DALHOUSIE UNIVERSITY

FACULTY OF COMPUTER SCIENCE

THERE WILL BE NO CONVERTING OF FINAL GRADES TO PASS/ILL THIS SEMESTER.
ALL LETTER GRADES WILL BE FINAL!

CSCI4181/CSCI6802 --- Algorithms in Bioinformatics Course Syllabus

Key Information

Lecturer:	Dr. Finlay Maguire (finlay.maguire@dal.ca)
Office:	Mona Campbell 4239
Meeting Time:	Tuesday & Thursday 8:35-9:55
Class Location:	Mona Campbell 1201
Homepage:	https://maguire-lab.github.io/bioinformatics_algorithms
Teaching Assistant:	Jee-In Kim (jeein.j.jim@dal.ca)

Note: We do not keep specific office hours – please contact us with any questions or if you would like to make an appointment

Important Dates

- First Class: **January 10th**
- Last Day to Add Winter Courses/Late Registration/Drop Without Financial Penalty: **January 20th**
- Munro Day (University Closed): **February 3rd**
- Last Day to Drop without "W"/Change Audit ↔ Credit: **February 6th**
- Reading Week (No Classes): **February 20th to 24th**
- Last Day to Drop with "W": **March 13th**
- End of Term: **April 12th**

- **Deadlines** (all deadlines are at the end of the day i.e., midnight apart from in-class oral presentations!):
    - Assignment #1: **February 9th**
    - Assignment #2: **March 2nd**
    - Assignment #3: **March 16th**
    - Assignment #4: **March 30th**
    - Selection of Paper for Review: **March 21st**
    - Paper Review (written): **April 4th**
    - Paper Review (Oral): **April 4-11th**

Course Description

Bioinformatics uses computational and statistical approaches to tackle questions of biological function and evolution. The goal of *Algorithms in Bioinformatics* is to introduce key applications of algorithms, data structures, and encodings to the analysis of large biological data sets. A recurring theme throughout the course will be the disconnect between algorithmic beauty and the horrifying realities of biological data. Every statistical model is violated and every classification comes with an asterisk, as we struggle with even the most basic concepts of 'gene' and 'species', and the challenges of understanding events that hap-

pened across ~3.5 billion years. In spite of these challenges, in this age of massive data sets we stand to learn a good deal if the computational tools we use are efficient, robust, properly validated, and correctly applied.

The course covers major challenge areas in bioinformatics, each focused on an aspect of DNA or protein sequence analysis. The goal in each case is to define an overarching problem, and then explore different approaches that have been applied to solving that problem, with an emphasis on the match (or mismatch) between the algorithm and the underlying biological system.

We will begin with an introductory section covering foundational biological and computer science background material. Then we will discuss 4 key challenge areas:

1. **Homology**: Comparing and searching biological sequences
2. **Assembly**: Recovering genomes from sequencing data
3. **Phylogenetics**: Inferring evolutionary relationships from biological sequences
4. **Machine Learning**: Overview of ML and challenges applying these methods to biological data

This course is largely lecture based with the main assessments in the form of 4 practical assignments as well as a (written and oral) summary of a current bioinformatics research paper.

Presenters

We're fortunate to have an exceptional set of individuals who will be giving guest lectures and short introductory presentations throughout the semester. Each of our presenters is conducting active hands-on research in the area they'll be presenting.

Jee In Kim is the course TA and will introduce you to her work in machine learning for the prediction of antimicrobial resistance.

Alex Manuele is a professional bioinformatics developer and an expert in the application of "deep learning" methods to sequence data and will give you an overview of applications from straightforward to more sophisticated embedding approaches.

Dr. Ryan Fink is an expert in food microbiology that will present his experiences applying pangenomic methods across massive bacterial genome datasets.

Prof. Robert Beiko is the original creator of this course and has broad expertise in bioinformatics algorithms (especially as applied to evolutionary microbiology). He will present key algorithms in the inference of phylogenetic networks and lateral gene transfer.

Learning Outcomes

There are three main learning objectives for the course. By the end you should have knowledge of the following:

- Key biological concepts such as genomics and evolution
- How different types of algorithms are appropriate to biological data analysis, and their shortcomings
- The use of widely adopted bioinformatics software tools

Point (1) is the focus of the Introduction, but new material and refinements will be presented throughout the course. Point (2) is the subject of most of the lectures. The tutorial assignments will form the basis of your learning in point (3).

