



Study Guide

Bioinformatics (BIF)

Semester 2, 2017

Prepared by:

Dr Nicola Armstrong
Mathematics and Statistics
Murdoch University

Copyright © Department of Statistics, Macquarie University; Melbourne School of Population and Global Health, The University of Melbourne.



THE UNIVERSITY OF

MELBOURNE

Table of Contents

1. Contacts	3
2. Overview	3
3. Unit objectives	4
4. Method of Delivery & Communication	4
5. Assessment and semester activity	5
5.1 Assignments	6
5.2 Examination	6
5.3 Exercises	6
5.4 Online participation	6
6. Timetable	7
7. Textbooks	7
8. Software	8
9. Immediately useful websites	8
10. Complaints policy	8

1. Contact details

Dr Nicola Armstrong

Mathematics and Statistics

Murdoch University, WA 6150

Ph: (08) 9360 2480

Fax: (08) 9360 6346

Email: n.armstrong@murdoch.edu.au

In previous years, the course has been run by Graham Wood and Steven Brown from Macquarie University and they developed most of the material for this unit.

2. Overview

Bioinformatics, the finding of patterns in very large sets of biological data, has grown very rapidly over the past few decades. In this unit the aim will be to provide an introduction to the main data sources, techniques and software that make up the area. The unit does not assume knowledge of biology, so begins by providing an introduction to the central biological ideas. It progresses in a natural way, stepping from DNA to evolution and phylogenetics and thence to RNA and microarrays, exploring key ideas at each stage; all work is directed at an understanding of biological systems. En route, web-based resources will be explored and used to provide data sets, and the statistical package R employed to perform analyses.

3. Unit Objectives

At the end of this unit you should be able to:

1. Explain the core dogma of molecular biology and the central ideas of population genetics
2. Access appropriate web based sources for data, and download the data in suitable format, when given a problem that requires genome or proteome data for its solution.
3. Understand and apply core bioinformatics techniques for the analysis of DNA and protein sequence data, such as global sequence alignment, BLAST, Hidden Markov Models, evolutionary models and phylogenetic tree fitting.
4. Process large quantities of data (such as the expression profiles of thousands of genes resulting from microarray experiments) using R, and communicate results in language suitable for presentation to both a bioinformatics journal and a lay audience.

The unit is divided into nine modules, with either one or two weeks devoted to each. These are listed below.

Module	Weeks	Content
1	1	Biology basics
2	2,3	Population genetics
3	4	Web-based data sources, data retrieval and tools
4	5,6	The analysis of single and multiple DNA or protein sequences
5	7,8	Hidden Markov Models and their applications
6	9	Evolutionary models
7	10	Phylogenetic trees
8	11,12	Analysis of microarrays
9	13	Introduction to RNASeq

4. Method of Delivery and Communication

The instructor will be available by email and phone to answer questions related to the module notes and exercises, and to address other issues requiring clarification. However, please note that instructors are not necessarily available every day of the week and you should expect a delay in response! In most instances, email will be the most effective form of communication for an initial response; when doing so please use "BIF:" in the subject line of your email.

The unit is offered in distance mode. Our core means of communication will be via printed notes (which will be available on Blackboard), e-mail and Blackboard discussion.

The unit relies to an extent on the prescribed text by Durbin, Eddy, Krogh and Mitchison (see below). The study notes provide a guide to readings in this text, and often to other readings, which will be provided. They also provide additional explanation where this is needed. In the study notes for each module, exercises or tasks are given, referring to exercises in the textbook or provided readings.

The study notes for the first module will be mailed to you. These will also be posted on the Blackboard site. You will be alerted by email when this material has been mailed out, so if you do not receive these within a few days then please contact the Unit co-ordinator (contact details given earlier).

Blackboard

We will be using Blackboard for online discussions, posting of course notes, solutions and data sets, and submission of assignments. The link is: <http://elearning.sydney.edu.au/> which you should bookmark.

As Blackboard is the primary medium for communication in this unit, we expect that students will access the site at least every second day during semester.

Student “Profiles”

In order to introduce ourselves to each other, we encourage you to create a profile on Blackboard. Just go to “Discussions” then “Introductions”, then “Create Thread” and away you go. You are encouraged to include at least the following:

- your preferred name;
- the degree program you are studying;
- your current job role, and the organisation for which you work;
- what you hope to gain from the unit;
- and a little about yourself!

You may include your e-mail address if you want to be contactable directly by other students. It would be good if you could do this by the end of the first week of semester.

Feedback

Your constructive comments are always valued and guide us in the continuous improvement of the unit.

