



BINF*6210 Software Tools for Biological Data Analysis and Organization

Fall 2020
Section(s): C01

College of Biological Science
Credit Weight: 0.50
Version 1.00 - September 09, 2020

1 Course Details

1.1 Calendar Description

This course will familiarize students with tools for the computational acquisition and analysis of molecular biological data. Key software for gene expression analyses, biological sequence analysis, and data acquisition and management will be presented. Laboratory exercises will guide students through application of relevant tools.

Restrictions: Restricted to students in Bioinformatics programs.

1.2 Course Description

Welcome Message: Welcome to BINF*6210! We look forward to working with you this semester. I love to teach this course, as students make such a large leap forward in just one semester in your ability to program and to conduct bioinformatics analysis on real biological data. Although the format is new this year due to the public health circumstances, we look forward to using technology to enable us to maintain the interactive structure of this course. Our students regularly "do things" rather than only listen. The best way to learn how to do bioinformatics analyses is to do bioinformatics analyses!

Overview: The main goal of this course is to guide graduate students through an introduction to the analysis of biological data using computational and statistical tools, with emphasis upon the analysis of molecular biology data. The course will largely focus upon developing programming skills in the R language for quality checking, analyzing, and visualizing data. The course also includes an introduction to several key web-based tools, SQL, and command-line software tools for phylogenetics. We also will cover how to acquire and analyze data from selected biological databases important for bioinformatics, including sequence databases, biodiversity databases, and functional gene annotation resources. It is important to recognize the origins and limitations of these data in addition to their utility. We will discuss core bioinformatics algorithms (e.g. for alignment, clustering, phylogenetics) and population

genetics metrics and principles that are important for making analytical decisions and interpreting results. We also will promote good practices for organizing your data and analyses, prepare reproducible analyses and well-commented code, and introduce software tools that facilitate version control and collaborative coding. As bioinformatics is a fast-moving discipline, we will also spend time practising strategies for how to learn to use new tools and to conduct new analyses.

Curriculum Note: This course is complementary to others in the bioinformatics graduate program. In the fall semester of 2020, programming in the Unix environment and in the Python language are covered in Bioinformatics Programming (BINF*6410). Students in the Master of Bioinformatics program must also take Topics in Bioinformatics (BINF*6890), which covers diverse concepts in bioinformatics and emphasizes critical thinking and communication skills. The winter semester core bioinformatics courses are Genomic Methods (BINF*6110), in which large-scale genomic analysis is covered, and Statistical Bioinformatics (BINF*6970). Students from other graduate programs may wish to discuss their background and the suitability of these courses with the instructors prior to enrolling.

Pre-Requisites: Students accepted into the Master of Bioinformatics and MSc in Bioinformatics programs should have the necessary background for this course. No prior programming experience is assumed, but students are expected to have taken at least one course at the undergraduate level in genetics or molecular biology as well as statistics (or have the equivalent experience).

1.3 Timetable

Where: Instruction will be in remote format, with access instructions provided through the CourseLink.

When: Tuesdays and Thursdays 11:30 AM - 12:50 PM Eastern time. September 10 - December 3, 2020.

(Note: There is no class on Tuesday October 13 for Fall Study Break.)

1.4 Final Exam

There is no final exam for this course.

2 Instructional Support

2.1 Instructional Support Team

Instructor:	Dr. Sarah Adamowicz Associate Professor, Department of Integrative Biology
Email:	sadamowi@uoguelph.ca
Telephone:	+1-519-824-4120 x53055
Office:	SSC 2447
Office Hours:	Office Hour: Tuesdays 1:00-2:00 PM from September 15 -

December 1, 2020 (except October 13).

For the benefit of all class members, I encourage students to post general questions about course content and the assignments to the Discussion board through CourseLink.

I will remain available remotely after class on Tuesdays for individual questions. Students requiring a more in-depth individual discussion should email Dr. Adamowicz for an appointment.

2.2 Teaching Assistant

Students will also benefit from interaction with a Teaching Assistant (to be determined). Students should contact the TA for help with trouble shooting software or code.

3 Learning Resources

3.1 Required Resources

R and RStudio (Software)

Prior to the first class, please install R on your computer:
<https://www.r-project.org/>

Prior to the second class, please install RStudio:
<https://www.rstudio.com/>

Announcements will be made throughout the semester regarding R packages or additional software to install prior to the next class.

