

BIO 224. Genomics, Proteomics, and Bioinformatics

Dr. Clayton J. Visger

Fall, 2018

E-mail: clayton.visger@csus.edu

Twitter: [@HoityPloidy](https://twitter.com/HoityPloidy)

Web: <https://cvisger.github.io/Visger-Lab/>

Office Hours: M/F 9:00-10:30 (or by appointment)

Lecture: M/W 200-215

Class Room: SQU 118

Office: HMB 211D or SQU 20

Lab: M/W 300-415

Course Description

Examination of current approaches in structural genomics, functional genomics and proteomics, and the bioinformatics tools utilized to understand genome organization, the regulation of gene expression, gene function and the evolutionary relationships within and between genomes. Lecture two hours; laboratory 3 hours.

Prerequisites

BIO 184, BIO 222, and graduate status or instructor permission.

Software

Lab computers will be made available to you, but you may prefer to use your own machine. The majority of labs will only require that your local computer can connect to Amazon EC2. Access to a shell/bash prompt, internet, and a decent text editor will suffice for all but a few labs.

You will need to sign up for an Amazon EC2 educate account (sign up as a student for free credits). We will cover this on day 1, but it can be done here:

<https://aws.amazon.com/education/awseducate/>

- Windows
 - **Bash for windows** or a suitable Bash terminal capable of sshing to remote servers (e.g., Xterm or Putty)

- **Sublime Text 3**
- **R**
- **Integrated Genome Viewer** (1.2gb java ver. and open it and have it download the 4gb hg19 reference genome)
- OSX (Mac)
 - **Sublime Text 3**
 - **R**
 - **Integrated Genome Viewer** (1.2gb java ver. and open it and have it download the 4gb hg19)
- Unix/Linux (see me)

Recommended resources

- **Biostars** (Biostars also sells a compiled handbook of bioinformatic workflows/case studies - exceptional resource and inexpensive)
- **StackExchange**
- **github** is an amazing resource for finding snippets of code to reuse/repurpose
- When in doubt, google!

Course Objectives

By the end of the course you should:

1. Gain basic computational skills
2. Understand how to perform a wide variety of typical genomic/bioinformatic analyses
3. Develop the ability and confidence to teach yourself new software/methods/computational skills
4. Apply bioinformatic skills to analyze publicly available genomic data

Expectations

- **Attendance:** *Please attend regularly and be on time*, frequent absence and/or lateness may be reflected in a decreased participation grade.
- **Respect:** *I respect all of my students and expect the same*, please do not talk while your instructor or others are speaking.
- **Communicate:** *If you are having difficulties with the course, life, whatever*, please come see me sooner rather than later. I am here to help you, but try to get that help before the semester is ending.

- **Cheating:** *Please don't cheat*, I take cheating seriously because it is unfair to other students, disrespectful, and undercuts your own learning. Penalties range from getting a 0 on the assignment to expulsion from the course. I expect you to know what cheating is: <http://www.csus.edu/um anual/student/stu-0100.htm>
- **Patience:** *Bioinformatics is an inherently finicky endeavor*, and we **WILL** run into technical issues. Software versions will change, breaking downstream compatibilities, and code will mysteriously stop working, etc. Instead of getting frustrated when problems arise, try to think of them as opportunities to get better at troubleshooting!

Grading Policy

I reserve the right to curve the scale dependent on overall class scores at the end of the semester. Any curve will only ever make it easier to obtain a certain letter grade. The grade will count the assessments using the following proportions:

	Points
Labs and participation	40
Homework (non-project)	20
Practicals (2 @ 20 pts each)	40
Software in research description	40
Software tutorial	40
Tutorial review	20
Point total	200

Grade assignments will be on a standard 90-80-70... scale.

Late Assignments

Late assignments will be accepted for no penalty if a valid excuse is communicated to the instructor before the deadline. After the deadline, assignments will be accepted for a 50% deduction to the score up to 2 days after the deadline. After this any assignments handed in will be given 0.

Accommodations for Disabilities

If you have a disability and require accommodations, please provide disability documentation to SSWD (Lassen Hall 1008, 916-278-6955). Please discuss your accommodation needs with me after class or during my office hours early in the semester.

