AFNS 508 APPLIED BIOINFORMATICS

College of Natural and Applied Sciences
Faculty of Agricultural, Life & Environmental Sciences
Department of Agricultural, Food & Nutritional Science
University of Alberta

Instructor: **Dr. Paul Stothard**
2-31 General Services Building (GSB)
Email: stothard@ualberta.ca

Office hours: By appointment—send an e-mail to set up a time.

Lecture/Seminar: GSB 7-11 Tuesday and Thursday, 12:30 to 13:50.

Calendar description: *Introduction to databases, software tools, and analysis methods used to characterize DNA and protein sequences. Topics include information retrieval from sequence databases, protein function prediction, assessing sequence similarity, measuring gene expression, and the analysis of high-throughput sequencing data.*

Course materials: Lecture materials and associated readings will be made available through the AFNS 508 eClass site. There is no required textbook for the course.

Learning outcomes

Through successful completion of this course students will be able to:

- Apply their knowledge of gene structure and function to predict the impacts of known and hypothetical DNA changes on phenotype.
- Extract and interpret sequence information using popular web-based databases and tools from NCBI, EBI, and UCSC.
- Use command-line-interface software to perform routine bioinformatics operations and analyses.
- Use R and Bioconductor to perform statistical analyses of data produced by high-throughput instruments.
- Process short-read and long-read sequence data to identify sequence variation, assemble genomes, and measure gene expression.
- Organize bioinformatics projects, and plan and document data transformations and analyses using Markdown, R Markdown, and Jupyter Notebooks.
- Execute Snakemake and Nextflow workflows.
- Analyze data using interactive and batch jobs on a computer cluster.

Course overview

Lectures will explore a variety of topics related to bioinformatics and will cover well-established as well as emerging technologies and techniques. Assignments and tutorials will be worked on in class and will give students the opportunity to apply a variety of software tools to genomics datasets. Students will be given access to a Linux-based server with bioinformatics software pre-installed.

Topics to be covered

- Sequence data formats
- Biological databases
- Command-line software tools
- R and Bioconductor
- High-throughput sequencing methods and associated data formats
- Bacterial genome assembly and annotation
- Gene expression analysis
- Data management strategies
- Bioinformatics workflows
- High-performance computing
- Python and JupyterLab

Course grading

Type	Due date	Marks	Total
Short-answer assignments (6)			
	Sept 19	10	
	Oct 3	10	
	Oct 17	10	
	Oct 31	10	
	Nov 10	10	
	Nov 28	10	60
Bioinformatics software review and protocol assignment	Dec 15	40	40
Total			**100**

Short-answer assignments

- Students will complete short-answer questions designed to provide hands-on experience with web-based or command-line software tools.

Bioinformatics software review and protocol assignment

- Students will write a review of a bioinformatics software tool of their choosing that includes a detailed analysis of a data set (a protocol). The document should explain the overall purpose of the software and why it is useful (i.e. what kinds of research questions can it address), and then guide the reader through a sample analysis using a publicly available data set (downloaded from NCBI for example), or simulated data. The analysis steps should be described such that another person could reproduce the analysis and obtain the same results. Relevant program output in the form of text excerpts, tables, and figures should be provided. The document should assist the reader with the interpretation of the results and should highlight particularly meaningful findings. Students are encouraged to choose or create a data set that highlights the program's capabilities and that produces an easily interpreted result. Small data sets (less that 1 GB in size) that can be processed in a short period of time on a standard laptop computer are preferred. Software tool / data set combinations already covered elsewhere (for example, on the software's website or in an existing review or tutorial) should be avoided. Sample reviews and a grading rubric will be available through eClass.

Assignment submission

- Unless stated otherwise in class, assignments are due at **11 pm on the specified date**.
- Assignments are to be submitted through eClass.

- Late assignments will be downgraded 50% per day (or portion thereof).

Final grade assessment is the responsibility of the instructor. Letter grades will be assigned only to the final distribution of aggregate raw scores. There will be no predetermined "curving" to assign final grades but instead cut-offs for different grades will be based on real breakpoints in the overall distribution of raw marks within a class for the current academic year.

Access to course materials

Recorded lectures, lecture materials, tutorials, assignments, and announcements will be accessible on-line through eClass to allow for remote and asynchronous participation in the course.

Access to representative evaluative material

Students will be given access to representative evaluative materials through eClass.

Academic integrity

"The University of Alberta is committed to the highest standards of academic integrity and honesty. Students are expected to be familiar with these standards regarding academic honesty and to uphold the policies of the University in this respect. Students are particularly urged to familiarize themselves with the provisions of the Code of Student Behaviour (online at www.governance.ualberta.ca) and avoid any behavior which could potentially result in suspicions of cheating, plagiarism, misrepresentation of facts and/or participation in an offence. Academic dishonesty is a serious offence and can result in suspension or expulsion from the University." (GFC 2003)

Code of student behavior

All students at the University of Alberta are subject to the Code of Student Behaviour, as outlined at: https://www.ualberta.ca/governance/resources/policies-standards-and-codes-of-conduct/code-of-student-behaviour. Please familiarize yourself with it and ensure that you do not participate in any inappropriate behavior as defined by the Code. Key components of the code include the following statements:

30.3.2(1) No Student shall submit the words, ideas, images or data of another person as the Student's own in any academic writing, essay, thesis, project, assignment, presentation or poster in a course or program of study.

30.3.2(2) c. No Student shall represent another's substantial editorial or compositional assistance on an assignment as the Student's own work.

Additional information

Policy about course outlines can be found in Course Requirements, Evaluation Procedures and Grading of the University Calendar.

Audio or video recording, digital or otherwise, of lectures, labs, seminars or any other teaching environment by students is allowed only with the prior written consent of the instructor or as a part of an approved accommodation plan. Student or instructor content, digital or otherwise, created and/or used within the context of the course is to be used solely for personal study, and is not to be used or distributed for any other purpose without prior written consent from the content author(s).