Papers and Textbook Chapters (Readings)

Relevant published articles related to the course content for each day will be posted through CourseLink. The first-listed article for each class is **required reading for that class**. The other posted readings are recommended or supplemental for students interested in more depth on that topic.

We will also be consulting a wide range of online resources, such as software manuals and vignettes for Bioconductor packages. Links to relevant resources will be posted in the class slides and in the comments sections of the example code.

Additionally, selected chapters from the following manuals and books will be recommended to accompany specific modules. All are available as open-access PDFs directly online or are available as a PDF book for download through the University of Guelph library site (<https://www.lib.uoguelph.ca/>).

1. Paradis E, 2005. R For Beginners. (Freely available through the following link: https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf)
 2. Wickham H & Golemund G, 2017. R for Data Science. O'Reilly. (Freely available through: <http://r4ds.had.co.nz/>)
 3. Xia X, 2018. Bioinformatics and the Cell: Modern Computational Approaches in Genomics, Proteomics and Transcriptomics. Second Edition. Springer. (Available through the library)
-

4 Learning Outcomes

4.1 Course Learning Outcomes

By the end of this course, you should be able to:

1. • obtain data from key databases relevant for bioinformatics and to understand the sources and limitations of these data.
 2. • filter, manipulate, analyze, and visualize bioinformatics data, with emphasis on the R programming language and software resources available through Bioconductor.
 3. • conduct reproducible analyses and use software tools for version control and collaboration.
 4. • understand and apply selected algorithms commonly used in bioinformatics, including for sequence alignment and clustering.
 5. • adapt the above skills to learn new tools and conduct new analyses not explicitly covered in class.
-

5 Teaching and Learning Activities

5.1 Lecture

Thu, Sep 10 - Thu, Dec 3

Topics: This course consists of both asynchronous components

(which you complete at your own pace in advance of class) and synchronous activities (i.e. conducted during our scheduled class time).

Each week, course materials will be uploaded to CourseLink for students to complete in advance of class. This will typically include a pre-recorded lecture or tutorial. Many weeks, there will also be one or two commented R scripts provided for students to go through at your own pace. There is also a required reading associated with each class.

5.2 Lab

Thu, Sep 10 - Thu, Dec 3

Topics:

The course also involves synchronous activities performed during class time (11:30 AM - 12:50 PM Tuesdays and Thursdays from Sept 10 - Dec 3; no class Oct 13). Our "computer lab" activities will take place remotely, enabled by technology. You will need a laptop and internet connection.

During class time, we will focus upon interactive learning activities, including critical thinking exercises, solving coding challenges, and sometimes even games. During many of the classes, we will use technology to form small "break-out groups" to enable small groups to work together for selected active learning exercises, with the Instructor and Teaching Assistant rotating among groups. We would then come together again as a complete class to discuss the exercise and to address questions from class members.

6 Assessments

Detailed instructions for each assignment will be posted to CourseLink at least two weeks before the due date and at least three weeks in the case of the Final Project. An overview is provided below.

Please note that the TurnItIn tool may be used to assess the originality of your work in comparison to that of your peers and to internet sources. If a high match to online sources is detected, please note that we would check to see where the matches are. We expect exact matches to assignment questions, to the references (e.g. journal article titles), and phrases that should be used exactly as in sources (e.g. long molecule names). Otherwise, be sure to phrase your work in your own words, and be sure to give credit to others for ideas from the

literature as well as to any online sources consulted for coding help.

Missed quizzes will receive a grade of 0, and your best 10 of 12 quizzes will be used for your quiz grade. For the 5 assignments, late work will receive a deduction of 5% for the first hour and otherwise a 10% deduction per 24-hour period.

In this course, assignments are used not only for assessment but also as learning tools. You should work on your assignments regularly. Do not leave these assignments to the night before the due date! We hope that you enjoy working on a variety of small projects throughout the course and expressing your creativity.

