BIOL 432

Introduction to Computation and Big Data in Biology

Instructor: Dr. Robert Colautti, 4325 BioSciences, robert.colautti@queensu.ca

Lectures: Mondays 2:30-5:30pm, MACCOR A309; See OnQ for more detail
Tutorial/Help: Tuesdays 11:30-12:30pm, MACCOR A309; See OnQ for more detail

Prerequisites: **NONE**
Co-requisite: **NONE**
Strongly recommended: BIOL 243 & BIOL 343
Optional: CISC 151, CISC 251, BIOL 331

Textbook: **None Required**. *Suggested:*
R for Data Science by Hadley Wickham
Practical Computing for Biologists by Haddock & Dunn (Oxford University Press)
Bioinformatics Data Skills by Buffalo (O'Reilly)

Assessments:

40%	Weekly Assignments & Pop Quizzes
10%	Final Project – Proposal (group)
20%	Final Project – Report (group)
10%	Final Project – Code (group)
20%	Participation and peer evaluation

Course Learning Outcomes

Students completing this course shall be able to:

1. Design and implement a strategy for project management in biological research, based on the philosophy that scientific research should be **OPEN** and **REPRODUCIBLE**
2. Use regular expressions to modify biological data files (e.g. automated error correction, file conversion, and data extraction)
3. Create publication-ready visualizations for succinct interpretation of biological data
4. Write custom scripts to curate, merge, subset, reformat, and parse large biological datasets
5. Write original scripts for data analysis pipelines using high-performance computers, particularly those maintained by Queen's **Centre for Advanced Computing**
6. Analyze and interpret 'big data' formats in biology (e.g. CSV, FASTA, FASTQ, SAM, BED, BAM, KML, XML, BMP, PNG, SVG, SHP) to address biological hypotheses.

Course content

This course introduces students to common practices in computer programming for the analysis of biological data, by working through specific problems. The first four weeks offer a 'crash course' review of fundamentals in programming, with an emphasis on the R coding environment. Applying this content, the remaining 8 weeks work through applied tutorials for DNA/RNA analysis, phylogenetics, metabarcoding, genome assembly and transcriptomics. A final group project provides an opportunity for students to apply their coding skills to address one or more biological questions of their choosing.

Philosophy

This course applies computational skills to a range of research topics in the biological sciences. The best way to learn coding is with extensive practice, trial and failure, using real-world examples. The philosophy of this course is that you won't learn to code by reading and memorization – only through extensive application and practice. For this reason, there are no major exams; instead, you are assessed through

regular assignments and quizzes designed to encourage skill development and problem-solving. In addition, you are **STRONGLY ENCOURAGED** to find opportunities to practice coding wherever possible. For example, write code instead of messing around with spreadsheet program like Microsoft Excel, use R Markdown instead of Microsoft Word to render final reports for your other classes. It will take longer at first, but it will save a lot of time in the long run. Finally, be prepared to get frustrated – you will make many errors and most of your coding time will be debugging and searching for answers on the internet. It is important to know that **this is COMPLETELY NORMAL** and self-directed research to solve coding problems is perhaps the most crucial skill you will learn in this course.

Assistance

This course has no lectures in the traditional sense. Weekly 'lectures' offer short introductory lectures followed by independent study and hands-on tutorial sessions where you interact with your group and your TA. Outside of synchronous meetings (i.e. lecture and help sessions), a very common and useful approach to solving errors or other problems is to search Google or Stack Overflow. Often, simply copying and pasting an error into an online search will produce a helpful link. Very often you can just type 'How do I X in R' (or the R package name like ggplot, dplyr) into Google and look for links to similar questions answered on the Stack Overflow website.

Grade Assessment

Weekly Assignments:
Individual assignments account for a relatively small portion of your final mark, so don't worry if you don't do well on a few of them. They are designed to motivate you to practice coding and to help you memorize and understand key concepts that will be tested through in-class pop quizzes. For this reason, it is important that you begin the assignment as soon as possible and complete it on your own. Assignments due within 2 business days of posting. **A late assignment receives 0%.** However, students who require accommodation for any reason will receive a 5-day grace period without penalty, **as long as official requests are submitted to the appropriate website** (see "Accommodations" and "Extenuating Circumstances", below).

