Bioinformatics 200
Introduction to Bioinformatics
Term 2 (Winter), 2015/16

Syllabus

Index

Description

This course is designed to be a capstone course in the bioinformatics program, as well as be a service class to students in the disciplines contributing to bioinformatics. It integrates material in the life sciences, computational science, and mathematics. The course introduces problems and major concepts in bioinformatics, trains students in a variety of basic techniques, and introduces terminology. It also illustrates the value of, and need for, cooperation, teamwork, and multi-disciplinary collaboration in areas such as bioinformatics. While BINF210 focuses mostly on bioinformatics tool use, BINF 200 provides more in-depth coverage of algorithms used by such tools and teaches relevant programming skills so that students can build software "pipelines" consisting of existing tools, as well as build their own basic tools.

Prerequisites

The prerequisites for BINF 200 are CMPT 111 and BMSC 200 (or equivalents). It is your responsibility to insure that you have met these prerequisites. If you don't, and you are in Arts & Science, you will not get credit for the course

for the course.

Scheduling

Class day & time: Monday, Wednesday, Friday; 10:30 a.m. - 11:20 a.m.
Class Location: Thorvaldson S311 (in the Spinks Addition)
Class duration: January 6 through April 6, 2016
Midterm Exam: in class, February 26, 2016 (the last day for withdrawls is March 15)
Lab day & time: Fridays, 3:30pm-4:50pm
Lab Location: Thorvaldson S311 (in the Spinks Addition)
Labs start: week of January 8th

Website

The website for this course is on the moodle server of the Department of Computer Science. The URL is `https://moodle.cs.usask.ca/course/view.php?id=314`

Instructor Information

Dr. Tony Kusalik
email: `kusalik@cs.usask.ca` (do not cut-and-paste this address as it has hidden, "garbage" characters in it)
office: Thorv S424 (in the Spinks Addition)
telephone: 966-4904

Office Hours

The professor is typically available immediately before and after the class for questions, advice, etc.

The professor has combined advising time / office hour every Friday between 11:30 and 12:30. At this time, he can be found in his office, Thorv S424. You can set up an appointment during this time, or you can just "drop in". Priority will be given to students who have made an appointment.

The prof is available to meet students at other times. Simply send an email message and set up an appointment!

Finally, remember that, if you need help, e-mail works 24 hours a day — and you'll probably even get a response in short order! Alternatively post something to one of the class forums on moodle.

Note

Any instance of my email address in the class web pages, including the instance above, has invisible garbage characters embedded in it to foil automated email address "harvesters". Thus you should not cut-and-paste those instances into the address field of a mail program. Type the address instead.

Teaching Assistant

Matthew Shannon `mfs953@mail.usask.ca`

Course Objectives

By the completion of this course, students will be expected to:

- be familiar with the nature of many of the main topic areas in bioinformatics;
- understand basic terminology in the field;
- understand fundamental concepts regarding pairwise sequence alignment, multiple sequence alignment, scoring matrices, heuristic algorithms for sequence-based search of biological databases, sequence patterns and motifs, evolutionary models and phylogenetics, homology-based protein functional prediction, visualization and prediction of protein secondary structure, visualization and prediction of protein tertiary structure, visualization and prediction of RNA secondary structure, genome sequencing and sequence assembly, and gene expression analysis;
- understand fundamental characteristics of dynamic programming algorithms, extreme value distributions, and regular expressions;
- understand the role and purpose of commonly-used bioinformatics databases and the relationships between them.

By the completion of this course, students will be expected to be able to:

- skillfully use UNIX/LINUX shell commands;
- write non-trivial scripts in UNIX shell (*bash*) and *Perl*;
- perform pairwise sequence alignments using dot plots, and the Needleman-Wunsch and Smith-Waterman algorithms;
- perform sequence-based searches of biological databases using various forms of BLAST, and interpret the results;
- select parameters for a sequence alignment algorithm based on the biological problem to be solved;
- effectively utilize various programs in the EMBOSS suite;
- perform a multiple sequence alignment, and build a sequence motif and sequence logo from the results;
- perform a pattern-based search of sequence database;
- effectively use multiple bioinformatics databases to answer biological questions;
- construct a phylogenetic tree from sequence information representing the relatedness of a group of organisms;
- proficiently use programs for predicting protein characteristics, protein secondary structure, protein tertiary structure, and RNA secondary structure;
- visualize biomolecular three-dimensional structures;
- assemble a small genome from a collection of sequence reads;

- and compare gene expression under various conditions or in different tissues.