6.1 Assessment Details

Weekly Online Quizzes (20%)

Throughout the semester, there will be a weekly online quiz, available through CourseLink. Quizzes will cover topics such as key terms, concepts, and code syntax. There will be 12 quizzes in total, and your best 10 will be used to calculate your quiz grade (2% valuation each). While each quiz bears modest weighting, we encourage all class members to keep up with the quizzes as they will add up to 20% of your total course grade, and grades in the project assignments are typically lower. Moreover, making a consistent weekly effort will help you to improve your knowledge and skills consistently throughout the semester. You should view the pre-recorded course materials and read the reading prior to attempting each quiz. While completing the quiz, you may consult course materials as well as online sources. You may complete each weekly quiz at your own pace, any time before the due date (midnight Sundays). Quizzes not completed by the due date will receive a grade of 0 but will remain available for viewing.

Quizzes should be completed before midnight each Sunday, preferably earlier.

Assignment #1 (15%)

Date: Fri, Oct 2, 5:00 PM

For assignments #1 and #2, you will apply your knowledge to solve new problems. You will design and complete a mini-project that builds upon the skills and concepts covered until that point in the course. Example mini-projects will be provided.

Code needs to be correct, do what it is meant to do (always check!), be well-commented, and reproducible. In your commenting, you should focus on being precise in your explanations of algorithms and functions. The assignment will include an introductory paragraph and a short written summary at the end interpreting your findings. You will additionally be assessed on the creativity and novelty of your mini-project in terms of going beyond the class materials.

Assignment #1 is due by 5:00 PM on Friday Oct. 2nd.

Assignment #2 (15%)

Date: Fri, Oct 23, 5:00 PM

See above for the description of Assignment #1.

Additionally, as the course progresses, you should aim to write code that is streamlined as well as computationally efficient. Examples would include using vectorized functions in R rather than repeating similar lines of code. You should also pay careful attention to the preparation of your visualizations, considering whether the main message is conveyed clearly, ensuring that you have used informative labeling, checking that your colour and symbol choices are clear and accessible, etc. The quality and sophistication of your work will improve over the course of the semester.

Assignment #2 is due by 5:00 PM on Friday October 23rd.

Assignment #3 (Group Project) (15%)

Date: Fri, Nov 13, 5:00 PM

For Assignment #3 (Group Project), you will swap code (from either Assignment #1 or Assignment #2) with a peer in your group. The assignment involves making improvements to your peer's code and using GitHub to manage the collaboration and code edits. You should discuss the project together and may work on the code together. Each person will individually prepare a short-write up about the code improvements and collaboration process, which is individually graded.

(Why GitHub? GitHub is an important code repository as well as a tool for version control and collaboration. By the end of your program, we would highly recommended that you post examples of your work to GitHub and provide a link to your GitHub page on your CV when applying for bioinformatics-related jobs.)

Assignment #3 is due by 5:00 PM on Friday November 13th.

Assignment #4 (Seminars) (5%)

Date: Fri, Dec 4, 5:00 PM

For students registered in the Software Tools class, attendance at the Bioinformatics Seminar Series is mandatory. Seminars will be held in virtual format for the fall 2020 semester. Attendance in real time is preferred to the maximum degree possible, to enable audience members to ask questions and to interact with members of the community afterwards. However, the seminars will be recorded for those who need to watch at a different time due to illness or personal circumstances. The seminar series will help you to expand your knowledge of the field of bioinformatics as well as increase your exposure to the diversity of careers possible.

Students should attend all seminars of the F20 semester; there are typically 3-4 seminars per semester. You will then choose any two for this short writing assignment about the seminars.

Assignment #4 is due by 5:00 PM on Friday December 4th.

Assignment #5 (Final Project) (30%)

Date: Fri, Dec 11, 5:00 PM

Assignment #5 involves completing a final course project consisting of written paragraphs

interspersed with commented code blocks and visualizations. Your project should include: introduction, description of dataset, data exploration and quality control (commented code block and visualizations), description of main software tool used, main analyses (commented code block and visualizations), interpretation of results and discussion.

You will choose your own topic, but you should clear your topic selection with the instructor at least three weeks before the due date. Several example topics will be provided. Your project must incorporate at least one software tool beyond those covered in class. Being able to read software documentation and do new analyses of interest to you is important in bioinformatics.

Assignment #5 is due by 5:00 PM on Friday December 11th.

7 Course Statements

7.1 Class Attendance

Class attendance is considered mandatory for Software Tools. While pre-recorded lectures and tutorials will be provided, participation in the synchronous learning activities during class time will help you to maximize your success in the course and beyond.

