BIO/EVO/MCB 598: Computing for Research

CONTACT INFORMATION

Instructors:
Dr. Melissa Wilson Sayres, melissa.wilsonsayres@asu.edu
Office hours: Tue 10:15am-11:15 in LSC 462 and by appointment via email

Kimberly Olney, kcolney@asu.edu
Office hours: Thur 10:15am-11:15 in LSC 462 and by appointment via email

COURSE DESCRIPTION

Course Meeting Times: Tuesday & Thursday from 9:00-10:15 in Life Sciences C (LSC) 182.

Course Number: BIO/EVO 598

Catalog Numbers: 30125/30126

Course Format: The course will work mainly with command-line programs, focusing on sequence-based analysis. This will be an interactive flipped course, with required hands-on aspects both in class debugging and trouble-shooting and for out-of-class assignments where students will work on problem-solving and implementation.

Attendance: *Student attendance is expected*. Please plan your travel to campus so you can arrive on time and stay for the entire class.

Required textbook: Biostar Handbook (available at: https://www.biostarhandbook.com)

Assignments and Readings: Assignments will be based on material presented in lectures and readings that you will complete at home. ***Readings are expected to be completed by classtime on the date they are listed in the lecture schedule. All assignments are due by 11:59pm on Saturday the week they are assigned*** *(e.g. Assignments on January 8 and 11 are due by 11:59pm on Saturday January 13).* Late assignments with unexcused absences will receive half if turned in on Sunday, one quarter credit if turned in on Monday, and no credit after Monday.

Quizzes: ***Quizzes associated with course content are due by 5am the same day as the lecture readings are assigned on before lecture begins*** (e.g., the quiz associated with Lecture 2 on Jan 11 is due by 5am on January 11). The only exception is the quiz associated with the first lecture, which will be due at the same time as the second quiz.

LEARNING OUTCOMES

Upon successful completion of this course, students will:

- have an account and be able to access the high performance computing cluster at ASU
- execute basic unix commands
- describe the utility of computing to scientific research and reproducibility

- describe ontologies
- identify common data formats in life sciences research
- download publicly accessible data
- describe current sequencing technology
- conduct basic data quality control
- implement advanced command line codes (e.g., awk, one-liners)
- debug code

ASSIGNMENTS

Two weekly assignments each week are expected to take between 1-2 hours per assignment to complete and include:

- Short answer
- Read and summarize material
- Collect primary data
- Multiple choice
- Primary analysis

GRADING POLICIES AND PERCENTAGES

Grading Policy: Assignments will account for 75% of the grade. Quizzes will count for 25% of the grade.

Your lowest quiz grade and lowest assignment grade will be dropped.

Because your lowest assignment grade and your lowest quiz will be dropped, there will be no make-up assignments and no make-up quizzes.

Final grades: Grades will be assigned on a standard letter system WITHOUT plusses and minuses:

A – 90% of total possible score and above
B – 80% to 89%
C – 70% to 79%
D – 60% to 69%
E – 59% or lower

If the instructor deems it necessary, a curve will be applied based on the distribution of all student grades. The curve cannot be determined until after the final exam is graded. A student's letter grade will not go down due to the curve.

ABSENCES

- As per above, there are no replacement or make-up exams.
- Information on excused absences related to religious observances/practices that are in accordance with ACD 304–04 "Accommodations for Religious Practices."
- Information on excused absences related to university sanctioned events activities that are in accord with ACD 304–02 "Missed Classes Due to University-Sanctioned Activities."

- Make-up assignments for will be available to students who contact their TA *PRIOR* to religious observances and University-sanctioned activities as listed above. No make-up assignments will be allowed for students who contact instructors *AFTER* the religious observance or University-sanctioned activity.

TECHNOLOGY SUPPORT

Blackboard: The course will use *Blackboard* for posting course information. Blackboard works both as a course website and as an interactive forum. Log in to Blackboard regularly to view announcements for the course. The syllabus is posted under Course Information.

Laptops: Computers are required for class assignments. There are three laptops per table for in-class trouble-shooting, but due to the intensive nature of the course, as a coding course, you are heavily encouraged to bring your own laptop to each class. Computers and digital devices should not be used during class other than for course-related content. *If you are found to be using digital devices for content unrelated to class, you will be asked to leave, and be marked with an unexcused absence.*

ACADEMIC INTEGRITY

ASU Policy on Academic Dishonesty: Academic honesty is expected of all students in all examinations, papers, and laboratory work, academic transactions and records. The possible sanctions include, but are not limited to, appropriate grade penalties, course failure (indicated on the transcript as a grade of E), course failure due to academic dishonesty (indicated on the transcript as a grade of XE), loss of registration privileges, disqualification and dismissal. For more information, see http://provost.asu.edu/academicintegrity

In the 'Student Academic Integrity Policy' manual, ASU defines ***plagiarism*** [as], 'using another's words, ideas, materials or work without properly acknowledging and documenting the source. Students are responsible for knowing the rules governing the use of another person(s) work or materials and for acknowledging and documenting the source appropriately. Academic dishonesty, including inappropriate collaboration, will not be tolerated. There are severe sanctions for cheating, plagiarizing and any other form of dishonesty.

Academic Integrity and Ethics: All work presented in this class must be a student's own, unless collaboration is specifically and explicitly permitted. *If a student is found to be cheating on an exam or assignment, the minimum penalty will be to* FAIL THE ENTIRE COURSE, and most likely a grade of XE (failure due to academic dishonesty). In serious cases, the instructor can request that a student be expelled from the university.

ACCOMODATING STUDENTS WITH DISABILITIES

Students who feel they will need disability accommodations in this class but have not registered with the Disability Resource Center (DRC) should contact DRC immediately. The DRC Tempe office is located on the first floor of the Matthews Center Building. DRC staff can also be reached at: (480) 965-1234 (V) or (480) 965-9000 (TTY). For additional information, visit: www.asu.edu/studentaffairs/ed/drc.