The assignments will give you experience with specific tools that support the following activities:

- Sequence data retrieval
- Homology search
- Multiple sequence alignment
- Phylogenetic analysis
- Machine-learning analysis of genomic data

We will likely use a Linux cluster for the sequence-assembly project, so you will gain experience with using bioinformatics software in an HPC environment.

Finally, the paper review will bring together things you have learnt from throughout the course (as well as exposing you to cutting-edge bioinformatics) and give you experience in presenting/critiquing research literature.

Class Format and Course Communication

- Lectures will be in person (please reach out if this is an issue and we develop accommodations).
- Syllabus/Materials/Schedule will be available via the course website listed above.
- Assignments will be submitted and evaluated via Brightspace.

Evaluation Criteria

We use the standard Dalhousie grading scheme (http://www.dal.ca/campus_life/student_services/academicsupport/grades-and-student-records/grade-scale-and-definitions.html). Note that the minimum passing grade in a graduate course is a B-, so any graduate student mark under 70% will be converted to an F.

Assignments must be submitted via Brightspace. **Late assignments will be penalized at 20% per day**. Assignments submitted after five days will still be evaluated if you would like feedback, although the final grade will still be zero.

Late submissions without penalty will be considered **only if** the appropriate channels (doctor's note, student declaration of absence) are used.

Tutorial Assignments (80% total):

- Four assignments, each worth 20%
- No collaboration is permitted on assignments

Paper review (20% total):

- Written review (10%)
- Oral presentation (10%)

Each of the four tutorial assignments will draw material from one module of the course and will involve the application of one or more methods to a problem data set. The challenge to the student will be to generate, evaluate and interpret the results obtained when different approaches are used.

Students will choose a recent (since 2020) research paper from the bioinformatics literature and produce a written review and 15-minute oral presentation

Tentative Schedule

Date	Week	Course Module	Type	Topic
10-Jan	1	**Introduction**	*Lecture*	Course Overview
			Lecture	Life at Resolution: DNA to Ecosystems
12-Jan			*Lecture*	Central Dogma: Copying and Interpretation of Genetic Information
			Lecture	Biochemical Pathways
17-Jan	2		*Lecture*	Molecular Evolution
19-Jan		**Homology**	*Lecture*	Encoding DNA and Proteins: Strings and Structures
24-Jan	3		*Lecture*	Sequence Alignment: Definitions and Scoring
26-Jan			*Lecture*	Optimal Alignment: Dynamic Programming
31-Jan	4		*Lecture*	Practical Approaches for Large Datasets: BLAST and Burrows-Wheeler Transform
02-Feb			*Lecture*	Multiple Sequence Alignment
			Assignment	Assignment 1: Alignment and Distant Homology (**due: 09-Feb**)
07-Feb	5		*Lecture*	HMMs and Gene Prediction
09-Feb		**Assembly**	*Lecture*	Assembly Part 1: Assembling 1 Genome
			Case Study	Pathogen Pangenomics Ryan Fink
14-Feb	6		*Lecture*	Assembly Part 2: Assembling Many Genomes
16-Feb			*Lecture*	Searching Massive Read-Sets Using K-mers and Graphs
			Assignment	Assignment 2: Genome Assembly (**Due: 02-Mar**)
				Reading Week (Feb 20-24)
28-Feb	7	**Phylogenetics**	*Lecture*	Basic concepts, parsimony, and distance.
02-Mar			*Lecture*	Maximum Likelihood
07-Mar	8		*Lecture*	Bayesian Phylogenetics
09-Mar			*Lecture*	Statistical Support
			Assignment	Assignment 3: SARS-CoV-2 Genome Analysis (**Due: 16-Mar**)
14-Mar	9		*Lecture*	Phylogenetic Networks and Lateral Gene Transfer Rob Beiko
16-Mar		**Machine Learning**	*Lecture*	Feature Selection and Visualisation
21-Mar	10		*Lecture*	How to Train Classifiers
			Assignment	**Deadline** for Paper Selection
23-Mar			*Lecture*	Basic approaches such as Naive Bayes
			Assignment	Assignment 4: Machine-Learning Prediction of AMR Phenotype (**Due: 30-Mar**) Jee-in Kim
28-Mar	11		*Lecture*	Support Vector Machines and Random Forests
30-Mar			*Lecture*	Artificial Neural Networks and "Deep Learning" Alex Manuele
04-Apr	12		*Assignment*	**Deadline** for Written Paper Reviews/ **Presentations**
06-Apr			*Assignment*	**Presentations**
11-Apr	13		*Assignment*	**Presentations**

Student Declaration of Absence

The Student Declaration of Absence policy shall apply. https://www.dal.ca/campus_life/safety-respect/student-rights-and-responsibilities/academic-policies/student-absence.html The student has a maximum of two (2) SDAs per course per semester. The student **must** notify the instructor of their inability to meet a deadline **before** the deadline by contacting the instructor or submitting the completed SDA. Upon notification the student has 3 days after the deadline to submit the SDA.