We will strive to record class materials to the maximum degree possible to assist any students who need to miss a class (e.g. due to illness, family circumstances, or other compassionate reasons). Instructor-led course presentations will be recorded. However, please note that small break-out groups of students will not be recorded. Therefore, if you need to miss a class, please go through the recorded components and also work through the commented example scripts and coding challenges posted to CourseLink. Also, review the example answers, once posted.

Throughout the semester, you should regularly consult CourseLink for announcements and posted course materials.

7.2 Group Activities

Throughout the course, we will engage in regular discussions and coding activities in pairs or small groups during class time. Pedagogical research indicates that you will learn better if you regularly work in groups and engage in active learning activities. So, please come to the virtual class prepared to engage with your peers.

We will change up the groups from time to time so that you can meet new people and work with individuals with varying personalities and academic backgrounds. Collaboration is common in bioinformatics in the workplace as well, and so this is good practice for your career beyond graduate studies.

I also encourage students to form peer study groups to review course materials outside class time and to engage in other activities beyond the course materials, such as analyzing additional datasets to develop your skills further. You may also work together to solve the "coding challenges" posed to you. Taking the time to work through problems of increasing difficulty will help you to improve.

For graded assignments, it is important to complete your work on your own. The exception is the single group assignment. For that assignment, you should discuss your assignment and can work together on the code. Each person submits an individual short write-up, which is graded individually. You need to prepare your own written remarks.

7.3 Course-Specific Statement on Academic Integrity

You are encouraged to work in peer groups to practise your coding skills, to discuss concepts, and to seek advice about useful software and information resources. However, you must complete your individual assignments independently. Electronic resources (such as TurnItIn) may be used to assess the originality of your assignments. Use quotations sparingly, such as for profound statements or definitions. Otherwise, you should paraphrase from any sources you cite for the written portions of your assignments. You are free to consult online resources to learn about various ways of coding and approaching bioinformatics problems. If you draw heavily from a specific source (such as a particular entry on Stack Overflow) when completing an assignment, then you should cite that source and indicate how you adapted the code for your purposes. You always need to check that your code works as intended.

You will work together in a small group for one group assignment (assignment #3). You should discuss your assignment and may work on the code together. You should complete the short write-up for that assignment on your own. The assignment is graded individually.

Please see below for the university-level statement on academic integrity for further information.

7.4 Due Dates

Assignments are due by the due **DATE AND TIME** listed for each assignment.

For the 5 assignments, late work that is submitted within one hour (exactly 60 minutes) of the deadline will receive a penalty of 5%. Otherwise, late work will receive a deduction of 10% per 24-hour period.

Among the 5 assignments, one time during the semester you may use a 48-hour grace period. Each student is permitted to submit one assignment within 48 hours of the due date without academic penalty. A reason does not need to be provided. We suggest to save this opportunity for when it is really needed!

Quizzes not completed by the due date receive a grade of zero. Your two lowest marks for the semester will be dropped from your quiz grade calculation. This would include cases of quizzes not attempted by the due date.

See below for the university guidelines regarding cases of missed work due to illness or compassionate reasons.

Assignments should be submitted to the course Dropbox available through CourseLink.

8 College of Biological Science Statements

8.1 Wellness

If you are struggling with personal or health issues:

- Counselling Services offers individualized appointments to help students work through personal struggles that may be impacting their academic performance.
- Student Health Services is located on campus and is available to provide medical attention.
- For support related to stress and anxiety, besides Health Services and Counselling Services, Kathy Somers runs training workshops and one-on-one sessions related to stress management and high performance situations.
<http://www.selfregulationskills.ca/>

8.2 Personal information

Personal information is collected under the authority of the University of Guelph Act (1964), and in accordance with Ontario's Freedom of Information and Protection of Privacy Act (FIPPA) <http://www.e-laws.gov.on.ca/index.html>. This information is used by University officials in order to carry out their authorized academic and administrative responsibilities and also to establish a relationship for alumni and development purposes.

For more information regarding the Collection, Use and Disclosure of Personal Information policies please see the Undergraduate Calendar.

(<https://www.uoguelph.ca/registrar/calendars/undergraduate/current/intro/index.shtml>)

9 University Statements

9.1 Email Communication

As per university regulations, all students are required to check their e-mail account regularly: e-mail is the official route of communication between the University and its students.