Note that the above two lists are not exclusive.

Class Format

In general, the class will involve lectures, in-class demonstrations, guided tutorials, lab exercises, assignments, and written examinations.

The course will be divided into a number of topic areas. Through lecture, the instructor will provide essential background information (either in Molecular Biology or Computer Science) on starting each area. Afterward, techniques and detailed subject material will be presented, either through lecture or demonstration. Demonstrations will be performed on the equipment in the classroom. Finally, students will be required to complete laboratory exercises and assignment problems. The solution of these exercises and problems are key to students' learning in the course. Through them students will learn more about the computer science techniques involved, the biological problem being tackled, and the use of existing applicable (software) tools.

There will be a distinction between lab exercises and assignments. Specifically, the former are intended for practice with basic bioinformatics tools and gaining experience with basic concepts and tools in an environment where a lab attendant is available. They are intended to be successfully completed in about 1 to 2 hours of work. Assignments, on the other hand, are intended to develop a deeper understanding of issues. Solutions to assignment problems will require more thought and preparation. Both lab exercises and assignments are to be completed on an individual basis. However, it is possible to discuss solution strategies with others.

A midterm will be held as well as a two-part final exam. The date of the midterm is tentatively February 26, 2016. It will be a closed-book exam, though students will be able to bring a "cheat sheet". The final exam will consist of both a written portion and an in-lab portion. Both portions of the final exam will be "open book".

Student Evaluation

Grading Scheme

	contribution
assigments	20%
labs	15%
class participation	2%
mid-term exam	18%
final exams	
lab final exam	12%
written final exam	33%

Important Notes

- The participation grade will be based on evidence of any constructive and/or contributive involvement in (postings to) the on-line discussion forums for the class.
- There will be 12 lab exercises, the first being made available during the week of January 8. There will

There will be 12 lab exercises, the first being made available during the week of January 3. There will also be 4 assignments, the first being made available in approximately the 4th week of January. See sections of this syllabus on Instructional Laboratory and Assignments, respectively, for more information.

- If a student does not write the midterm exam, the corresponding portion of the final grade allocation will be re-assigned to the final exam.
- As mentioned above, the date of the midterm is tentatively February 26, 2016.
- If students know in advance that they will miss a lab, an assignment, or exam due to extenuating circumstances, they should contact the professor well in advance so that alternate arrangements can be made.
- The midterm exam will be closed book, with the exception of students being allowed to bring a one-page "cheat sheet". Both portions of the final exam will will be "open book". Students needing special aids or allowances for writing exams (i.e. DSS students) must make arrangements prior to the exam with the instructor for arrangement of these aids or allowances.
- It is preferable that students complete labs and assignments even if it means submitting them late; the potential for knowledge gain is more important than strict enforcement of submission deadlines. However, it is not fair to other students or to the marker if some students are allowed to submit late assignments with no consequences. Hence, late assignments and labs will be accepted, but they will be docked a penalty that will grow exponentially with the duration late. The specifics of the functions are given under the Assignments and Labs sections, respectively, of the moodle pages for the class.

Criteria That Must Be Met To Pass

- Laboratory exercises and assignments are required (mandatory) elements of the class. Therefore, in accordance with the University's "academic courses policy on course delivery, examinations & assessment of student learning" lab exercises and assignments must be completed in order to achieve a passing grade. For instance, students who skip the labs or don't bother doing the assignments will not pass the class no matter how well they "ace" the final exam. However, students may be exempted from having to complete a particular lab or assignment due to extenuating circumstances on a case-by-case basis (e.g. for medical reasons or serious personal matters). Students should contact the instructor to set up special provisions for the lab or assignment as soon as they are able.

  For the purposes of the University's "academic courses policy on course delivery, examinations & assessment of student learning", if more than one assignment is missed and no exemptions have been arranged, a student will be deemed to have not successfully completed a required element of the class (the assignment portion). This, in turn, will mean that the student will not receive a passing grade in the

  course. That is, unless special provisions have been arranged, no more than one assignment can be missed for a student to pass the class.

  For the purposes of the University's "academic courses policy on course delivery, examinations & assessment of student learning", if more than two labs are missed and no exemptions have been arranged, a student will be deemed to have not successfully completed a required element of the class (the laboratory exercises) and will not receive a passing grade in the course. That is, unless special provisions have been arranged, no more than two labs can be missed for a student to pass the class.