Software tutorial project

Bioinformatics is rapidly evolving, and my primary goal is for you to walk away with the ability to teach yourself to use whatever method your future work/research dictates. To encourage and evaluate your ability to do this, the culminating experience of the course will be to:

- Identify a research article that uses a piece of genomic/bioinformatic software (Shell, Python, or R based) that we haven't covered in the course
- Prepare a 2-3 page (single spaced) summary of the software in the context of the paper, with the goal of showing the reader why it is a useful/important analysis to learn and use. The following should be included:
 - Summarized goal of the paper
 - Sampling strategy
 - Dataset generation
 - Drawing from the developer documentation provide an in-depth description of the analysis/software, including how it works, underlying assumptions/biases about the input data, and how it was used to address the question explored in your selected research article
 - Discuss potential research questions/applications beyond the scope of your selected paper that the software/analysis can be used to address (at least in part)
 - While your selected research article will be a focal point, throughout the paper relevant works should be referenced. For example, a software which provides divergence time estimates should cite related work on estimating rates of molecular evolution. (There should be no fewer than 10 relevant journal articles referenced)
- Learn the ins and outs of the software, and use that knowledge to prepare a tutorial/lab similar to those I provide you each week, with the target audience being your fellow classmates. You should provide a step by step guide to using the piece of software with a new dataset you have found (not the same data from your selected paper). You should annotate what each step is doing, and why.
- You will test run and provide substantive feedback on two other student generated software tutorials
- Using feedback from other students you will modify/improve your tutorial

Schedule and weekly learning goals

The schedule is tentative and subject to change. Lecture / Lab

Week 01, 08/27 - 08/31:

- Monday: Course introductions / setup
- Wednesday: Unix Shell basics / Shell tutorial

Week 02, 09/03 - 09/07:

- Monday: **Labor Day** class canceled
- Wednesday: Sequencing platforms / Fasta, Fastq, Phred, and quality control

Week 03, 09/10 - 09/14:

- Monday: Databases and the SRA / Accessing Databases
- Wednesday: Alignments (classical) / Sequence Alignments

Week 04, 09/17 - 09/21: Practical 1

- Monday: Intro to BLAST / BLAST tutorial
- Wednesday: **Practical 1**

Week 05, 09/24 - 09/28:

- Monday: Phylogenetics, gene trees, and species trees / Overflow lab
- Wednesday: Gene families / Homeobox gene family

Week 06, 10/01 - 10/05:

- Monday: Pattern searching within genomes / Pattern searching cont.
- Wednesday: Regular expressions / regex

Week 07, 10/08 - 10/12:

- Monday: Short-read alignment / Visualizing genomic alignment using IGV
- Wednesday: Genomic variation, structural variation, and more / Read Mapping

Week 08, 10/15 - 10/19:

- Monday: More info on Software project / Variant calling
- Wednesday: Unix basics part 2 / Shell scripts, if/than, and loops

Week 09, 10/22 - 10/26: Read sequence assembly demystified prior to Wednesday

- Monday: Paper assembly exercise / Paper assembly exercise
- Wednesday: Genome assembly / Assembly tutorial

Week 10, 10/29 - 11/02:

- Monday: Review / Catch up day / Mole Rat papers assigned
- Wednesday: **Practical 2**

Week 11, 11/05 - 11/09:

- Monday: Naked Mole Rate Jigsaw exercise
- Wednesday: Intro to R / R tutorial

Week 12, 11/12 - 11/16: Software in research description due Monday

- Monday: Gene expression - approaches and experimental design / RNAseq prep
- Wednesday: Interpreting RNAseq results / Differential expression analysis

Week 13, 11/19 - 11/23:

- Monday: RNAseq biases, standards, and Gene Ontology / GO enrichment analysis
- Wednesday: **Thanksgiving - No Class**

Week 14, 11/26 - 11/30:

- Project time
- Project time

Week 15, 12/03 - 12/07: Turn in Software tutorial Monday

- Software workshop peer review
- Software workshop peer review

Week 16, 12/10 - 12/14: Turn in revised software tutorial by the designated Final Exam period