Tests:
There are no exams in this course. Instead, you will be tested on your understanding of the course content through OnQ quizzes available in-class. The purpose of the quizzes is to reinforce learning concepts from tutorials and assignments that you work through during help sessions. **There are no make-up quizzes**, however students requiring accommodation for any reason may miss up to two quizzes, or drop their two lowest quiz marks without penalty, **as long as official requests are submitted to the appropriate website** (see "Accommodations" and "Extenuating Circumstances", below).

Final Project:
A large part of your final grade depends on a group project in which you work collaboratively and apply methods learned in this course to write original code for managing and interpreting a large dataset. The project is due the last week of classes. **A late project receives 0%.** However, groups with students requiring accommodation for any reason receive a 2-week grace period without penalty.

Participation:
Part of your final grade will be based on participation marks provided to students who contribute to discussion in class in a collegial and respectful manner. Each member of your group will also provide a peer assessment grade based on their opinion of your contributions to the group according to the following statement: "This group member was respectful, helpful, and made important contributions to the group assignments without dominating the discussion."

Other Considerations

Grievances:
It is the student's responsibility to raise any concerns about grading errors within 5 business days of receiving a grade on any quiz or assignment. Requests must first be made in writing via email to the TA,

and must include a very clear explanation of the reason for concern. If the situation is not resolved, a follow-up request can be made to the course coordinator to re-grade the assignment, which may result in a lower or higher grade. Other grade adjustments not will NOT BE CONSIDERED without prior approval of accommodations or extenuating circumstances, as follows:

Accommodations:
Queen's University is committed to achieving full accessibility for people with disabilities. Part of this commitment includes arranging academic accommodations for students with disabilities to ensure they have an equitable opportunity to participate in all of their academic activities. If you are a student with a disability and think you may need accommodations, you are strongly encouraged to contact the Queen's Student Accessibility Services (QSAS) and register as early as possible. For more information, including deadlines and application forms, please visit the QSAS website:
http://www.queensu.ca/studentwellness/accessibilityservices/

1. Submit a request
2. Email the course co-ordinator and TAs when a request has been submitted

Extenuating circumstances:
For excused absences due to short-term extenuating circumstances (e.g. flu, injury, short personal leave), please submit a formal Academic Consideration Request Portal (ACRP) on the Arts & Sciences accommodations website: https://www.queensu.ca/artsci/accommodations

1. Submit a request
2. Email the course co-ordinator and TAs when a request has been submitted

Academic Integrity (http://www.academicintegrity.org)**:**
Queen's students, faculty, administrators and staff all have responsibilities for supporting and upholding the fundamental values of academic integrity. Academic integrity is constituted by the five core fundamental values of honesty, trust, fairness, respect and responsibility and by the quality of courage. These values and qualities are central to the building, nurturing and sustaining of an academic community in which all members of the community will thrive. Adherence to the values expressed through academic integrity forms a foundation for the "freedom of inquiry and exchange of ideas" essential to the intellectual life of the University. Students are responsible for familiarizing themselves with and adhering to the regulations concerning academic integrity. General information on academic integrity is available at **Academic Integrity @ Queen's University**, along with Faculty or School specific information. Departures from academic integrity include, but are not limited to, plagiarism, use of unauthorized materials, facilitation, forgery and falsification. Actions which contravene the regulation on academic integrity carry sanctions that can range from a warning, to loss of grades on an assignment, to failure of a course, to requirement to withdraw from the university

Turnitin:
Queen's University has partnered with the third-party application Turnitin to help maintain our standards of excellence in academic integrity. Turnitin is a suite of tools that provide instructors with information about the authenticity of submitted work and facilitates the process of grading. Submitted files are compared against an extensive database of content, and Turnitin produces a similarity report and a similarity score for each assignment. A similarity score is the percentage of a document that is similar to content held within the database. Turnitin does not determine if an instance of plagiarism has occurred. Instead, it gives instructors the information they need to determine the authenticity of work as a part of a larger process.
Terms: http://turnitin.com/en_us/about-us/privacy.
Privacy Policy: http://turnitin.com/en_us/about-us/privacy#policy

Copyright:
The material in this course is the property of the instructor unless otherwise stated. Third party copyrighted materials (such as book chapters and articles) have either been licensed for use in this course or fall under an exception or limitation in Canadian Copyright law. Copying this material for distribution (e.g. uploading material to a commercial third-party website) can lead to a violation of Copyright law. More information: http://library.queensu.ca/copyright/students