CSI 5126 (COMP 5108). Algorithms in Bioinformatics

Marcel Turcotte
University of Ottawa
School of Electrical Engineering and Computer Science (EECS)
marcel.turcotte@uottawa.ca

Fall 2018

[PDF]

Course Hours

- **Lectures**: Tuesday 13:00–14:30 and Thursday 11:30–13:00 SMD 402
- **Office hours**: Thursday, 14:30–16:00

The first lecture will be held on September 6, 2018. Information about the schedules, lecture notes, and more can be found on the course Web page.

Professor

- Marcel Turcotte (marcel.turcotte@uottawa.ca)
- **Contact details**: STE 5-106, (613) 562-5800 ext. 7441
- **URL**: `http://www.eecs.uottawa.ca/~turcotte/teaching/csi-5126`

Calendar Description

Fundamental mathematical and algorithmic concepts underlying computational molecular biology; physical and genetic mapping, sequence analysis (including alignment and probabilistic models), genomic rearrangements, phylogenetic inference, computational proteomics and systemic modeling of the whole cell. Prerequisites: CSI 3105, COMP 3804 or equivalent.

Course Learning Outcomes

Upon completion of the course, student will be able to:

- List and describe the fundamental algorithms in bioinformatics
- Articulate the trade-offs behind algorithms in bioinformatics
- Write computer programs for solving large scale bioinformatics problems
- Critically review scientific publications in this field
- Locate and critically evaluate scientific information
- Apply one of the paradigms presented in class to solve real-world problems
- Present scientific content to a small technical audience

Course Calendar

Here is a draft of the week by week content.

1. **Syllabus and introduction, course objectives and students expectations.**
2. **Essential cell biology.**
   Wiesława Widłak (2013). Molecular Biology: Not Only for Bioinformaticians (Vol. 8248). Springer. Chapters 1, 2, and 3.
3. **Essential Cell Biology.**
   Wiesława Widłak (2013). Molecular Biology: Not Only for Bioinformaticians (Vol. 8248). Springer. Chapters 4, 5, and 9.
4. **Suffix Trees.**
   Bernhard Haubold and Thomas Wiehe (2006). Introduction to computational biology: an evolutionary approach. Birkhäuser Basel. Pages 43-53.
   Dan Gusfield (1997) Algorithms on strings, trees, and sequences : computer science and computational biology. Cambridge University Press. Chapters 5, 6 (optional), 7.
5. **Suffix Trees.**
   Bernhard Haubold and Thomas Wiehe (2006). Introduction to computational biology: an evolutionary approach. Birkhäuser Basel. Pages 43-53.
   Dan Gusfield (1997) Algorithms on strings, trees, and sequences : computer science and computational biology. Cambridge University Press. Chapters 8 (optional), 9.
6. **Molecular Sequence Alignment**.
   Bernhard Haubold and Thomas Wiehe (2006). Introduction to computational biology: an evolutionary approach. Birkhäuser Basel. Pages 11-15, 30-33.
7. **Molecular Sequence Alignment.**
   Bernhard Haubold and Thomas Wiehe (2006). Introduction to computational biology: an evolutionary approach. Birkhäuser Basel. Pages 35-39.
8. **Substition Matrices.**
   Warren J. Ewens, Gregory R. Grant (2001) Statistical Methods in Bioinformatics: An Introduction. Springer. Pages: 238-249.
9. **Inferring Phylogenies.**
   Bernhard Haubold and Thomas Wiehe (2006). Introduction to computational biology: an evolutionary approach. Birkhäuser Basel. Pages 143-168.
10. **Inferring Phylogenies.**
    Bernhard Haubold and Thomas Wiehe (2006). Introduction to computational biology: an evolutionary approach. Birkhäuser Basel. Pages 143-168.
11. **RNA Secondary Structure. Dynamic programming.**
12. **RNA Secondary Structure. Cocke-Younger-Kasami.**
13. **Sequence Motif (Median).**
14. **Sequence Motif (Hidden Markov Model).**
    Bernhard Haubold and Thomas Wiehe (2006). Introduction to computational biology: an evolutionary approach. Birkhäuser Basel. Pages 101-116.
15. **MotifGP.**
16. **ModuleInducer.**

Evaluations

- Assignments (2) (20%)
- Midterm (20%)
- Review and oral presentation of a scientific publication (10%)
- Project (50%)
  - Outline (10%)
  - Presentation (10%)
  - Report (30%)

Oral Presentations

Papers in (refereed) journals and conference proceedings are the main vehicules for communicating scientific information. You must select a publication that presents either a specialized application or a more efficient algorithm on a topic that has been presented in class.

Objectives

- Thoroughly study of a specific topic in bioinformatics
- Familiarity with the modes of communicating research
- Develop your presentation skills

Deliverable

- One or two pages summary of the publication
- 15–20 minutes presentation

Project

Objectives

- Thorough study of a specific bioinformatics topic
- Learning to study autonomously
- Develop your presentation skills

Deliverable

The format of the projects is quite flexible. I foresee three broad types of work: the development of a novel application, the analysis of a new data set, or a (thorough) review of the literature on a specific topic. For all three types of work, I would like to see a review of the literature, sample data and a prototype implementation (where applicable). The main difference between each type of work will be the relative importance of each of the components.

Teamwork

Teams will be made of 1 or 2 members. Larger teams will have to produce proportionally more work! Complementary work between teams is also welcomed, i.e. two or more teams working on a related but complementary topic, leading to a more realistic application.

