

FACULTY of ENGINEERING

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING,

BINF2010

INTRODUCTION TO BIOINFORMATICS

TERM 3, 2019

Course staff

Course Convener/Lecturer in Charge:
Bruno Gaëta (BG) – bgaeta@unsw.edu.au
School of Computer Science and Engineering: K17-401C, x57213
Lecturers
MB- Mike Bain (mike@cse.unsw.edu.au)
RE – Rich Edwards (richard.edwards@unsw.edu.au)
FL – Fabio Luciani (luciani@unsw.edu.au)
MW - Marc Wilkins (marc.wilkins@unsw.edu.au)
SW – Susan Wilson (sue.wilson@unsw.edu.au)
YY – Yunki Yau (Yunki.Yau@health.nsw.gov.au)

Course details

This course is worth 6 units of credit and includes 3 hours of lectures and 2 hours of practicals per week. Labs are not held every week, and some are optional/unassessed. Lectures are held on Mondays 4-6 in Ainsworth 102 (K-J17-102) and on Thursdays 9-10 also in Ainsworth 102. Labs are scheduled after the Thursday lecture, on Thursdays 10-12, in the Strings and Brass labs (K-J17-302 and 305).

Course summary

See course schedule for syllabus

Course aims and learning outcomes

To provide an overview of the field of bioinformatics and its impact on modern biology, together with its industrial and computing context.

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

L01 Define bioinformatics and provide examples of common uses in analysing genome, protein and expression data and in modelling biological systems
L02 Describe some common application areas of bioinformatics and the techniques used therein (genome annotation, rational drug design, medical genomics etc)
L03 Explain the use of and perform common bioinformatics procedures using commonly available software and websites, including: retrieving relevant sequences and structures from databases; identifying ORFs in a DNA sequence; identifying the function of an unknown sequence by similarity searching; identifying the function of an unknown sequence by pattern searches or based on physicochemical properties; creating a multiple sequence alignment and a phylogenetic tree;
L04 Use the UNIX command line to perform basic file management tasks and launch bioinformatics programs
L05 Describe and explain some common bioinformatics algorithms for sequence analysis, genome assembly, phylogenetic tree inference and high-throughput data mining and visualisation
L06 Describe the nature and format of the common data types used in bioinformatics (sequences, structures, expression arrays etc)
L07 Discuss the impact of bioinformatics on modern biology and relevant ethical and social issues
L08 Identify and discuss some engineering challenges common in bioinformatics
L09 Write simple shell scripts to perform semi-automated high-throughput bioinformatics analyses
L10 Use R to visualise complex data and analyse transcriptomics data

Assumed knowledge

- The course assumes an understanding of basic molecular biology including cells, genes, DNA, RNA and proteins, as taught in the prerequisite course BABS1201
- Some knowledge of basic probabilities and statistics (HSC level) is also assumed

Teaching strategies and rationale

- Lectures are structured to emphasise the bioinformatics analysis process. Each lecture module starts with a discussion of the biology context, followed by a discussion of bioinformatics approaches, then of relevant computer science algorithms. Guest lecturers provide illustrations of applications in industry and current research.
- Lectures are supplemented by computer laboratory work and assignments that demonstrate the use of software and interpretation of outputs.

UNSW graduate attributes especially developed in the course include:

- *The skills involved in scholarly enquiry* – students need to research, compare and evaluate different bioinformatics methods as part of the practical work
- *An in-depth engagement with the relevant disciplinary knowledge in its interdisciplinary context* – bioinformatics is presented in the context of its applications to biology, and of the computer science methods it draws on
- *The capacity for analytical and critical thinking and for creative problem-solving* – laboratory work and assignments requires students to solve a range of problems by choosing appropriate bioinformatics methods and applying them
- *The ability to engage in independent and reflective learning* – the final examination especially requires students to reflect and provide a critical synthesis of the course contents
- *A respect for ethical practice and social responsibility* – the course includes discussion and reflection on ethical considerations of the Human Genome Project and other bioinformatics applications
- *The skills of effective communication* – written communication is assessed principally through laboratory reports, and the midsession and final examinations.

Assessment

- Scripting assignment 15%
  Write a shell or Python/Perl script to perform a high-throughput or complex multi-step bioinformatics annotation task. Scripts will be evaluated on their ability to complete the set task accurately on a provided dataset as well as a new dataset of the same type, both in terms of choice of programs to use and on code correctness. (L03, L04, L09)
- Laboratory work 25%
  Marks from quizzes on laboratory work (L03, L04, L10)
- Midsession exam 20%
  An exam in the middle of session, covering the weeks 2-5 material (L01, L02, L05, L06)
- Final exam 40%
  A formal exam in the exam period (L01, L02, L05, L06, L07, L08)

Laboratory work

Due to the high demand for the course, our available lab space may be limited. Most of the labs will therefore be run in "flipped" mode, where the details of the lab are released one week before the lab session. Students are encouraged to attempt the lab on their own before the lab session so they can bring their questions and make the lab time more productive.

Supplementary exams and assessment

A student will be offered a supplementary exam for the midsession or final exam only if they have missed the exam due to a well-documented medical reason. Students who sit a supplementary exam will have their marks calculated in the same way as other students. No supplementary exam will be awarded if the student has already sat the exam. Students must apply for special consideration in order to be considered for a supplementary exam.

Students with a final course mark of 47-49 may also be considered for supplementary assessment if it is allowed by UNSW policy. In this case, if the student passes the supplementary assessment they will receive a final mark of 50.

Course schedule

Lectures are held on Mondays 4-6 in Ainsworth 102 (K-J17-102) and on Thursdays 9-10 also in Ainsworth 102. Labs are scheduled after the Thursday lecture, on Thursdays 10-12, in the Strings and Brass labs (K-J17-302 and 305).