5. Assessment and semester activity

Assessment for the unit consists of: three assignments and a written examination. The weighting is as follows:

Three assignments (20% each)	60%
Examination	40%

5.1 Assignments

Rules for assignment submission are given in the document “BCA Assessment Guide” which is available on the BCA Blackboard site. *Please read this document carefully.* All students are required to sign an “own work” declaration before submission of assignments is allowed via Blackboard. Please see the “[Assessment Guide](#)” for more information.

Late submission: Requests for an extension of the due date for an assignment must be made in advance of the due date for that assignment. These requests must be made directly to the unit coordinator by email. The unit coordinator will reply by email with the decision as to whether an extension has been granted and the new due date.

Late penalty: If no extension has been given, 5% of the earned mark for an assignment will be deducted for each day that an assignment is late, up to a maximum of 50%.

NOTE: It is not the intention of this late penalty policy to cause a student to fail the unit when otherwise they would have passed. If deductions for late assignments result in the final unit mark for a student being less than 50, when otherwise it would have been 50 or greater, the student's final mark will be exactly 50.

5.2 Examination

The examination will essentially be an extended assignment, to be completed over the period of a week. It will be released on Blackboard at 5pm on Monday 6 November, and is due at 5pm on Monday 13 November. You should consider yourself under examination conditions, i.e. there is to be no communication between yourself and other students or any other person, in connection with the examination.

5.3 Exercises

There are exercises in the notes. Exercises should be attempted. Online discussions around each module's exercises can be conducted on Blackboard. You are encouraged to make discussion contributions, on any of the questions, including (but we hope not limited to) those designated for submission.

5.4 Online participation

You are strongly encouraged to participate in online discussions. We are a dispersed group and this forum can bring a little unity!

Instructors will generally let discussion flow between the students, except where key points need resolution. The instructors will not usually make comments on the exercises required for submission, at least until after the deadline.

Any general Discussion items, in particular on the module notes, can be posted to the other Discussion areas, e.g. R, Module 1, Module 2, etc. If you want the attention of the

lecturers you are very strongly advised to be in contact directly, using email. This is monitored more regularly than the discussion area on Blackboard.

6. Timetable

Week	Starting date	Module	Assessment due
1	31 July	1	
2	7 August	2	
3	14 August	2	
4	21 August	3	Assignment 1
5	28 August	4	
6	4 September	4	
7	11 September	5	
8	18 September	5	Assignment 2
Mid semester break			
9	2 October	6	
10	9 October	7	
11	16 October	8	
12	23 October	8	Assignment 3
13	30 October	9	

Assignments are due on the Monday of the indicated weeks. Assignments will be available on Blackboard at least two weeks before the due date. The examination (see above) will be an extended assignment, due Monday 13 November.

7. Textbooks

The prescribed text is:

R. Durbin, S. Eddy, A. Krogh and G. Mitchison (1998). *Biological Sequence Analysis*, Cambridge University Press. ISBN 0 521 62041 4 (hardcover) or 0 521 62971 3 (paperback).

Copies are available from the Co-op bookshop (can be ordered online). You will start to need the text in Module 4, so it is a good idea to order it early.

There are numerous texts on bioinformatics. The following may be useful:

1. M. Zvelebil and J. Baum, (2008) *Understanding bioinformatics*, Garland Science.
2. Warren J. Ewens and Gregory R. Grant (2005) *Statistical methods in bioinformatics - an introduction*, Springer-Verlag. (Excerpts from this texts will be provided.)
3. Ben Hui Liu, (1998) *Statistical Genomics: Linkage, Mapping and QTL analysis*, CRC Press.
4. Anna Tramontano (2007) *Introduction to bioinformatics*, Chapman and Hall. (Good on data sources and sequence analysis; strong on protein analysis, not in this course.)
5. D.P. Snustad and M.J.Simmons (2003) *Principles of Genetics*, John Wiley and Sons. (Excellent as a molecular biology reference.)

6. Cynthia Gibas and Per Jambeck (2001) Developing Bioinformatics Computer Skills, O'Reilly. (Useful if you move into software development in bioinformatics.)

8. Software

We will be using the statistical package R. You can download and install a version of this reliable freeware from the internet at <http://www.r-project.org/>. You may also want to install RStudio, a free GUI (graphical user interface) for R, but this is not essential.

9. Immediately useful websites

- BCA Blackboard homepage: <http://elearning.sydney.edu.au/>
- BCA homepage: <http://www.bca.edu.au>
- BCA Student Resources (including the Guide for Reporting Statistical Results): <http://www.bca.edu.au/currentstudents.html>
- R homepage: <http://www.r-project.org/>

10. Complaints policy

Please see the BCA complaints policy in the Assessment Guide

<http://www.bca.edu.au/currentstudents.html#assessmentguide>