Report

Reports should be sufficiently detailed that it should be possible to implement the approach on the basis of the text alone. Having said that, you should also make every conceivable effort to keep the report concise. Assuming a team of size 2, a 10–15 page report should be appropriate. Structure for the reports:

- Introduction
  - Background
  - Problem definition
  - Describing the data
- Methods
- Results
- Conclusions
- Future research
- Full list of references

Material and Resources

Lecture notes (slides) and complementary resources will be posted on the course Web site: http://www.eecs.uottawa.ca/~turcotte/teaching/csi-5126/lectures.

Monographs

There is no required textook. However, the following book is particularly relevant in terms of the topics that it covers and the emphasis on algorithms.

- Wing-Kin Sung (2010) Algorithms in Bioinformatics: A Practical Introduction. Chapman & Hall/CRC. QH 324.2 .S86 2010

The third edition of the following textbook has been published at the end of August 2018. I will certainly draw inspiration from that book:

- Pavel A. Pevzner and Phillip Compeau (2018) Bioinformatics Algorithms: An Active Learning Approach. Active Learning Publishers. http://bioinformaticsalgorithms.com

The following two monographs are part of my short list of essential bioinformatics books.

- Durbin, R. *et al* (1998, 2000) Biological sequence analysis: probabilistic models of proteins and nucleic acids. Cambridge University Press.
  (QP 620 .B576 1998)
- Gusfield, D. (1997) Algorithms on strings, trees, and sequences: computer science and computational biology. Cambridge University Press.
  (QA 76.9 .A43 G87 1997)

Documents on the Web

Springer Link provides our community with access to journals, books, series, protocols and reference documents (access restricted to University of Ottawa based on your IP address).

- Bernhard Haubold and Thomas Wiehe (2006). Introduction to computational biology: an evolutionary approach. Birkhäuser Basel.
  - http://link.springer.com/book/10.1007%2F3-7643-7387-3
- Böckenhauer, H.-J. and Bongartz, D. (2007) Algorithmic Aspects of Bioinformatics: Springer Ebooks
  - www.springerlink.com/content/v53674
- Sperschneider, Volker (2008) Bioinformatics: Problem Solving Paradigms. Springer.
  - www.springerlink.com/content/978-3-540-78505-7
- Aluru S. (2006) Handbook of Computational Molecular Biology. Chapman & Hall/CRC.
  (QH 324.2 .H357 2006)
  - http://www.engnetbase.com/books/4726/c061fm.pdf
- Warren J. Ewens, Gregory R. Grant (2001) Statistical Methods in Bioinformatics: An Introduction. Springer.
  - http://link.springer.com/book/10.1007%2F978-1-4757-3247-4

Essential Cell Biology

- Wiesława Widłak (2013). Molecular Biology: Not Only for Bioinformaticians (Vol. 8248). Springer.
  - http://link.springer.com/book/10.1007/978-3-642-45361-8
- Brown, T. A. (2007) Genomes. Oxford: BIOS Scientific Publishing. (QH 447 .B76 2007)
  - http://www.ncbi.nlm.nih.gov/books/bv.fcgi?call=bv.View..ShowTOC&rid=genomes.TOC&depth=2
- Rodden Robinson, T. (2005) Genetics For Dummies: For Dummies.

Plagiarism

Academic fraud is an act by a student that may result in a false evaluation (including papers, tests, examinations, etc.). It is not tolerated by the University. Any person found guilty of academic fraud will be subject to severe sanctions.

Here are some examples of academic fraud:

- Plagiarism or cheating of any kind;
- Present research data that has been falsified;
- Submit a work for which you are not the author, in whole or part;
- Submit the same piece of work for more than one course without the written consent of the professors concerned.
- Please consult this webpage: it contains regulations and tool to help you avoid plagiarism.

An individual who commits or attempts to commit academic fraud, or who is an accomplice, will be penalized. Here are some examples of possible sanctions:

- Receive an "F" for the work or in the course in question;
- Imposition of additional requirements (from 3 to 30 credits) to the program of study;
- Suspension or expulsion from the Faculty.
- You can refer to the regulations on this webpage.

Student Services

Academic Writing Help Centre

At the AWHC you will learn how to identify, correct and ultimately avoid errors in your writing and become an autonomous writer.

In working with our Writing Advisors, you will be able to acquire the abilities, strategies and writing tools that will enable you to:

- Master the written language of your choice
- Expand your critical thinking abilities
- Develop your argumentation skills
- Learn what the expectations are for academic writing
- http://www.sass.uottawa.ca/writing/

Career Services

Career Services offers various services and a career development program to enable you to recognize and enhance the employability skills you need in today's world of work.

- http://www.sass.uottawa.ca/careers/

Counselling Service

There are many reasons to take advantage of the Counselling Service. We offer:

- Personal counselling
- Career counselling
- Study skills counselling

- http://www.sass.uottawa.ca/personal/

Access Service

The Access Service acts as intermediary between students, their faculty and other University offices to ensure that the special needs of these students are addressed and that the best possible learning conditions are being offered.

Note that the University of Ottawa is affiliated with AERO and ACE services for the adaptation of accessible academic materials for students with perceptual disabilities. If you have any questions, please contact the Accessibility Librarian or the Access services for textbooks.

- http://www.sass.uottawa.ca/access/

Policy – Prevention of Sexual Violence

The University of Ottawa will not tolerate any act of sexual violence. This includes acts such as rape and sexual harassment, as well as misconduct that take place without consent, which includes cyberbullying. The University, as well as various employee and student groups, offers a variety of services and resources to ensure that all uOttawa community members have access to confidential support and information, and to procedures for reporting an incident or filing a complaint. For more information, please visit www.uOttawa.ca/sexual-violence-support-and-prevention.

About the Instructor

I have been conducting research in bioinformatics since 1989. I am particularly interested in the inference of structural motifs, mostly for ribonucleic acids.