

**Applied Bioinformatics
PATH9577B and BIOL9919B**

Course Schedule

Regularly Scheduled Class Meeting Time

January 6th – March 31st 2025
Mondays 12:30-2:30 p.m.
Room: to be provided

Class Presentations

Monday April 7th, 12:30 to 3:30 pm

Exam

Individual exams, to be scheduled April 14 to 17

Course Instructors and Coordinators

- **Dr. Vera Tai**
Department of Biology, Faculty of Science
vtai4@uwo.ca

- **Dr. Christina Castellani**
Department of Pathology and Laboratory Medicine, Schulich School of Medicine & Dentistry
christina.castellani@schulich.uwo.ca

Course Summary

Bioinformatics has become an essential skill set in modern biology and biomedical research. This is largely driven by new genetic sequencing technologies that generate gigabytes of data overnight, but other technologies such as remote sensing and image/signal processing are driving similar challenges in broader fields including ecology, pathology and neuroscience. Although commercial software feature graphical user interfaces and 'one click' analysis workflows, these are often expensive, proprietary (closed source) programs that are constrained to a narrowly defined selection of the most popular analyses. However, biology is diverse (different organisms don't play by the same rules as model species), and research is driven by innovation, customization and asking new questions. Consequently, there is a widespread demand for the ability to customize bioinformatic workflows.

The objectives of this course are to provide students with diverse backgrounds and no prior experience with programming with a basic foundation in bioinformatics. However, greater emphasis will be placed on programming and developing skills to customize analyses, than on using existing programs. This course will train students to process biological data sets by developing basic 'command-line' and script-based computing skills. Topics covered include the UNIX file system, basic UNIX commands, running programs and building pipelines, and data exploration and visualization in R.

The course is structured to resemble a standard bioinformatic workflow: data files are obtained and managed within a UNIX-like system; open source tools and a scripting language such as bash or Python is used to manipulate and clean the data; and the processed data are analyzed and visualized using a language like R.

Students will also investigate existing bioinformatics tools or programs and develop skills to understand their parameters, how they are operated, and how to interpret their output. Students will develop the skills needed to understand the purpose, the parameters, and interpret the output of bioinformatic tools for their field of study.

Course Learning Objectives

1. understand the role of bioinformatics in modern biological and biomedical research
2. develop knowledge and skills to operate bioinformatics tools and interpret their output
3. manage files with a user account in a remote UNIX-like computing environment
4. run programs on the command line with arguments, and control data streams
5. analyze genomic sequence data using various bioinformatic tools
6. understand how to reduce a complex problem down to the simplest steps
7. write concise and informative comments and documentation within and outside a script
8. import and manipulate data frames in R
9. explore data by drawing and customizing plots using the base R graphics package
10. generate visualizations using projections and dimensionality reduction methods in R

Course Materials

There is no textbook for this course. For course readings, we refer students to a small number of online resources that can be accessed at no cost, including materials that are developed and maintained by the course instructors:

- Basic UNIX commands: <https://github.com/PoonLab/courses/blob/master/PATH9577Q/Readings/basicunixcommands.md>
- The Linux Command Line by William Shotts, <https://linuxcommand.org/index.php>
- Fundamentals of Data Visualization by Claus O. Wilke, <https://serialmentor.com/dataviz/>

Detailed Course Schedule

The course schedule comprises one 2-hour session per week for 12 weeks (one term). There will be an additional class for presentations and individual exams scheduled during the exam period.

Day	Topic	Instructor
January 6	Introduction, review of syllabus. What is bioinformatics? The history and philosophy of UNIX. Connecting to a remote system. UNIX file system structure and navigation. Absolute and relative paths.	VT

January 13	Managing files (cp, mv, rm) and directories (mkdir, rmdir). User permissions. Editing files with nano.	VT
January 20	Running programs on the command line. Redirecting data streams. Working with plain text files and understanding data formats.	VT
January 27	Running standalone BLAST. Bash scripting.	VT
February 3	Installing open source software or downloading data.	VT
February 10	Example workflow - Next generation sequence (NGS) analysis. Reference mapping. Variant calling.	VT
February 17	<i>Spring Reading Week</i>	
February 24	Introduction to R	CC
March 3	Data visualization in R	CC
March 10	Data exploration and quality control	CC
March 17	Linear models and linear mixed effects models	CC
March 24	Working with high dimensionality data	CC
March 31	Summary statistics and extracting biological information from gene lists	CC
April 7	Presentation Day	CC/VT
April 14-17	<i>Oral examinations</i>	

Methods of Evaluation

- **Paper review (10%).** Find an exemplary paper from your research field of interest with a significant bioinformatics workflow and write a short (1-2 page) description of the role that bioinformatics played in that study.
 - **Due February 24th 2025**
- **Assignments (50%).** Students will be expected to complete homework assignments using the methods presented in class to carry out bioinformatic data processing and analysis.
- **Oral presentations (15%).** Based on their paper review assignments, students will choose or will be assigned a command-line bioinformatic program or tool to present to the class. Presentations should be 10 minutes and include a description of the tool's purpose, similar tools and what features distinguish your tool, inputs, sample code, an explanation of parameter settings, outputs, and interpretation of outputs.
 - Presentations on **April 7th 2025, 12:30-3:30 pm**
 - ***please note the additional hour of class time on this day**
- **Oral examination (25%).** Students will be evaluated on an individual (one-on-one) basis by one of the course instructors on methods covered in the class.

Statement on Academic Offences

Scholastic offences are taken seriously and students are directed to read the appropriate policy, specifically, the definition of what constitutes a Scholastic Offence, at the following Web site:
http://www.uwo.ca/univsec/pdf/academic_policies/appeals/scholastic_discipline_grad.pdf

All required papers may be subject to submission for textual similarity review to the commercial plagiarism-detection software under license to the University for the detection of plagiarism. All papers submitted for such checking will be included as source documents in the reference database for the purpose of detecting plagiarism of papers subsequently submitted to the system. Use of the service is subject to the licensing agreement, currently between The University of Western Ontario and Turnitin.com (<http://www.turnitin.com>).