

# Bioinf 702: Comparative Bioinformatics

## Course Guide 2017

Alexei Drummond

March 6, 2017

### 1 Introduction

This course sets out to give a foundation in some of the main areas of modern bioinformatics. It will build on your previous knowledge of computer science, biology and the basics of bioinformatics. We will attempt to cover a large amount of material quite quickly so it is important that you understand the material as the course progresses — if you do not understand something, do not hesitate to ask.

The class size is small which means there is plenty of opportunity for discussion and to ask questions.

Your feedback is very important to us to help us improve and refine the course, so please let us know what works and what doesn't or if there is anything you would like to cover that we have not.

Enjoy the course!

Professor Alexei Drummond, Course co-ordinator.

### 2 Lectures and Labs

Lectures are scheduled for Monday, Tuesday, Wednesday 11:00-12:00, McGregor 2 (106-113, Biology building)

Labs are scheduled for Friday 1-4pm, 302-150 (Science centre) every second week. Depending on the lecturer, this time may or may not be used. The last 3 labs will also be located in the Protein Structure labs (to be advised).

### 3 Lecturers

- Alexei Drummond (co-ordinator), Room 303S.479 (Computer Science), Phone 88298, Email alexei@cs.auckland.ac.nz
- Tim G. Vaughan, Room 303S.467 (Computer Science), Email tgvaughan@gmail.com
- Verne Lee, Room 110.424 (Thomas Building), Phone 83888, Email t.lee@auckland.ac.nz

Lecturers operate an open door policy. Feel free to drop by or make an appointment.

## 4 Learning outcomes

After successfully completing this course, you will be able to:

- Be familiar with the important features of proteins and protein structures
- Understand and make critical judgements about methods of protein sequence analysis and their strengths and weaknesses compared to methods based on protein structure.
- Be able to draw biological inference about protein function using a variety of sequence database searching and predictive techniques.
- Be familiar with the principles of the main experimental methods for the determination of protein structures, their respective strengths and weaknesses, and what errors may exist in the data.
- Be familiar with the organization and use of protein structure databases.
- Understand the methods of protein structure prediction and their strengths and weaknesses
- Understand the classification of proteins by structure and methods of structural alignment used in that classification.
- Describe the application of computational methods to the analysis of phylogenetic relationships and to population genetic analysis using the coalescent.
- Describe the application of computational methods to the inference of pairwise and multiple sequence alignments and, hence, of positional homology
- Write scripts in Perl, Python or Java that implement some of the principle algorithms used in comparative bioinformatics and phylogenetics.
- Describe the design and operation of Markov models and understand how they may be applied in sequence alignment, and for modelling sequence evolution and phylogenetic relationships.
- Critically evaluate the appropriate techniques and methods to perform comparative analysis of biological data

## 5 Assessment

60% of the course will be internally assessed. The assessment will comprise 6 assignments/student seminars of equal value. Assignments will include data analysis, some short programming problems and reading and presenting research articles. Assignments will be evenly spaced across the semester.

40% of the course will be assessed in the final 2 hour exam that will cover all the major concepts in the course.

To pass the course, students must pass both the internal assessment and the exam.

## 6 Resources

All relevant readings will be posted on Canvas (<http://canvas.auckland.ac.nz>) and available from the University Library (<http://www.library.auckland.ac.nz/>).

## 7 Lecture outline

The outline below gives a guide of course content and structure. Specific lecture order and topics covered may change, so please keep up to date by attending lectures and checking Canvas. If you miss a lecture, please contact the relevant lecturer to find out how to catch-up. There will be an initial introductory lecture and a final revision lecture to make 36 lectures in total.

### 7.1 Sequence alignment and homology (3 Lectures TGV)

A review of the idea of homology. A tour of dynamic programming algorithms used for alignment including an introduction to Hidden Markov Models.

### 7.2 Basic concepts in phylogenetics (3 Lectures AJD)

How are trees used and interpreted in bioinformatics, how are they constructed and how are they modelled.

### 7.3 Bayesian inference and phylogenetic inference (10 Lectures AJD + TGV)

Probability theory, Models of sequence evolution, maximum likelihood phylogenetic inference, Bayesian phylogenetic inference, BEAST2, Applications.

### 7.4 Weeks 7-12 Structural bioinformatics (17 lectures by VL)

Topics to be covered:

- Introduction to protein structures and their features
- Comparative protein sequence analysis
- Methods of experimental structure determination and assessment of errors in protein structures
- Structural databases
- Prediction of protein structures and comparative protein structure modelling
- Alignment and classification of protein structures

## 8 Policy on plagiarism and cheating

Plagiarism means using the work of others in preparing an assignment and presenting it as your own without explicitly acknowledging - or referencing - where it came from. Plagiarism can also mean not acknowledging the full extent of indebtedness to a source.

Work can be plagiarised from many sources - including books, articles, the world wide web, and other students' assignments. Plagiarism can also occur unconsciously or inadvertently. Direct copying is definitely plagiarism. Paraphrasing of another work without attribution is also plagiarism. Submitting someone else's unattributed or less than fully attributed work or ideas is not evidence of your own grasp of the material and cannot earn you marks.

The general principle for determining whether cheating has occurred is whether work that is submitted for grading is the students own work, reflecting his or her learning. The best way to avoid being accused of plagiarism is to know what it is, how to cite references correctly and how to paraphrase information appropriately. In the context of laboratory assignments, avoid submitting written answers that are a collaborative effort.

Cheating and plagiarism will be penalised by deduction of some or all of the marks for the assignment. In addition, details of offences involving cheating and plagiarism will be kept on file in SBS and serious cases will be reported to the Discipline Committee. For further help or clarification you should consult:

- your lecturer or course coordinator,
- the University guidelines on Conduct of Coursework at [www.auckland.ac.nz/uaa/policies-procedures](http://www.auckland.ac.nz/uaa/policies-procedures),
- the Student Learning Centre (SLC) located in the Kate Edger Information Commons and at [www.slc.auckland.ac.nz](http://www.slc.auckland.ac.nz), or
- websites like [www.plagiarism.org](http://www.plagiarism.org).