

# Advanced Topics in Genomic Analysis

## Bi 623 Summer 2019 Syllabus

### **General Course Information**

#### **Instructor:**

Dr. Clay Small  
csmall@uoregon.edu

#### **GE Teaching Assistant:**

#### **Class meetings**

MTWHF                      10:00 AM – 12:00 PM and 1:30 PM – 4:00 PM                      B042 PSC

### **Course description**

This is the third of three workshop-style courses in the summer term. In this course, you will learn about common problems and questions addressed in genome biology and the fundamental conceptual bases for solutions to these problems. Most importantly, you will gain practical experience applying advanced tools in pursuit of these solutions. This course assumes a base knowledge of the Unix shell, programming principles, working in shared computing environments such as Talapas, and next generation sequencing data. It is helpful if students have experience with the R statistical language, but this topic will be introduced and frequently revisited during the course. In general, the foci of Bi 623 are data visualization, and interpretation of genomic data. The emphasis is on application of existing approaches and software, specifically with respect to the types of analyses students will most likely be called upon to conduct during their fall projects, internships, and careers.

### **Course goals**

- Continue to build mastery of the command line environment using Unix/Linux
- Understand the value and power of comparative genomics
- Understand and perform phylogenetic analysis to generate species and gene trees
- Develop proficiency in the R statistical programming environment
- Understand k-mer-based genomic metrics; perform transcriptome *de novo* and evaluate its quality
- Perform differential expression analysis using RNA-seq data, in addition to downstream analyses such as Gene Set Enrichment and multivariate visualization of high-dimensional data
- Understand how bacterial communities are quantified using different data types (i.e. 16S amplicon, metagenomic, metatranscriptomic), and test for community differences using 16S data
- Demonstrate proficiency in reading, discussing, and critiquing primary research literature

### **Course Organization**

**In-class assignments:** During the course of the class, you will gain practice via in-class assignments. It is anticipated that you will be able to complete most of these

during in-class lab time, but you are free to work on them outside of class. While these assignments will not be graded per se, students will receive participation points based on their commitment to these exercises during class – **5% of grade**

**Problem sets:** Students will be assigned eight problem sets to complete during and outside of class time. – **60% of grade**

**Quizzes:** There will be four quizzes during class over material covered in lectures. – **16% of grade**

**Lab notebook:** Students should keep a digital lab notebook to be turned in at the end of the course. It should contain notes/comments on your scripting/analysis problems and solutions throughout the course. This should be readable to a peer, and not simply be a concatenation of your assignments! – **4% of grade**

**Group presentations:** Three of the Friday afternoons will be dedicated to journal club-style group presentations on assigned papers (5% of grade). For the two Fridays a student is not presenting, she/he will write an “executive summary” for one of the papers (5% of grade). Students are responsible for reading all assigned papers, and for participating in discussing the paper during the journal club (5% of grade). – **15% of grade**

**Preparing and submitting assignments:** The information required to complete all in-class assignments and problem sets will be given in class. If you find yourself Googling how to complete entire workflows or key programmatic tasks in the assignments, you are most likely going about the problem wrong and will cheat yourself out of development as an algorithmic thinker. You should be able to describe how any piece of code you turn in operates, preferably via comments in your code.

Students are encouraged to work together and share information. In particular, some students will have a higher skill level than others, and we encourage those students with more experience to help their peers. **However, no direct sharing of code is allowed – each student must write their own code.**

Assignments will be submitted on Canvas in the requested format. Be sure to include your name on the document. Be professional – appropriately name your files, make sure they are organized, and submit only the information requested. Late assignments will not be accepted.

### **Text Book**

“Bioinformatics and Functional Genomics.” 2009. Jonathan Pevsner. Wiley-Blackwell.

Selected chapters from Pevsner, along with other supplemental reading materials will be assigned throughout the course.

9	Day	Morning	Afternoon	Assigned	Due	reading
	Mon, Aug 19	Introduction to Genomic Inference Topics / Public Sequence Databases	Public Sequence Databases Continued	Assignment 1 (Data Presentations); In-class exercise 1 (database queries / retrieval)		Pevsner, chapter 2, pp 13-42
	Tue, Aug 20	Introduction to R	R Continued, Data Visualization	Assignment 2 - R Problem Set; In-class exercise 2 (R plotting practice)		
	Wed, Aug 21	Student Data Summary Presentations	Presentations, cont. and Open Lab (work on Assignment 2)		Assignment 1 due (presentations in class)	
	Thu, Aug 22	Comparative Genomics / Homology	Phylogenetic Thinking and Inference	Assignment 3 - Reciprocal Best Blast and Synteny Plots	Assignment 2 due	Koonin 2005, pp 1-32; Baum & Offner 2008, pp 1-9
9	Fri, Aug 23	Open Lab (work on Assignment 3); Quiz 1	Open Lab (work on Assignment 3)	Papers for JC8		
	Mon, Aug 26	Maximum Likelihood Phylogenetic Inference	Phylogenetic Discordance	In-class exercises 3&4 (phyML trees);		
	Tue, Aug 27	Introduction to R Markdown	Open Lab (Assignment 3; RMD exercise)		Assignment 3 due	
	Wed, Aug 28	Introduction to RNA-seq	Open Lab (Assignment 4)	Assignment 4 - RNA-seq data QC		
	Thu, Aug 29	k-mer frequency analysis / de novo Transcriptome Assembly	k-mer frequency analysis / de novo Transcriptome Assembly cont.	Assignment 5 - k-mer-based genome estimates / Transcriptome Assembly and Evaluation	Assignment 4 due	Grabherr et al. 2011 pp 1-11
10	Fri, Aug 30	Open Lab (work on Assignment 5); Quiz 2	JC 8 (GE - Max)	Papers for JC9		
	Mon, Sep 2	No class				
	Tue, Sep 3	RNA-seq Differential Expression Analysis	RNA-seq Differential Expression Analysis cont.	Assignment 6 - RNA-seq alignment and enumeration	Assignment 5 due	Mortazavi et al. 2008 pp 1-8
	Wed, Sep 4	Open Lab (Work on Assignments 6-7)	Special topic: Emily Sutton's Research and accompanying RNA-seq In-class exercise	Assignment 7 - Differential gene expression analysis		
	Thu, Sep 5	Open Lab (Work on Assignments 6-7)	Intro to linear algebra lecture (Nick Maggio)?			
11	Fri, Sep 6	Open Lab (work on Assignment 7); Quiz 3	JC 9 (GE - Max)	Papers for JC10	Assignment 6 due	
	Mon, Sep 9	Gene Set Enrichment Analysis	Open Lab (work on Assignments 7-8)	Assignment 8 - Gene Set Enrichment Analysis; and RNA-seq Data Viz		Pevsner, chapter 9 pp 354-370
	Tue, Sep 10	Technical vs. biological variation and estimating repeatability (with stickleback microbiome In-class exercise)	Stilianos Louca guest lecture on microbial evolution Open Lab (Work on Assignment 8)	In-class exercise 5 (Basic QIIME and OTU Analysis)	Assignment 7 due	
	Wed, Sep 11	Open Lab (Work on Assignment 8)	Metagenomics and 16S Sequencing	In-class exercise 6 (Estimating repeatability)		
	Thu, Sep 12	Discussion of Fall Group Projects?	Discussion of Fall Group Projects?		Assignment 8 due	
12	Fri, Sep 13	Open Lab (work on remaining assignments); Quiz 4	JC 10 (GE - Max)			