Week	Lec Mon 4-6 (2 hours)	Lec Thu 9-10	Lab Thu 10-12
1: 16/9	Introduction to bioinformatics. Sequence data, evolution and alignment (BG)	Sequence databases, RESTful APIs (BG)	**Introduction to UNIX command line (optional) (BG)**
2: 23/9	Sequence alignment by dynamic programming (BG)	Database searching and multiple sequence alignment (BG)	**Basic sequence analysis using EMBOSS (BG)**
3: 30/9	Genome informatics: applications and assembly algorithms, annotation and genome databases (BG)	Phylogeny inference and UPGMA algorithm (BG)	**Phylogeny and UPGMA (BG)**
4: 7/10	*PUBLIC HOLIDAY*	The human genome project (flipped lecture and discussion) (BG)	**Writing a shell script (optional) (BG)**
5: 14/10	Proteomics (MW)	Proteomics (MW)	**Proteomics (MW)**
6: 21/10	*Midsession exam (BG)*	Case study: a medical application of human genomics (FL)	*No lab*
7: 28/10	Structural bioinformatics (BG)	Case study: characterising a new protein (BG)	**Structural bioinformatics (BG)**
8: 4/11	Transcriptomics (MB)	Transcriptomics (MB)	**Transcriptome analysis using R (MB)**
9: 11/11	Experimental design (SW)	Case study: long read assembly of diploid yeast genomes (RE)	**Data visualisation using R (RE)**
10: 18/11	Bioinformatics in drug design. Bioinformatics in Translational Medicine: Case Studies in Gastroenterological Diseases (YY)	Distributed and application-specific hardware in bioinformatics (BG)	**Workflow management/Galaxy (BG)**

Resources for students

- There is no required textbook however a recommended book for this course is "Digital Code of Life – How Bioinformatics is Revolutionizing Science, Medicine and Business" by Glyn Moody. This is quite an entertaining book that discusses the history and context of bioinformatics. This book does not cover the same materials as the course, but provides background information and context on most sections of the course, in a very readable format.
- Lecture slides, discussion forums, announcements and assignment specs will be made available on the course moodle website (accessible through my.unsw)

Course evaluation and development

This course will be evaluated through the online MyExperience process at the end of session. Individual lecturers may also distribute surveys on their own teaching. Feedback from these surveys is taken seriously and you are encouraged to respond. Feedback from 2018 indicated some issues with the format of lecture slides which will be addressed by better listing the learning outcomes. The scripting assignment is also being revised in response to student feedback.

Student Conduct

The Student Code of Conduct (Information at https://student.unsw.edu.au/conduct, Policy at https://www.gs.unsw.edu.au/policy/documents/studentcodepolicy.pdf) sets out what the University expects from students as members of the UNSW community. As well as the learning, teaching and research environment, the University aims to provide an environment that enables students to achieve their full potential and to provide an experience consistent with the University's values and guiding principles. A condition of enrolment is that students *inform themselves* of the University's rules and policies affecting them, and conduct themselves accordingly.
In particular, students have the responsibility to observe standards of equity and respect in dealing with every member of the University community. This applies to all activities on UNSW premises and all external activities related to study and research. This includes behaviour in person as well as behaviour on social media, for example Facebook groups set up for the purpose of discussing UNSW courses or course work. Behaviour that is considered in breach of the Student Code Policy as discriminatory, sexually inappropriate, bullying, harassing, invading another's privacy or causing any person to fear for their personal safety is serious misconduct and can lead to severe penalties, including suspension or exclusion from UNSW.
If you have any concerns, you may raise them with your lecturer, or approach the School Ethics Officer (ethics-officer@cse.unsw.edu.au), Grievance Officer (grievance-officer@cse.unsw.edu.au), or one of the student representatives.
Plagiarism is defined as using the words or ideas of others and presenting them as your own. UNSW and CSE treat plagiarism as academic misconduct, which means that it carries penalties as severe as being excluded from further study at UNSW. There are several on-line sources to help you understand what plagiarism is and how it is dealt with at UNSW:

Plagiarism and Academic Integrity: https://student.unsw.edu.au/plagiarism

UNSW Plagiarism Procedure: https://www.gs.unsw.edu.au/policy/documents/plagiarismprocedure.pdf

Make sure that you read and understand these. Ignorance is not accepted as an excuse for plagiarism. In particular, you are also responsible that your assignment files are not accessible by anyone but you by setting the correct permissions in your CSE directory and code repository, if using. Note also that plagiarism includes paying or asking another person to do a piece of work for you and then submitting it as your own work.

UNSW has an ongoing commitment to fostering a culture of learning informed by academic integrity. All UNSW staff and students have a responsibility to adhere to this principle of academic integrity. Plagiarism undermines academic integrity and is not tolerated at UNSW. Plagiarism at UNSW is defined as using the words or ideas of others and passing them off as your own.

If you haven't done so yet, please take the time to read the full text of UNSW's policy regarding academic honesty and plagiarism (https://student.unsw.edu.au/plagiarism)
The pages below describe the policies and procedures in more detail:

Student Code Policy:
https://www.gs.unsw.edu.au/policy/documents/studentcodepolicy.pdf

Student Misconduct Procedure:
https://www.gs.unsw.edu.au/policy/documents/studentmisconductprocedures.pdf

Plagiarism Policy Statement:
https://www.gs.unsw.edu.au/policy/documents/plagiarismpolicy.pdf

Plagiarism Procedure:
https://www.gs.unsw.edu.au/policy/documents/plagiarismprocedure.pdf

You should also read the following page, which describes your rights and responsibilities in the CSE context:
https://www.engineering.unsw.edu.au/computer-science-engineering/about-us/organisational-structure/student-services/policies/essential-advice-for-cse-students