hypothesis testing, correcting for multiple comparisons, etc.
- Characterize the limitations of data to answer questions of interests
- Code basic scripts to accomplish the goals above
- Obtain resources to learn new languages and algorithms

Textbooks, materials, and websites

Textbooks (same as lecture)

- **OpenIntro Statistics**, available to download for free online at http://www.openintro.org/stat/ or purchase hard copy through Amazon (see link).
- **Practical Computing for Biologists**, Haddock and Dunn, ISBN 978-0-87893-391-4

Web resources

- **Blackboard:** I will use the site for this course provided by the University (https://blackboard.uttyler.edu/) to share lecture slides, PDFs of additional readings, lists of resources, etc. Your grades will also be posted here, and a number of assessments throughout the semester will utilize the tools available in the site.
- **Google:** All of the material we will cover this semester has been discussed at length on the internet. A quick search will probably yield an answer to most technical questions (from sites like StackOverflow and SeqAnswers), and searching for more information about conceptual topics will help you develop a deeper understanding of the nuances of the problems we are trying to solve. We'll talk in class about how to make your searches more efficient, as well as ways to discern between "good" and "bad" answers.
- **GitHub class repository:** We will be referencing documents, data, and code included in a repository for this class, found at https://github.com/k8hertweck/Bioinformatics.
- **GibHub Student Developer Pack:** available for free with your UT Tyler student email and ID card (https://education.github.com/pack). We will set this up together in your first lab meeting.

Other materials

- **Flashdrive**, at least 1 GB
- **Desktop computers in BEP 139:** The computers in our lab have been pre-installed with the software necessary to perform all lab assignments. You should be able to complete homework requiring this software during class. If you require additional time, please contact me to schedule access to the lab.
- **Using your personal computer (laptop):** Please see the documentation in the GitHub class repository for more information about installing the software used in class on your personal computer.

Grading

Your performance in this class will be based on rubric-based scoring of assignments submitted throughout the semester. Rubric scores on all assignments will be averaged and converted to a percentage with final grades assigned using the following scale: 100-90%: A; 89-80%: B; 79-70%: C; 69-60%: D; <59: F

Weekly assignments will occur each week in lab, and will be due the following Tuesday at 5 pm. The description for each assignment includes point allocations for rubric scoring. Ten points will be deducted for each day an assessment is late. After five days, the grade will be automatically entered as zero. Contact your instructor for a deadline extension in the event of extenuating circumstances.

A class project will be conducted in the second part of the semester. You will be assigned a unique dataset to analyze using the tools learned each week. You will submit preliminary results for grading prior to a short presentation at the end of the semester. Each submission for the class

project will be assessed the same as a regular assignment.

Professionalism

This course was designed to emulate a collaborative research environment you may encounter professionally after graduating. You are expected to attend and participate in class discussions as would suit a working environment.

Given the relevance of technology to the study of bioinformatics, you are encouraged to use laptops or tablets to take notes and share information during class. However, the use of technology should be a supplement to in-class learning and should never interfere with other students' ability to concentrate, so be judicious in your use of cell phones (i.e., texting during class is considered rude).

I encourage you to work in groups to study and discuss class materials. You assume full responsibility for all submitted assignments, however, which means that you are expected to fully and accurately document your sources for relevant code and concepts. Please see the supplemental document on Blackboard for more information about plagiarism and appropriate citations.

The University of Texas at Tyler's honor code compels us to embrace: "Honor and integrity that will not allow me to lie, cheat, or steal, nor to accept the actions of those who do." I will not tolerate violations of academic integrity; possibly penalties may include failure of this course and University disciplinary action as described by the Manual of Policies and Procedures for Student Affairs, Chapter 8: Student Conduct and Discipline (http://www.uttyler.edu/mopp/).

Schedule

Week 1: Jan 13-15 Introduction

Week 2: Jan 20-22 Data

Week 3: Jan 27-29 Pipelines (Jan 26: deadline for registration)

Week 4: Feb 3-5 Hypothesis testing I

Week 5: Feb 10-12 Hypothesis testing II

Week 6: Feb 17-19 Combining tools and data presentation

Week 7: Feb 24-26 Mid-semester evaluations, introduction to class project

Week 8: Mar 3-5 Sequence searching

SPRING BREAK

Week 9: Mar 17-19 Sequence alignment

Week 10: Mar 24-26 Genome assembly (March 23: last day to withdraw)

Week 11: Mar 31-Apr 2 Genome annotation

Week 12: Apr 7-9 Clustering and phylogenetics

Week 13: Apr 14-16 Work day: student projects

Week 14: Apr 21-22 Student project presentations

University Mandated Policies

The following policies are mandated by the University. Text describing them can be found at http://www.uttyler.edu/academicaffairs/syllabuspolicies.pdf.

- Students Rights and Responsibilities
- Grade Replacement/Forgiveness and Census Date Policies
- State-Mandated Course Drop Policy
- Disability Services
- Student Absence due to Religious Observance
- Student Absence for University-Sponsored Events and Activities
- Social Security and FERPA Statement
- Emergency Exits and Evacuation

According to the UT Tyler Handbook of Operating Procedures: "The information contained in the course syllabus, other than the grading criteria and absence and make-up policies, may be subject to change with reasonable advance notice as long as the change is without prejudice to the students."

Acknowledgements

This syllabus was based on a draft by Srini Kambhampati at UTT, and incorporated content from the bioinformatics syllabi of both Mike Barker (U Arizona) and Matt MacManes (U New Hampshire). For full version history, see https://github.com/k8hertweck/Bioinformatics.

License