

FACULTY of ENGINEERING

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

BINF3010/9010

APPLIED BIOINFORMATICS

TERM 2, 2019

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:
Susan Corley, School of Biotechnology and Biomolecular Sciences – s.corley@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
Fatemeh Vafaee, School of Biotechnology and Biomolecular Sciences - f.vafaee@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@unsw.edu.au

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

Course summary and 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 BINF6111 and BINF6112). 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:

- 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
- Use the UNIX shell to manage and analyse biological sequence and functional genomics data
- Research, identify, select and use bioinformatics tools and databases to analyse biological sequences, structures and functional genomics data
- 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
- Outline the strengths and limitations of the main approaches used in sequence and structural bioinformatics, functional genomics and systems biology.
- 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
- Interpret and analyze data generated by proteomics experiments using bioinformatics

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.

Rationale and teaching strategies

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.

- 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
- 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.

Course schedule

Week	Starting	Lec 1 Tue 2-4	Lab Tue 4-6	Lec 2 Thu 2-4
1	Jun 3	Sequence Analysis **Gaeta**	**Intro lab** (optional) *Intro tute* (optional) **Gaeta**	Sequence Analysis **Gaeta**
2	Jun 10	Sequence Analysis **Gaeta**	**Alignment and databases** Gaeta	Sequence Analysis **Gaeta**
3	Jun 17	Genome Informatics **Lan**	**MSA and phylogeny** Gaeta	RNA-Seq **Corley**
4	Jun 24	*To be confirmed*	**Genome Assembly** Gaeta	Statistics **Wilson**
5	Jul 1	Statistics **Wilson**	*Midterm exam*	Statistics **Wilson**
6	Jul 8	Proteomics **Raftery**	**Visualisation and analysis using R** Gaeta	Structural Bioinformatics **Curmi**
7	Jul 15	Structural Bioinformatics **Curmi**	**RNA-Seq** Corley	Structural Bioinformatics **Curmi**
8	Jul 22	PTMs and Proteogenomics **Wilkins**	**Proteomics** Wilkins	Structural Bioinformatics **Curmi**
9	Jul 29	Structural Bioinformatics **Curmi**	*No lab*	Protein Interactions and Networks **Wilkins**
10	Aug 5	ncRNA analysis **Vafaee**	**Structure Comparison** Gaeta	Industry lecture **TBC**

Class times and locations

This course is worth 6 units of credit. Lectures are held on:
Tue 14:00 - 16:00 Red Centre Central Wing M032 (K-H13-M032)
Thu 14:00 - 16:00 Old Main Building 230 (K-K15-230)

Laboratories (shown in **Bold** on the course schedule) will be held on Tuesdays 4-6pm in Bongo Lab K17 (CSE) G07B (K-K17-G07A). These are meant to be for consultation and discussion, and it is recommended that you attempt the labs *before* these sessions so you can bring your questions and issues.
The midterm exam will be scheduled during a lab slot at a location to be advised closer to the date.

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 attempt the optional Intro lab scheduled in week 1.

The into tutorial scheduled for week 1 at the same time as the intro lab is specifically targeted at BINF9010 students who need a refresher on the biology foundations required for the course. Students who have completed the BINF3010 prerequisite BIOC2201 or equivalent will not need this optional tutorial.

Assessment: BINF3010

Computer laboratory work – 30%

Computer laboratories will emphasise the use of common bioinformatics applications. Most labs will be assessed using an online moodle quiz. Laboratory quizzes will not be accepted after their set submission deadline.

The labs evaluate your ability to use bioinformatics software and interpret results in the light of the theoretical background discussed in the lectures. The lab room is not large enough to sit all the enrolled students, so the labs have been structured so that they can be completed on your own on your own computer as much as possible. The lab sessions are meant to be for consultation and discussion of any difficulties you have encountered doing the lab. It is recommended that you attempt to complete the lab by yourself before the corresponding lab session so you can bring your questions for the lab session.