- The final exam is another required element of the class. Failure to write the final exam will result in failure of the course.

Final Exam Scheduling

The Registrar schedules all final examinations, including deferred and supplemental examinations. Students are advised not to make travel arrangements for the exam period until the official exam schedule has been posted.

Texts and Library Materials

Textbooks

There are two recommended textbooks for the class:

- *Bioinformatics and Functional Genomics, 3rd edition*, by Jonathan Pevsner, published by Wiley-Blackwell (2015) and
- *UNIX and Perl to the Rescue!: A Field Guide for the Life Sciences (and Other Data-rich Pursuits)* by K. Bradnam & I. Korf, published by Cambridge University Press (2012).

Both textbooks should be available in the Bookstore. The are also available in electronic form (for purchase) from the publishers, as well as retailers such as `amazon.ca`. For example, see `http://ca.wiley.com/WileyCDA/WileyTitle/productCd-1118581784.html`, `http://ca.wiley.com/WileyCDA/WileyTitle/productCd-0470085851.html`. The University Library has the Bradnam & Korf text as well as earlier editions of the Pevsner text.

Students can substitute the 2nd edition of the Pevsner text (published in 2009). It has less coverage and is less current, but is available is paperback form. Also, used copies of the 2nd edition will have greater availability. For instance, the 2nd edition of the Pevsner text has been used for several years as the textbook in BINF 210.

The author of the Pevsner textbook has resources for students using the text at `http://www.bioinfbook.org`. Bradnam & Korf have a website to accompany their text. They also have a free, on-line primer.

References in the Library

The University libraries have large collections of books covering the areas of bioinformatics, computational biology, using LINUX/UNIX, and programming in Perl. Students are encouraged to make use of them. Many of the items are "E-books".

Lecture Schedule and Readings

The following table provides a tentative schedule of high level topics for the class, and when approximately that topic will be covered. It also describes where relevant reference material can be found in the recommended textbooks for the class.

Week	Topic	Pevsner		Bradnam & Korf
		2nd ed.	3rd ed.	
1	Course Introduction			
	What is "Bioinformatics"?	pp. 4-8	pp. 10-15	
	Review of Relevant Molecular Biology	pp. 279-301, 381-386, 421-430	pp. 433-451, 543-550, 589-598	

2	The Bioinformatics Workstation / LINUX			Part 3, Part 5
	The UNIX/LINUX shell			
3	The UNIX/LINUX shell			
	Perl for Bioinformatics			Part 4
4	Perl for Bioinformatics			
	Biological Resources on the WWW	Ch 2	Ch 2	
5	EMBOSS	pg. 81		
	Pairwise Sequence Alignment	Ch 3	Ch 3	
6	Pairwise Sequence Alignment			
	Database Searching with BLAST	Ch 4, pp. 141-144	Ch 4, pp. 167-170	
7	Database Searching with BLAST			
	PSSMs and PSI-BLAST	pp. 145-161	pp. 171-179	
	midterm			
8	Multiple Sequence Alignment	Ch 6	Ch 6	
	Profiles and Motifs	pp. 389-397	pp. 552-558	
9	Molecular Evolution	Ch 7	Ch 7	
	Phylogenetics			
10	Sequence Assembly	pp. 525-552	pp. 378-395	
	Whole Genome Comparison	pp. 161-169	pp. 186-196	
	Gene Finding	pp. 169-172, 555-558, 617-619	pp. 334-338, 737-742, 819-824	
11	Visualizing Protein Structure	Ch 11, pp. 397-401	Ch 13, pp. 559-566	
	Predicting Protein Secondary Structure			
12	Predicting Protein Tertiary Structure			
	Predicting RNA Secondary Structure			
13	Gene Expression	Ch 9	Ch 11	
	Gene Expression Analysis			

Instructional Laboratory

As described under "Class Format", the laboratory session is for gaining practical experience in the application of concepts introduced in the lectures. The lab periods are 90 minutes in length. An initial portion of the lab time may be devoted to providing introductory information. The remainder of the period will be for completion of the laboratory exercise. The lab instructor will be available during this time to provide assistance as necessary.

Students are required to attend the weekly laboratory session. Each week's lab component is to be completed in

Students are required to attend the weekly laboratory session. Each week's lab component is to be completed in the time period specified and a solution either shown to the instructor for immediate evaluation, or submitted electronically. A grace period, beyond the end of the laboratory session, for submission of the solution will be given. The duration of the grace period will be specified as part of the lab specification. Beyond the grace period, late submissions will be accepted, though they will be subject to an exponentially increasing penalty.