Academic Standards

Failure to properly attribute sources in your work will be treated as an academic standards issue and points may be deducted for not following citation requirements. For example, forgetting to quote text taken from other sources, failure to include in-text citations, or a failure to include required information in the citations or references. Please see the resources on proper citation provided by the Dalhousie Writing Center (https://dal.ca.libguides.com/c.php?g=257176&p=5001261).

Please note that if it appears that the error was made with intent to claim other people's work as your own such as a lack of both citations and references, an allegation of plagiarism will be submitted to the Faculty Academic Integrity Officer, which could result in consequences such as a course failure.

Responsible Computing Policy

Usage of all computing resources in the Faculty of Computer Science must be within the Dalhousie Acceptable Use Policies (https://www.dal.ca/dept/university_secretariat/policies/information-management-and-technology/acceptable-use-policy-.html) and the Faculty of Computer Science Responsible Computing Policy. For more information please see https://www.dal.ca/content/dam/dalhousie/pdf/faculty/computerscience/policies-procedures/fcs_policy_local.pdf

Use of Plagiarism Detection Software

All submitted code may be passed through a plagiarism detection software, such as the plagiarism detector embedded in Codio, the Moss (https://theory.stanford.edu/~aiken/moss/) Software Similarity Detection System, or similar systems. If a student does not wish to have their assignments passed through plagiarism detection software, they should contact the instructor for an alternative. Please note, that code not passed through plagiarism detection software will necessarily receive closer scrutiny. https://cdn.dal.ca/content/dam/dalhousie/pdf/dept/university_secretariat/policy-repository/OriginalitySoftwarePolicy.pdf

Student Health and Wellness

Taking care of your health is important. As a Dalhousie student, you have access to a wide range of resources to support your health and wellbeing. Students looking to access physical or mental health & wellness services at Dalhousie can go to the Student Health & Wellness Centre in the LeMarchant Building. The team includes: registered nurses, doctors, counsellors and a social worker. Visit **dal.ca/studenthealth** to learn more and book an appointment today.

Students also have access to a variety of online mental health resources, including telephone/texting counselling and workshops/training programs. Learn more and access these resources at d**al.ca/mentalhealth**.

Culture of Respect[1]

Every person has a right to respect and safety. We believe inclusiveness is fundamental to education and learning. Misogyny and other disrespectful behaviour in our classrooms, on our campus, on social media, and in our community is unacceptable. As a community, we must stand for equality and hold ourselves to a higher standard.

[1] Source: Speak Up! © 2005 Southern Poverty Law Center. First Printing. This publication was produced by Teaching Tolerance, a project of the Southern Poverty Law Center. Full "Speak Up" document found at: http://www.dal.ca/dept/dalrespect.html. Revised by Susan Holmes from a document provided April 2015 by Lyndsay Anderson, Manager, Student Dispute Resolution, Dalhousie University, 902.494.4140, lyndsay.anderson@dal.ca www.dal.ca/think.

What we all need to do:

1. **Be Ready to Act:** This starts with promising yourself to speak up to help prevent it from happening again. Whatever it takes, summon your courage to address the issue. Try to approach the issue with open-ended questions like "Why did you say that?" or "How did you develop that belief?"
2. **Identify the Behaviour:** Use reflective listening and avoid labeling, name-calling, or assigning blame to the person. Focus the conversation on the behaviour, not on the person. For example, "The comment you just made sounded racist, is that what you intended?" is a better approach than "You're a racist if you make comments like that."
3. A**ppeal to Principles:** This can work well if the person is known to you, like a friend, sibling, or co-worker. For example, "I have always thought of you as a fair-minded person, so it shocks me when I hear you say something like that."
4. **Set Limits:** You cannot control another person's actions, but you can control what happens in your space. Do not be afraid to ask someone "Please do not tell racist jokes in my presence anymore" or state "This classroom is not a place where I allow homophobia to occur." After you have set that expectation, make sure you consistently maintain it.
5. **Find or be an Ally:** Seek out like-minded people that support your views, and help support others in their challenges. Leading by example can be a powerful way to inspire others to do the same.
6. **Be Vigilant:** Change can happen slowly, but do not let this deter you. Stay prepared, keep speaking up, and do not let yourself be silenced.