EXPECTED CLASSROOM BEHAVIOR

Classroom behavior: Be sure to arrive on time for class. Excessive tardiness will be subject to sanctions. Under no circumstances should you allow your cell phone to ring during class. Any disruptive behavior, which includes ringing cell phones, listening to your mp3/iPod player, text messaging, constant talking, eating food noisily, reading a newspaper will not be tolerated. If you are conducting in any of this behavior, you will be asked to leave class and will receive an unexcused absence for the day.

POLICY AGAINST THREATENING BEHAVIOR

All incidents and allegations of violent or threatening conduct by an ASU student (whether on-or off campus) must be reported to the ASU Police Department (ASU PD) and the Office of the Dean of Students. If either office determines that the behavior poses or has posed a serious threat to personal safety or to the welfare of the campus, the student will not be permitted to return to campus or reside in any ASU residence hall until an appropriate threat assessment has been completed and, if necessary, conditions for return are imposed. ASU PD, the Office of the Dean of Students, and other appropriate offices will coordinate the assessment in light of the relevant circumstances. For more, see Student Services Manual SSM 104–02 "Handling Disruptive, Threatening or Violent Individuals on Campus."

The course syllabus is a "contract" between student and instructor. We will do our best to uphold our end of the deal, and we expect you to behave in a mature, responsible and ethical fashion as outlined in this document. **If you do not accept the rules for conduct in this class, please drop the course now.**

Tentative Lecture Schedule: This is the planned schedule for the semester. Specific lecture topics may be added or removed depending on how long we spend on earlier topics.

Lecture slides are available here:
https://www.biostarhandbook.com/edu/course/1/

References are approximate chapters for each set of lectures, but you should follow links that are available at each set of lecture slides for specific sections.

Please bookmark this dynamic outline: https://docs.google.com/document/d/1AAW4HPt-AyanQpT62wF9JS00WcAhs09XXoiWZxn_s4s/edit?usp=sharing

Date	Topic	References
9 January (Tue)	**Lecture 1: How is Bioinformatics practiced?** Course structure; How is bioinformatics practiced; Computer set-up	Ch1 & 2
11 January	**Lecture 2: How do I use the command line?** Unix command line use. Find help on commands. Flag System	Ch4
16 January	Trouble-shooting computer setup. Grad students set up HPC at ASU.	
18 January	**Lecture 3: How are Unix commands used when analyzing data?** Examples of processing biological data from the command line	Ch4
23 January	**Lesson 4: What do all these words mean?** Making sense of data. Sequence and gene ontologies	Ch5
25 January	**Lesson 5: How to interpret a list of genes?** Functional enrichment	Ch5
30 January	**Lecture 6: How to access published data from the command line** Reproducibility. Data repositories. Accessing data with Entrez Direct	Ch6
1 February	**Lecture 7: Data formats. Genbank, FASTA and FASTQ** Accessing and manipulating sequencing data.	Ch7 & 8
6 February	**Lecture 8: Quality control of high throughput sequencing data** Quality visualization. Improving data quality. Adapter removal.	Ch9
8 February	**Lecture 9: Advanced quality control of FASTQ data** Sequence duplication, read merging, MultiQC, error correction.	Ch9
13 February	**Lecture 10: Sequencing concepts, methods, coverage formula** Single end and paired-end sequencing, computing sequencing depth	Ch8
15 February	**Lecture 11: Scripting and Automation** Automating tasks. Make analyses reproducible. Access the Short Read Archive.	Ch15
20 February	**Lecture 12: Accessing the Short Read Archive. Advanced scripting.** Short read archive, fastq-dump, repeating commands	Ch7
22 February	**Lecture 13: Sequence Alignments** Alignment scoring, global, local alignments	Ch11
27 February	**Lecture 14: BLAST, Basic Local Alignment Search Tool** Using blast online and at the command line	Ch12
1 March	**Lecture 15: BLAST databases** Make blast databases. BLAST search tasks	Ch12
6 March	*Spring Break*	-
8 March	*Spring Break*	-

13 March	**Lecture 16: Short Read Aligners** High throughput aligners and mappers: bwa, bowtie	Ch13
15 March	**Lecture 17: Sequence Alignment Maps (SAM)** Understanding the SAM format	Ch14
20 March	**Lecture 18: Paired end reads in BAM files.** Create and filter BAM files	Ch14
22 March	**Lecture 19: Visualizing BAM alignments.** How to use IGV the Integrative Genomics Viewer	Ch16
27 March	**Lecture 20: Visualizing Large Genomic Variation** Large insertions, deletions, copy number variations	Ch16
29 March	**Lecture 21: Filtering SAM files.** Select alignments by their attributes	Ch14
3 April	**Lecture 22: Processing SAM/BAMfiles.** Picard tools. Unaligned BAM files	Ch14
5 April	**Lecture 23: Short Genomic Variations** First steps in detecting short variations.	Ch17
10 April	**Lecture 24: Let's call some SNPs** SNP calling with bcftools and freebayes	Ch17
12 April	**Lecture 25: The Variant Call Format** Understand the VCF format.	Ch17 & 18
17 April	**Lecture 26: Making sense of variants** Variant effect prediction, interval datatypes, BED, GFF	Ch18
	Lectures 27-28 are optional.	-
19 April	**Lecture 29: Introduction to RNA-Seq data analysis** RNA-seq data concepts	Ch19
24 April	**Lecture 30: RNA-Seq statistical analysis** RNAseq statistics: R and bioconductor	Ch19
30 April-7May	**Finals Week – All Assignments and Quizzes due by April 30th 11:59pm**	