Important Notes

- Laboratory exercises are a required (mandatory) element of the class. Therefore, in accordance with the University's "academic courses policy on course delivery, examinations & assessment of student learning" lab exercises must be completed and submitted in order to achieve a passing grade. However, students may be exempted from having to complete a particular lab due to extenuating circumstances on a case-by-case basis (e.g. for medical reasons or serious personal matters). Alternatively, the student may be given an extension on the due date. Students should contact the instructor to set up special provisions for the lab as soon as they are able.

- For the purposes of the University's "academic courses policy on course delivery, examinations & assessment of student learning", if more than two labs are missed and no exemptions have been arranged, a student will be deemed to have not successfully completed a required element of the class (the laboratory exercises) and will not receive a passing grade in the course. That is, unless special provisions have been arranged, no more than two labs can be missed for a student to pass the class.

- Late labs will be accepted, though they will be docked a penalty. The penalties will be determined by an exponential function, the specifics of which are given under the Labs section of the moodle pages for the class.

Teaching Assistant

Name: Matthew Shannon
E-mail: `mfs953@mail.usask.ca` (do not cut-and-paste this address as it has hidden, "garbage" characters in it)

Assignments

There will be about 4 assignments in the class. They will not be long assignments, though the problems will be harder than those in the laboratory. Students will typically have 2 weeks to complete each assignment. Assignment solutions will be submitted electronically via the the moodle pages for the class. Late submissions will be accepted, though they will be docked accordingly. The late penalty will be determined using an exponential function.

Important Notes

- Assignments are required (mandatory) element of the class. Therefore, in accordance with the University's "academic courses policy on course delivery, examinations & assessment of student learning", assignments must be completed in order to achieve a passing grade. However, students may be exempted from having to complete a particular assignment due to extenuating circumstances on a case-by-case basis (e.g. for medical reasons or serious personal matters). Alternatively, the student may be given an extension on the due date. Students should contact the instructor to set up special provisions for the assignment as soon as they are able.

- For the purposes of the University's "academic courses policy on course delivery, examinations & assessment of student learning", if more than one assignment is missed and no exemptions have been

arranged, a student will be deemed to have not successfully completed a required element of the class (the assignment portion). This, in turn, will mean that the student will not receive a passing grade in the course. That is, unless special provisions have been arranged, no more than one assignment can be missed for a student to pass the class.

- Late assignment submissions will be accepted, though they will be docked a penalty. The penalties will be determined by an exponential function, the specifics of which are given under the Assignments section of the moodle pages for the class.

Laboratory Resources

The main computational resources for this course will be the Mac OS and LINUX operating systems, the EMBOSS package of bioinformatics software, the Perl scripting language, and various other bioinformatics software packages.

Students will have accounts in the Computer Science computing laboratories (e.g. rooms THORV S311, S320, S360). The class will make use of the Apple Macintosh workstations in S311. However, any of the LINUX or Mac OS X workstations on the 3rd floor can be used at other times. For remote access (e.g. from home) to a LINUX system, log in to `tuxworld.usask.ca` using the *ssh* program or a virtual terminal program that supports the *ssh* protocol. Information on how to set up remote access to the Computer Science Department's computational lab resources can be found at `http://www.cs.usask.ca/support/index.php#all`, under "Remote Access to LINUX" or "Remote Access to Apple Mac".

On-Line Resources

On-line forums on topics related to the class are hosted on the moodle pages for the class.
The author of the Pevsner textbook also has resources for students using the text at `http://www.bioinfbook.org`.

The website to accompany the Bradnam & Korf is `http://unixandperl.com`.

The website `http://www.ee.surrey.ac.uk/Teaching/Unix/` has a good tutorial introduction to UNIX and LINUX.

The following are the URLs of institutions hosting major bioinformatics resources:

- NCBI
- EBI
- EMBL
- PIR (Protein Identification Resource)
- PDB (Protein Data Bank)
- The Sanger Center

Policies

Late Assignments and Labs

- Policies on late assignments are outlined in the Student Evaluation section of this document, and re-

iterated in the Assignments section.

- Policies on late labs are outlined in the Student Evaluation section of this document, and re-iterated in the Instructional Laboratory section.

Missed Assignments and Labs

- Policies on missed assignments are outlined in the Student Evaluation section of this document, and re-iterated in the Assignments section.
- Policies on missed labs are outlined in the Student Evaluation section of this document, and re-iterated in the Instructional Laboratory section.

