

FACULTY of ENGINEERING

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

BINF3010/9010

BIOINFORMATICS METHODS AND APPLICATIONS

SESSION 1, 2017

Course staff

Course Convener:
Bruno Gaëta (BG) – bgaeta@unsw.edu.au
School of Computer Science and Engineering: K17-401C, x57213
Consultation times – By appointment

Lecturers:
Mike Bain (MB), School of Computer Science and Engineering – mike@cse.unsw.edu.au
Paul Curmi (PC), School of Physics – p.curmi@unsw.edu.au
Ruiting Lan (RL), School of Biotechnology and Biomolecular Sciences – r.lan@unsw.edu.au
Mark Raftery (MR), Biomolecular Research Facility UNSW – m.raftery@unsw.edu.au
Marc Wilkins (MW), School of Biotechnology and Biomolecular Sciences – m.wilkins@unsw.edu.au
Sue Wilson (SW), Stats Central UNSW and ANU – Sue.Wilson@anu.edu.au

Guest Lecturers
Warren Kaplan (WK), Garvan Institute of Medical Research

Course details

This course is worth 6 units of credit. Lectures are held on:

- Tuesday 10-11 in Australian School Business 232 (K-E12-232)
- Thursday 14-15 in Australian School Business 232 (K-E12-232)
- Thursday 17-18 in Australian School Business 232 (K-E12-232)

Laboratories (shown in **Bold** on the course schedule) will be held on Tuesdays 11-14 in Tabla Lab K17 (CSE) G07B (K-K17-G07B)with overflow in Drum Lab K17 (CSE) B08 (K-K17-B08)

Important note: **most labs will require the use of computers running the Linux operating system, and it is recommended that students who are not familiar with this environment attend the "Lab 0" session held during O-week and partner with a student familiar with Linux in the labs.**

Some weeks (shown in *italics* on the course schedule) the laboratory will be replaced by a 2 hour extra lecture/tutorial or exam in Quadrangle G044 (K-E15-G044).

Course aims

Bioinformatics (the use of computing methods for the management and analysis of molecular biology data) has become an integral component of biomolecular sciences, especially genomics and proteomics. This course focuses on the practical use of bioinformatics methods and resources for the analysis of DNA and protein sequences and structures, as well as results from microarray and proteomics, with emphasis on their evolutionary underpinnings and statistical foundations.

General course aims

- Provide an introduction to commonly used bioinformatics resources, methods and software, with an emphasis on their use, capabilities and limitations
- Provide an introduction to bioinformatics as an application area of mathematical and computational sciences

BINF3010 is a core course in the BE (Bioinformatics) program, and an elective in Science programs.
BINF9010 is a core course in the MIT (bioinformatics) program, and an elective in a number of other graduate programs including Biomedical Engineering and Biostatistics.
BINF3010/9010 does not cover bioinformatics algorithms in depth (these are covered in BINF3020/9020). The course focuses on the intelligent application of common bioinformatics methods to assist in biological discovery and is primarily targeted at students with a biology background. For students with a computing background the course provides an introduction to bioinformatics and its biological context.

Student learning outcomes
At the end of this course, students should be able to:
- LO1: Outline the fundamental biology concepts that provide the context for bioinformatics, including sequence, structure and function as they relate to biological information macromolecules and molecular evolution
- LO2: Describe some of the computational representations commonly used in bioinformatics including sequences (strings), alignments and three-dimensional structure coordinates
- LO3: List common bioinformatics programs, resources and protocols used to analyse DNA and protein sequences, outline their strengths and limitations
- LO4: Research, identify, select and use bioinformatics tools and databases to analyse biological sequences, structures and functional genomics data.
- LO5: Apply statistical thinking to the design and analysis of high-throughput molecular biology experiments, and perform a range of data processing and analysis tasks using the R environment
- LO6: Interpret and analyse data generated by proteomics experiments using bioinformatics
- LO7: Discuss the experimental basis of protein structure determination and discuss, compare and apply a range of computational methods for analysing and predicting protein tertiary structure

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 and final examination
- *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 require 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 midsession and final examinations require students to reflect and provide a critical synthesis of the course contents
- *The skills required for collaborative and multidisciplinary work* – the laboratory exercises are to be carried out in teams of mixed student background
- *The skills of effective communication* – written communication is assessed principally through laboratory reports and the final examination. Effective communication between students of different backgrounds is also necessary for carrying out the laboratory assignments.