University Statements

This course is governed by the academic rules and regulations set forth in the University Calendar and the Senate. https://academiccalendar.dal.ca/Catalog/ViewCatalog.aspx?pageid=viewcatalog&catalogid=117&loaduseredits=False

Territorial Acknowledgement

Dalhousie University is located in Mi'kma'ki, the ancestral and unceded territory of the Mi'kmaq. We are all Treaty people.

Dalhousie acknowledges the histories, contributions, and legacies of the African Nova Scotia people and communities who have been here for over 400 years.

Internationalization

At Dalhousie, 'thinking and acting globally' enhances the quality and impact of education, supporting learning that is "interdisciplinary, cross-cultural, global in reach, and orientated toward solving problems that extend across national borders." https://www.dal.ca/about-dal/internationalization.html

Academic Integrity

At Dalhousie University, we are guided in all of our work by the values of academic integrity: honesty, trust, fairness, responsibility and respect. As a student, you are required to demonstrate these values in all of the work you do. The University provides policies and procedures that every member of the university community is required to follow to ensure academic integrity. (read more: http://www.dal.ca/dept/university_secretariat/academic-integrity.html)

Accessibility

The Student Accessibility Centre is Dalhousie's centre of expertise for matters related to student accessibility and accommodation. If there are aspects of the design, instruction, and/or experiences within this course (online or in-person) that result in barriers to your inclusion please contact: https://www.dal.ca/campus_life/academic-support/accessibility.html for all courses offered by Dalhousie with the exception of Truro.

Conduct in the Classroom — Culture of Respect

Substantial and constructive dialogue on challenging issues is an important part of academic inquiry and exchange. It requires willingness to listen and tolerance of opposing points of view. Consideration of individual differences and alternative viewpoints is required of all class members, towards each other, towards instructors, and towards guest speakers. While expressions of differing perspectives are welcome and encouraged, the words and language used should remain within acceptable bounds of civility and respect.

Diversity and Inclusion — Culture of Respect

Every person at Dalhousie has a right to be respected and safe. We believe inclusiveness is fundamental to education. We stand for equality. Dalhousie is strengthened in our diversity. We are a respectful and inclusive community. We are committed to being a place where everyone feels welcome and supported, which is why our Strategic Direction prioritizes fostering a culture of diversity and inclusiveness (Strategic Priority 5.2). (read more: http://www.dal.ca/cultureofrespect.html)

Student Code of Conduct

Everyone at Dalhousie is expected to treat others with dignity and respect. The Code of Student Conduct allows Dalhousie to take disciplinary action if students don't follow this community expectation. When appropriate, violations of the code can be resolved in a reasonable and informal manner—perhaps through a restorative justice process. If an informal resolution can't be reached, or would be inappropriate, procedures exist for formal dispute resolution. (read more: https://cdn.dal.ca/content/dam/dalhousie/pdf/dept/university_secretariat/policy-repository/Code%20of%20Student%20Conduct%20rev%20Sept%202021.pdf)

Fair Dealing Policy

The Dalhousie University Fair Dealing Policy provides guidance for the limited use of copyright protected material without the risk of infringement and without having to seek the permission of copyright owners. It is intended to provide a balance between the rights of creators and the rights of users at Dalhousie. (read more: https://www.dal.ca/dept/university_secretariat/policies/academic/fair-dealing-policy-.html)

Originality Checking Software

The course instructor may use Dalhousie's approved originality checking software and Google to check the originality of any work submitted for credit, in accordance with the Student Submission of Assignments and Use of Originality Checking Software Policy. Students are free, without penalty of grade, to choose an alternative method of attesting to the authenticity of their work, and must inform the instructor no later than the last day to add/drop classes of their intent to choose an alternate method. (read more: https://cdn.dal.ca/content/dam/dalhousie/pdf/dept/university_secretariat/policy-repository/OriginalitySoftware-Policy.pdf)

Student Use of Course Materials

These course materials are designed for use as part of the CSCI courses at Dalhousie University and are the property of the instructor unless otherwise stated. Third party copyrighted materials (such as books, journal articles, music, videos, etc.) have either been licensed for use in this course or fall under an excep-

tion or limitation in Canadian Copyright law. Copying this course material for distribution (e.g. uploading material to a commercial third party website) may lead to a violation of Copyright law.

Learning and Support Resources

Please see https://www.dal.ca/campus_life/academic-support.html