UNIVERSITY of TASMANIA

**School of Physical Sciences**

**Faculty of Science, Engineering and Technology**

# KMA712
# Introduction to Bioinformatics

**Semester 1, 2017**

**Unit Outline**

**Michael Charleston**

CRICOS Provider Code: 00586B

# CONTACT DETAILS

## Unit coordinator


Unit coordinator: Michael Charleston

Campus: Sandy Bay

Email: michael.charleston@utas.edu.au

Phone: (03) 6226 2444

Room location and number: Maths & Physics Building, Clark Road, Rm443

Consultation hours: Open door policy

# WHAT IS THE UNIT ABOUT?

## Unit description*

Bioinformatics is a core skill that is desperately needed by practitioners in the life sciences, from health to pathology and epidemiology to fundamental biological research. It is a discipline that brings together tools from mathematics, statistics, and computer science, to answer challenging questions that arise in the modern life sciences.

The unit is aimed at those pursuing research in the life sciences and/or working with large biological data sets, and is intended to provide them with sufficient knowledge to effectively use existing bioinformatics tools, to understand how the bioinformatics tools work and what can or cannot be done with them, and to develop and critically assess bioinformatics workflows and methodologies. These tools and skills will enhance students' research capacity in life sciences.

## Intended Learning Outcomes*

On completion of this unit, students will be able to

1. Effectively use a range of *current bioinformatics tools and workflows* to solve problems in life sciences ("software");
2. Explain and apply the fundamental *computational methods* underlying the solution of bioinformatics problems ("methods");
3. Classify the *major classes* of bioinformatics problems and choose appropriate tools to solve them ("major problems");
4. Evaluate and *communicate* bioinformatics methodologies clearly ("communication").

## Graduate Quality Statement

Successful completion of this unit supports your development of course learning outcomes, which describe what a graduate of a course knows, understands and is able to do. Course learning outcomes are available from the Course Coordinator. Course learning outcomes are developed with reference to national discipline standards, Australian Qualifications Framework (AQF), any professional accreditation requirements and the University of Tasmania's Graduate Quality Statement.

> The University of Tasmania experience unlocks the potential of individuals. Our graduates are equipped and inspired to shape and respond to the opportunities and challenges of the future as accomplished communicators, highly regarded professionals and culturally competent citizens in local, national, and global society. University of Tasmania graduates acquire subject and multidisciplinary knowledge and skills and develop creative and critical literacies and skills of inquiry. Our graduates recognise and critically evaluate issues of social responsibility, ethical conduct and sustainability. Through respect for diversity and by working in individual and collaborative ways, our graduates reflect the values of the University of Tasmania.

## Alterations to the unit as a result of student feedback*

This unit is a re-design of the previous version, which was a short intensive "block" course. In response to feedback, this redesign is more spread out to give students time to absorb material better and to give more time for practising the skills that are being taught. Also being offered in blended mode containing pre-recorded lectures and with some online assessment, the new design is intended to maintain flexibility for research students who are not always able to be on campus due to field work etc.

## Prior knowledge and/or skills

**Assumed knowledge:**

Students enrolling in it should have a BSc or equivalent.

Prior to beginning the unit you should have a *working knowledge* of the fundamentals of molecular biology (such as that covered by KPZ164), in particular the structure of DNA and RNA, the fundamental dogma of molecular biology, and essential evolutionary theory. You should have an *appreciation* of simple mathematical terminology and symbols and be able to carry out simple statistical analyses.

It is not assumed that you will have any programming experience or advanced understanding in mathematics or statistics.