The rationale behind your approach to learning and teaching

Bioinformatics now pervades biological research, and new methods and technologies are constantly developed. This course is aimed at teaching bioinformatics from a *user*'s perspective (as opposed to that of a *developer*), to emphasise the use of bioinformatics to assist in biological discovery. Since bioinformatics constantly evolves the goal is not to teach the use of specific tools and methods but to focus on *principles*, *limitations* and *assumptions* of common approaches to provide the means for students to research and evaluate new methods and apply them intelligently to produce meaningful results.

Teaching approach and strategies

- In order to establish the link between the topics being covered and the current state of the art in bioinformatics research, each topic is presented by a lecturer active in research in that area or its applications
- Hands-on practical computer laboratories require the students to use a range of bioinformatics methods described in the lectures and reflect on the results they obtain
- Group assignments assess research and teamwork skills
- Bioinformatics is a new discipline that changes quickly, and it is important for graduates to be able to keep in touch with the state of the art in bioinformatics methods, rather than just learn about a standard set of tools that will soon be obsolete. The course therefore emphasises fundamental principles and requires the students to demonstrate the ability to research and evaluate new methods.

Assessment: BINF3010

Computer laboratory work – 30%

Computer laboratories will emphasise the use of common bioinformatics applications. Most labs will require a written lab report typically submitted one week after the lab. These assignments evaluate your ability to use bioinformatics software and interpret results in the light of the theoretical background discussed in the lectures. Submission method will be advised: online submission, when available, will be through the course website. Submission of printed reports will be directly to the lecturer during the lecture period. Laboratory reports will not be accepted after their set submission deadline. Learning outcomes assessed: LO3, LO4, LO5, LO6

Important notes:

- **Most labs will require the use of computers running the Linux operating system, and it is recommended that students who are not familiar with this environment attend the "Lab 0" session held during O-week or at** https://wiki.csesoc.unsw.edu.au/Lab0 **and partner with a student familiar with Linux in the labs.**
- **The Week 12 lab requires the use of the R statistics/scripting language. If you have not used R before, make sure to read Chapter 1 of the Manual "An Introduction to R" (including working through the introductory session given in Appendix A) that is available at http://www.r-project.org**

Protein modelling assignment – 10%

An assignment requiring you to predict the 3D structure of a protein given its sequence. Predicted structures will be compared to the "real" structure in a laboratory session. Learning outcomes assessed: LO4, LO7

Midsession examination – 25%

An examination held in week 7 covering the weeks 1-5 lecture and lab material. Learning outcomes assessed: LO1, LO2, LO3

Final examination – 35%

A 2 hour examination covering the weeks 5-12 lecture material (and some material covered in the labs) taking place in the exam period. Learning outcomes assessed: LO1, LO2, LO3, LO4, LO5, LO6, LO7

Assessment: BINF9010

Biology quizzes – 2%

Short quizzes to remind you of biology fundamentals relevant to the bioinformatics methods being studied. Quizzes will be administered online through the course website. Learning outcomes assessed: LO1

Computer laboratory work – 30%

Computer laboratories will emphasise the use of common bioinformatics applications. Most labs will require a written lab report typically submitted one week after the lab. These assignments evaluate your ability to use bioinformatics software and interpret results in the light of the theoretical background discussed in the lectures. Submission method will be advised: online submission, when available, will be through the course website. Submission of printed reports will be directly to the lecturer during the lecture period. Laboratory reports will not be accepted after their set submission deadline. Learning outcomes assessed: LO3, LO4, LO5, LO6