Important notes:

- **Most labs will require the use of computers running the Linux operating system. It is recommended that students with no familiarity with Linux go through the optional intro to UNIX lab, and partner with a student familiar with Linux in the labs.**
- **Some labs require 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.

Midsession examination – 20%

An examination held during the term covering the first part of the lecture and lab material.

Final examination – 40%

A 2-hour examination covering the second part of the lecture material (and some material covered in the labs) taking place in the exam period.

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.

Computer laboratory work – 28%

Computer laboratories will emphasise the use of common bioinformatics applications. Most labs will be assessed using an online moodle quiz. Laboratory quizzes will not be accepted after their set submission deadline.

The labs evaluate your ability to use bioinformatics software and interpret results in the light of the theoretical background discussed in the lectures. The lab room is not large enough to sit all the enrolled students, so the labs have been structured so that they can be completed on your own on your own computer as much as possible. The lab sessions are meant to be for consultation and discussion of any difficulties you have encountered doing the lab. It is recommended that you attempt to complete the lab by yourself before the corresponding lab session so you can bring your questions for the lab session.

Important notes:

- **Most labs will require the use of computers running the Linux operating system. It is recommended that students with no familiarity with Linux go through the optional intro to UNIX lab, and partner with a student familiar with Linux in the labs.**
- **Some labs require 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.

Midsession examination – 20%

An examination held during the term covering the first part of the lecture and lab material.

Final examination – 40%

A 2 hour examination covering the second part of the lecture material (and some material covered in the labs) taking place in the exam period.

Relevant resources

- 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 Moodle.
- 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

Student Conduct

The **Student Code of Conduct** (https://student.unsw.edu.au/conduct) 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)
- The Grievance Officer (grievance-officer@cse.unsw.edu.au)
- One of the student representatives (stureps.ug@cse.unsw.edu.au)

Plagiarism is defined as using the words or ideas of others and presenting them as your own (https://student.unsw.edu.au/plagiarism). 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.

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:

- Essential Advice for CSE Students: https://www.engineering.unsw.edu.au/computer-science-engineering/about-us/organisational-structure/student-services/policies/essential-advice-for-cse-students

Special Consideration

If your work in this course is affected by unforeseen adverse circumstances, you should apply for Special Consideration through MyUNSW, including documentation on how you have been affected. If your request is reasonable and your work has clearly been impacted, then

- for an assignment, you may be granted an extension
- for the Final Exam, you may be offered a Supplementary Exam

Note the use of the word "may". None of the above is guaranteed. It depends on you making a convincing case that the circumstances have clearly impacted your ability to work.

If you are registered with Disability Services, please forward your documentation to Bruno Gaeta within the first two weeks of semester.

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.
-

Main feedback from 2018:

1. Sample exam questions would help with revision
2. Assumed biology background was too deep for some students
3. Issues with lecture recordings
4. Protein structure prediction assignment is challenging – more guidance would be useful

Response to this feedback:

1. Different lecturers have quite different styles of exam questions, and in some cases the exam questions are quite varied in style. We will provide some sample exam questions however students should be aware that these will represent only some examples and the actual questions may be quite different.
2. The assumed biology knowledge is made very clear in the course outline and handbook description. BINF9010 students with limited biology experience should really not take the course if they are not ready to undertake significant additional study on their own.
3. Unfortunately, we have no control over lecture recordings, and it is always risky to rely on them. We encourage students to attend the lecture in person if possible as they gain to benefit by asking questions as well.
4. The protein structure assignment is indeed difficult and is meant to reflect the real challenges of protein structure prediction. This is taken into account when the assignment is marked, and many students do very well in it. One difficulty is that we often do not have access to a true “unknown” protein and have to rely on a published protein, asking the class not to use the published structure in their answer, which precludes the use of fully automated methods. It is worth noting than in previous years when a true “unknown” structure was used the marks were generally lower as many students put too much trust in automated methods.

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