

FACULTY of ENGINEERING, FACULTY of SCIENCE

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING, SCHOOL OF BIOTECHNOLOGY AND BIOMOLECULAR SCIENCES

BINF3001

BIOINFORMATICS 3

SESSION 2, 2008

Course Convener:

Bruno Gaëta (BG) – <u>bgaeta@unsw.edu.au</u> School of Computer Science and Engineering: K17-401C, x57213 School of Biotechnology and Biomolecular Science: Biosciences 231B1, x52056

Course lecturers

- MB Mike Bain (mike@cse.unsw.edu.au)
- PC Paul Curmi (<u>p.curmi@unsw.edu.au</u>)
- WK Warren Kaplan (<u>w.kaplan@garvan.org.au</u>)
- MR Mark Raftery (<u>m.raftery@unsw.edu.au</u>)
- MT Mark Tanaka (<u>m.tanaka@unsw.edu.au</u>)
- MW Marc Wilkins (<u>m.wilkins@unsw.edu.au</u>)
- SW Sue Wilson
- AW Alan Wilton (<u>a.wilton@unsw.edu.au</u>)
- RW Robert Womersley (<u>r.womersley@UNSW.EDU.AU</u>)

Course details

This course is worth 6 units of credit and includes 3 hours of lectures and either 3 hours of computer laboratory work or 2 hours of tutorials per week.

Course aims and learning outcomes

BINF3001 builds on the foundations laid in BINF1001 and BINF2001, and covers a range of advanced topics and current research areas in bioinformatics. The course demonstrates the application of computational, mathematical and statistical methods to problems in modern life sciences, through examples in nucleotide and protein structure, function and expression.

At the end of this course, students should be able to:

- 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
- Apply statistical methods to the design and analysis of DNA microarray experiments, and perform a range of data processing and analysis tasks using the R language
- Define hidden Markov models, describe their associated algorithms, and discuss their application to biological sequence analysis
- Describe and compare algorithms for identifying conserved patterns in sets of DNA or protein sequences
- Research and explain the use of statistical models to represent causal processes and their application in computational biology
- Outline the foundations of statistical methods used in gene mapping and population genetics
- Interpret and analyse data generated by proteomics experiments using bioinformatics
- Discuss and present the range of open source development environments available to bioinformatics software developers and use at least one of these environments in their own software development projects.

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
- Lectures are supplemented by computer laboratory work and assignments that demonstrate the use of software and interpretation of outputs, and programming assignments that require students to implement algorithms discussed in lectures and tutorials
- Group assignments and class presentations assess research, teamwork and communication skills

UNSW graduate attributes especially developed in the course include:

- *The skills involved in scholarly enquiry* students are required to research and compare different approaches and methods as part of assignments
- An in-depth engagement with the relevant disciplinary knowledge in its interdisciplinary context the course covers advanced applications of bioinformatics in a range of biological domains
- *The capacity for analytical and critical thinking and for creative problem-solving* several assessment tasks require students to analyse a problem, evaluate different approaches and develop a potential solution
- *The ability to engage in independent and reflective learning* assignments and examinations require students to provide a critical synthesis of course contents
- Information literacy the skills to appropriately locate, evaluate and use relevant information several assessment tasks, and especially the ones on causal modeling, open source development environments and

population genetics, require the students to search through the scientific literature as well as software documentation and present a synthesis of their research

- A capacity to contribute to, and work within, the international community the open source assignment especially requires students to examine large, international collaborative projects and how they can contribute to them
- *The skills required for collaborative and multidisciplinary work* a number of assignments are carried out in groups
- The skills of effective communication several assessment tasks require written reports or oral presentation

Resources for students

- There is no textbook for this course. Individual lecturers will recommend resources and textbooks for their section of the course.
- Lecture slides, discussion forums, announcements and assignment specs will be made available on the course website: <u>http://www.cse.unsw.edu.au/~bi3001</u>
- The enrolment key for the website is DURBIN

Course evaluation and development

This course will be evaluated through the online CATEI process at the end of the 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 previous years suggested that the workload for this course was heavy relative to other courses. In response to this feedback the continuous assessment load has been reduced this year.