Important notes:

- **Most labs will require the use of computers running the Linux operating system, and it is recommended that students who are not familiar with this environment attend the "Lab 0" session held during O-week and partner with a student familiar with Linux in the labs.**
- **The Week 12 lab requires the use of the R statistics/scripting language. If you have not used R before, make sure to read Chapter 1 of the Manual "An Introduction to R" (including working through the introductory session given in Appendix A) that is available at http://www.r-project.org**

Protein modelling assignment – 10%

An assignment requiring you to predict the 3D structure of a protein given its sequence. Predicted structures will be compared to the "real" structure in a laboratory session. Learning outcomes assessed: LO4, LO7

Midsession examination – 24%

An examination held in week 7 covering the weeks 1-5 lecture and lab material. Learning outcomes assessed: LO1, LO2, LO3

Final examination – 34%

A 2 hour examination covering the weeks 5-12 lecture material (and some material covered in the labs) taking place in the exam period. Learning outcomes assessed: LO1, LO2, LO3, LO4, LO5, LO6, LO7

Academic honesty and plagiarism

What is Plagiarism?

Plagiarism is the presentation of the thoughts or work of another as one's own.*
Examples include:

- direct duplication of the thoughts or work of another, including by copying material, ideas or concepts from a book, article, report or other written document (whether published or unpublished), composition, artwork, design, drawing, circuitry, computer program or software, web site, Internet, other electronic resource, or another person's assignment without appropriate acknowledgement;
- paraphrasing another person's work with very minor changes keeping the meaning, form and/or progression of ideas of the original;
- piecing together sections of the work of others into a new whole;
- presenting an assessment item as independent work when it has been produced in whole or part in collusion with other people, for example, another student or a tutor; and
- claiming credit for a proportion a work contributed to a group assessment item that is greater than that actually contributed.†

For the purposes of this policy, submitting an assessment item that has already been submitted for academic credit elsewhere may be considered plagiarism.

Knowingly permitting your work to be copied by another student may also be considered to be plagiarism.

Note that an assessment item produced in oral, not written, form, or involving live presentation, may similarly contain plagiarised material.

The inclusion of the thoughts or work of another with attribution appropriate to the academic discipline does *not* amount to plagiarism.

The Learning Centre website is main repository for resources for staff and students on plagiarism and academic honesty. These resources can be located via:

www.lc.unsw.edu.au/plagiarism

The Learning Centre also provides substantial educational written materials, workshops, and tutorials to aid students, for example, in:

- correct referencing practices;
- paraphrasing, summarising, essay writing, and time management;
- appropriate use of, and attribution for, a range of materials including text, images, formulae and concepts.

Individual assistance is available on request from The Learning Centre.

Students are also reminded that careful time management is an important part of study and one of the identified causes of plagiarism is poor time management. Students should allow sufficient time for research, drafting, and the proper referencing of sources in preparing all assessment items.

* Based on that proposed to the University of Newcastle by the St James Ethics Centre. Used with kind permission from the University of Newcastle
† Adapted with kind permission from the University of Melbourne.