9.2 When You Cannot Meet a Course Requirement

When you find yourself unable to meet an in-course requirement because of illness or compassionate reasons please advise the course instructor (or designated person, such as a teaching assistant) in writing, with your name, id#, and e-mail contact. The grounds for Academic Consideration are detailed in the Undergraduate and Graduate Calendars.

Undergraduate Calendar - Academic Consideration and Appeals

<https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08-ac.shtml>

Graduate Calendar - Grounds for Academic Consideration

<https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/index.shtml>

Associate Diploma Calendar - Academic Consideration, Appeals and Petitions

<https://www.uoguelph.ca/registrar/calendars/diploma/current/index.shtml>

9.3 Drop Date

Students will have until the last day of classes to drop courses without academic penalty. The deadline to drop two-semester courses will be the last day of classes in the second semester. This applies to all students (undergraduate, graduate and diploma) except for Doctor of Veterinary Medicine and Associate Diploma in Veterinary Technology (conventional and alternative delivery) students. The regulations and procedures for course registration are available in their respective Academic Calendars.

Undergraduate Calendar - Dropping Courses

<https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08-drop.shtml>

Graduate Calendar - Registration Changes

<https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/genreg-reg-regchg.shtml>

Associate Diploma Calendar - Dropping Courses

<https://www.uoguelph.ca/registrar/calendars/diploma/current/c08/c08-drop.shtml>

9.4 Copies of Out-of-class Assignments

Keep paper and/or other reliable back-up copies of all out-of-class assignments: you may be asked to resubmit work at any time.

9.5 Accessibility

The University promotes the full participation of students who experience disabilities in their academic programs. To that end, the provision of academic accommodation is a shared responsibility between the University and the student.

When accommodations are needed, the student is required to first register with Student Accessibility Services (SAS). Documentation to substantiate the existence of a disability is required; however, interim accommodations may be possible while that process is underway.

Accommodations are available for both permanent and temporary disabilities. It should be noted that common illnesses such as a cold or the flu do not constitute a disability.

Use of the SAS Exam Centre requires students to book their exams at least 7 days in advance and not later than the 40th Class Day.

For Guelph students, information can be found on the SAS website
<https://www.uoguelph.ca/sas>

For Ridgetown students, information can be found on the Ridgetown SAS website
<https://www.ridgetownc.com/services/accessibilityservices.cfm>

9.6 Academic Integrity

The University of Guelph is committed to upholding the highest standards of academic integrity, and it is the responsibility of all members of the University community-faculty, staff, and students-to be aware of what constitutes academic misconduct and to do as much as possible to prevent academic offences from occurring. University of Guelph students have the responsibility of abiding by the University's policy on academic misconduct regardless of their location of study; faculty, staff, and students have the responsibility of supporting an environment that encourages academic integrity. Students need to remain aware that instructors have access to and the right to use electronic and other means of detection.

Please note: Whether or not a student intended to commit academic misconduct is not relevant for a finding of guilt. Hurried or careless submission of assignments does not excuse students from responsibility for verifying the academic integrity of their work before submitting it. Students who are in any doubt as to whether an action on their part could be construed as an academic offence should consult with a faculty member or faculty advisor.

Undergraduate Calendar - Academic Misconduct
<https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08-amisconduct.shtml>

Graduate Calendar - Academic Misconduct
<https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/index.shtml>

9.7 Recording of Materials

Presentations that are made in relation to course work - including lectures - cannot be recorded or copied without the permission of the presenter, whether the instructor, a student, or guest lecturer. Material recorded with permission is restricted to use for that course unless further permission is granted.

9.8 Resources

The Academic Calendars are the source of information about the University of Guelph's procedures, policies, and regulations that apply to undergraduate, graduate, and diploma programs.

Academic Calendars
<https://www.uoguelph.ca/academics/calendars>

9.9 Disclaimer

Please note that the ongoing COVID-19 pandemic may necessitate a revision of the format of course offerings and academic schedules. Any such changes will be announced via CourseLink and/or class email. All University-wide decisions will be posted on the COVID-19 website (<https://news.uoguelph.ca/2019-novel-coronavirus-information/>) and circulated by email.

9.10 Illness

The University will not normally require verification of illness (doctor's notes) for fall 2020 or winter 2021 semester courses. However, requests for Academic Consideration may still require medical documentation as appropriate.