Assessment

- Protein structure programming assignment 15%
 - Requires implementing an algorithm commonly used in protein structure modeling, to demonstrate understanding of the geometry involved in dealing with 3-dimensional structures and of the format of PDB entries. This assignment is marked entirely on the ability of the resulting software to provide a correct output with both provided and new test cases.
- Open source environments presentations 15%
 - A group assignment where teams use a selected open source development environment (eg BioPerl, BioJava) and present it to the class to discuss how this resource can assist in bioinformatics software development. Marked on content and quality of presentations, including a peer assessment component.
- Microarray assignment 10% An assignment involving using the R statistical language to analyse microarray data
- Causal modeling assignment 10% Requires students to work in teams to research a causal modeling method and present a summary of their findings to the class
- Other practical/tutorial work and assignments 10%
 - Reports from the computer laboratory or tutorial work
- Mid-session exam 20%
 - Covers all the material up to week 6
- Final exam 20%
 - Covers all the material not covered in the mid-session exam

Course schedule

The Friday lab/tute is either a 2-hour lecture/tutorial held in Mathews 310 (shown in *italics*) or a 3-hour lab held in Bell lab (Mech Eng Undercroft) (shown in **bold**).

Woolz	Storting	Loc1 (Wod 1nm)	Loo? (Thu Qam)	Log3 (Fri 3nm)	Lab (Ball lab) (Eri 11
WCCK	Starting	Mathews 207	Ovedrengle 1040	Maryan Draym LC2	Lab (Bell lab) ($\Gamma \Pi \Pi \Pi$ -
		(K, E22, 207)	Quadrangle 1049	Wolven Blown LG2	2) of tutorial (Mai 510)
		(K-F23-307)	(K-E15-1049)	(K-C20-LG2)	(Fri 11-1)
1	28/07/08	Protein structure	Protein structure (PC)	Protein structure (PC)	Optimisation (RW)
		(PC)			
2	4/08/08	Protein structure	Protein structure (PC)	Protein structure (PC)	Protein structure (WK)
		(PC)			
3	11/08/08	Protein structure	Protein structure (PC)	Protein structure (PC)	Protein structure
		(PC)			assignment (BG)
4	18/08/08	Protein structure	Proteomics (MR)	Proteomics (MR)	Proteomics (MW)
		(PC)			
5	25/08/08	Population genetics	Population genetics	Population genetics	Statistical genetics
		(MT)	(MT)	(MT)	(AW)
6	1/09/08	Population genetics	HMMs (BG)	HMMs (BG)	Population genetics
		(MT)			(MT)
7	8/09/08	HMMs (BG)	HMMs (BG)	HMMs (BG)	MIDSESSION EXAM
-					
8	15/09/08	Statistical analysis	Statistical analysis of	Statistical analysis of	R (SW)
		of microarrays (SW)	microarrays (SW)	microarrays (SW)	
9	22/09/08	Statistical analysis	Statistical analysis of	Statistical analysis of	R (SW)
		of microarrays (SW)	microarrays (SW)	microarrays (SW)	
10	6/10/08	Causal modeling	Causal modeling	Causal modeling	Causal modeling (MB)
		(MB)	(MB)	(MB)	
11	13/10/08	Causal modeling	Causal modeling	Causal modeling	Open source
		(MB)	(MB)	(MB)	environments (BG)
12	20/10/08	PTMs and	PTMs and glycomics	PTMs and glycomics	PTMs and glycomics
		glycomics (MW)	(MW)	(MW)	(MW)

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.

Other matters

Please refer to the "yellow form" (<u>http://www.cse.unsw.edu.au/~studentoffice/policies/yellowform.html</u>) for important information and policies regarding the use of CSE facilities, OHS, special consideration and supplementary assessment, accessibility etc