Course schedule

Week	Starting	LEC TUE 10-11	TUE 11-13 (**LAB** or */Lecture/Tutorial*)	THU 14-15	THU 17-18
1	27/02/17	Sequence analysis (BG)		Sequence analysis (BG)	Sequence analysis (BG)
2	6/03/17	Sequence analysis (BG)	**Databases** (BG)	Sequence analysis (BG)	Sequence analysis (BG)
3	13/03/17	Sequence analysis (BG)	**Alignment** (BG)	Sequence analysis (BG)	Sequence analysis (BG)
4	20/03/17	Genome informatics (RL)	**Similarity searching** (BG)	Genome informatics (RL)	Genome informatics (RL)
5	27/03/17	Proteomics (MR)	**Phylogeny and MSA** (BG)	Proteomics (MR)	PTMs, PPIs and glycomics (MW)
6	03/04/17	PTMs, PPIs and glycomics (MW)	**DNA Sequencing** (BG)	PTMs, PPIs and glycomics (MW)	PTMs, PPIs and glycomics (MW)
7	10/04/17	Systems Biology (MB)	*Midsession exam*	*TBC*	Structural bioinformatics (PC)
	17/04/17	**BREAK**			
8	24/04/17	ANZAC DAY (no lecture)	ANZAC DAY (no lab)	*Informatics for population scale genome sequencing* (WK)	Structural bioinformatics (PC)
9	1/05/17	Structural bioinformatics (PC)	**Proteomics** (MW)	Structural bioinformatics (PC)	Structural bioinformatics (PC)
10	8/05/17	Structural bioinformatics (PC)	*Structural bioinformatics (PC) NOTE: room to be confirmed*	Structural bioinformatics (PC)	Structural bioinformatics (PC)
11	15/05/17	Statistical analysis of high-throughput data (SW)	**PTMs** (MW)	Statistical analysis of high-throughput data (SW)	Statistical analysis of high-throughput data (SW)
12	22/05/17	Statistical analysis of high-throughput data (SW)	**R** (SW)	Statistical analysis of high-throughput data (SW)	Statistical analysis of high-throughput data (SW)
13	29/05/17		**Modelling evaluation** (BG)		

Resources for students

- There is no textbook for this course. Individual lecturers will provide lists of reference books and articles.
- Readings and discussion boards will be made available on the course website, which can be accessed through the moodle website linked from: https://moodle.telt.unsw.edu.au/course/view.php?id=24678
- A number of bioinformatics textbooks are available through the UNSW Library for reference reading. One starting point for assistance is: info.library.unsw.edu.au/web/services/services.html

Course evaluation and development

- Feedback on this course and on individual lecturers will be gathered through a survey at the end of session, as part of the MyExperience process. Feedback from this survey is the basis for improving the course in subsequent years.
- ***Feedback from 2016:***
  1. Difficulties with the required biology knowledge.

2. More step-by-step instructions in the labs
3. A way to ask questions about the labs without making a public forum post

Response to this feedback:

1. This course has a very diverse cohort of students and finding a way to cater to everybody without dumbing down the material is not easy. **Remember that the assumed knowledge for this course is a decent grounding in molecular biology, equivalent to completion of BIOC2201.** This is specified in the BINF3010 prerequisites but unfortunately we cannot specify an equivalent prerequisite for BINF9010, as UNSW does not have a postgraduate introductory molecular biology course. **BINF9010 students must ensure they have the required knowledge or they will have difficulties in the course**. Also remember that working in interdisciplinary groups in laboratories is encouraged so that students from different backgrounds can complement and learn from each other.
2. To some extent the instructions in the labs are not too specific on purpose. We do want students to think about what they are doing and be able to apply what they are doing to other data and different analyses rather than blindly following a recipe step by step. But the labs are revised every year to try to improve clarity.
3. The policy of accepting only questions asked on the forum, and only before a reasonable deadline, is not only so that lecturers do not have to answer the same question many times, but also because it is unfair for some students to get answers that other students do not get to see. A public forum also allows students to assist each other, which offers additional learning benefits.

Other matters

- Occupational Health and Safety: students are reminded of the university's OHS policies and recommendations, which are accessible at
  www.riskman.unsw.edu.au/ohs/ohs.shtml
  Information specific to OHS in the school of CSE, and especially of ergonomics issues related to use of computers can be accessed at http://www.cse.unsw.edu.au/~ohs/

- Equity and diversity: note that students who have a disability that requires some adjustment in their learning and teaching environment are encouraged to discuss their study needs with the course convener prior to, or at the commencement of the course, or with the Equity Officer (Disability) in the Equity and Diversity Unit (9385 4734). Information for students with disabilities is available at:
  www.equity.unsw.edu.au/disabil.html