Missed Examinations

Policies of the College of Arts & Science apply to this course in the case of missed examinations. Those policies are described at http://www.usask.ca/programs/colleges-schools/arts-science/#AcademicInformationampPolicies. Expand the item entitled "Deferred and Supplemental Examinations".

Incomplete Course Work and Final Grades

The procedures and policies regarding incomplete course work and final grades described in item 8.5 of the University's "Academic Courses Policy on Course Delivery, Examinations & Assessment of Student Learning" document apply to this class. Section II of the same document describes provisions governing examinations and

grading.

Academic Honesty

The University of Saskatchewan is committed to the highest standards of academic integrity and honesty. Students are expected to be familiar with these standards regarding academic honesty and to uphold the policies of the University in this respect. Students are particularly urged to familiarize themselves with the provisions of the Student Conduct & Appeals section of the University Secretary Website and avoid any behaviour that could potentially result in suspicions of cheating, plagiarism, misrepresentation of facts and/or participation in an offence. Academic dishonesty is a serious offence and can result in suspension or expulsion from the University.

All students should read and be familiar with the regulations on academic student misconduct (http://www.usask.ca/secretariat/student-conduct-appeals/academic-misconduct.php) as well as the policies and procedures regarding non-academic misconduct (http://www.usask.ca/secretariat/student-conduct-appeals/non-academic-misconduct.php). Academic honesty is also defined and described in the Department of Computer Science statement on academic honesty (http://www.cs.usask.ca/students/academic-honesty/index.php).

For more information on what academic integrity means for students see the highlights brochure from the University Secretary website at http://www.usask.ca/secretariat/student-conduct-appeals/forms/IntegrityDefined.pdf.

In BINF200 all assignments and laboratory exercises are to be completed on an individual basis, unless specified otherwise. For the purposes of this class, and this class only, the interpretation of what constitutes "an individual basis" is:

- You may work on certain aspects of an assignment or laboratory as a member of a group. However, other aspects of your work must be done independently. Which aspects must be done independently are described below. In any event, if you worked with a group, it is good practice to clearly identify the other members of the group in your submission, and acknowledge the contributions of other group members (who did or contributed what).

- You may work within a group to verify that the members of the group have understood the problem. You may even confirm your design and results together. However, each of you must perform the implementation phase independently. Failure to perform an independent implementation constitutes academic dishonesty.

- You may work within a group on designing testing strategies and criteria for the selection of test cases. However, each of you must perform the testing phase independently. Failure to perform an independent test constitutes academic dishonesty.

- Each of you must compose your documentation (internal documentation, external documentation, and testing documentation) independently. Failure to compose your own documentation constitutes academic dishonesty.

Please understand that students without the practical experience gained when performing their own implementation, testing, and documentation will be at a significant disadvantage in their examinations when compared to those students who have that experience. It will also be a serious detriment to the former students in subsequent courses, and in an employment situation.

Examinations with Disability Services for Students (DSS)

Students who have disabilities (learning, medical, physical, or mental health) are strongly encouraged to register with Disability Services for Students (DSS) if they have not already done so. Students who suspect they may have disabilities should contact DSS for advice and referrals. In order to access DSS programs and supports, students must follow DSS policy and procedures. For more information, check http://students.usask.ca/health/centres/disability-services-for-students.php, or contact DSS at 306-966-7273 or dss@usask.ca.

Students registered with DSS may request alternative arrangements for midterm and final examinations. Students must arrange such accommodations through DSS by the stated deadlines. Instructors shall provide the examinations for students who are being accommodated by the deadlines established by DSS.

Class Recordings

Students wishing to record any portion of lectures, either audio or video plus audio, must contact the instructor beforehand and obtain permission to do so. Note that Section 5 of the University's "Academic Courses Policy on Class Delivery, Examinations, and Assessment of Student Learning" document lays out policies and requirements associated with such recordings.

Moodle Discussion Forums

Students are encouraged to make use of the moodle discussion forums set up for the class. The professor may on occasion move or remove a posting where he deems such action to be in the best interests of the class.

Other Policies

- All students must be properly registered in order to attend lectures and receive credit for this course.
- Students activities involving University computational facilities are governed by the Department of Computer Science Laboratory Policies. These are available at http://www.cs.usask.ca/support/index.php#all, under "Computer Lab Policies".

Luxuriantly hand-crafted from only the finest HTML tagsby ...
kusalik@cs.usask.ca