# HOW WILL I BE ASSESSED?*

## Assessment schedule*

| Assessment task | Date due | Percent weighting | Intended Learning Outcomes measured |
|---|---|---|---|
| Assessment Task 0: Self-assessment quiz | Week 0 | 0% | |
| Assessment Task 1: Ongoing online quiz | Weekly except week 6 | 30% | ILO 1, 2, 3, 4 |
| Assessment Task 2: Practical exercises | Weeks 2, 4, 7, 9, 12 | 15% | ILO 1, 2 |
| Assessment Task 3: Mid-semester quiz | Week 6 | 5% | ILO 2, 3, 4 |
| Assessment Task 4: Assignment 1 | Week 8 | 20% | ILO 1, 2, 4 |
| Assessment Task 5: Assignment 2 | Exam period | 30% | ILO 1, 2, 3, 4 |

## Assessment details*

### Assessment Task 0: Self-assessment quiz

| | |
|---|---|
| Description / conditions | This is an online self-assessment of the students' own ability, which will provide a baseline of their perceived knowledge level prior to starting the course. |
| Assessment criteria | This is an opportunity for students to rate their ability in areas pertinent to this unit, such as their own confidence in using command-line on a computer and their knowledge of underlying concepts. At the end of the unit they will have the opportunity to reassess their ability. It will take the form of an online survey on MyLO. |
| Intended Learning Outcomes measured | This has no weighting to the students' final grade. |
| Duration | Approximately one hour. |
| Date | Week 0 |

### Assessment Task 1: Ongoing online quiz

| | |
|---|---|
| Description / conditions | This ongoing, online quiz will have questions each week to assess students' understanding of material just covered. Each will contribute 2.5% to the total grade, coming to 30%. The regular testing will serve both to give rapid feedback to the students on their progress, and to keep them on track with the current topics so they don't fall behind. However, to maintain flexibility for the students, each set of questions will remain open for two weeks once made available. The questions will not go into depth on practical matters such as implementation or running particular analyses, as these will be covered by other assessment tasks. |
| Assessment criteria | There will be approximately five questions each week, designed to assess students understanding on each week's material. The questions will be available on MyLO and auto-marked where appropriate. Students will be able to monitor their progress throughout the semester. |
| Intended Learning Outcomes measured | ILO1 (software): 20%<br>ILO2 (methods): 40%<br>ILO3 (major problems): 30%<br>ILO4 (communication): 10% |
| Duration | Expected to take no more than about 30 minutes each week. |
| Date | Throughout the semester: weeks 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, NOTE: slightly larger "half-way" quiz in week 6 (see later). |

## Assessment Task 2: Ongoing practical exercises

| | |
|---|---|
| Task description | These practical exercises are designed to help students get hands-on experience in applying their knowledge to solve real problems in bioinformatics. They will work in a lab or remotely with online help from the Unit Coordinator or agent thereof, with time to complete the exercises before submission for assessment. **Students are expected to attend a majority of these practicals in person.** |
| Assessment criteria | The assessment will be graded according to the following criteria:<br>Correct use of software tools for the task (80%).<br>Interpretation of results (20%). |
| Intended Learning Outcomes measured | ILO1 (software): 80%<br>ILO2 (methods): 20% |
| Task length | Approximately 1-2 hours each task |
| Date due | Weeks 2, 4, 7, 9 and 12 only. |

## Assessment Task 3: Mid-semester Quiz

| | |
|---|---|
| Task description | This pen-and-paper quiz will be for students to demonstrate that they understand, and can communicate, the concepts and methods that have been covered in the material up to the previous week. **Students must attend the quiz in person.** |
| Assessment criteria | The assessment will be graded on demonstrated understanding of the major bioinformatics problems covered thus far and the computational methods that are used to solve them. |
| Intended Learning Outcomes measured | ILO2 (methods): 40%<br>ILO3 (major problems): 40%<br>ILO4 (communication): 20% |
| Task length | One hour |
| Date due | Week 6. |

## Assessment Task 4: Assignment 1

| | |
|---|---|
| Task description | Assignment 1 is to use a given set of bioinformatics tools to perform an analysis on a publicly available data set, for example, to find the set of all possible best matches between a set of query sequences and a set of target sequences.<br><br>The analysis will require students to download data from a public database, use command-line tools to chain together results from multiple tools, and to interpret the overall results.<br><br>The results should be presented clearly, critically assessed, and should be communicated with an intent that they be easily repeatable. |
| Assessment criteria | The assessment will be graded according to the following criteria:<br><br>Software (ILO1): appropriate choice (10%) and correct use (20%) of software tools for the task.<br><br>Methods (ILO2): quality of the way in which the analysis was carried out (20%) and interpretation of results (20%)<br><br>Communication (ILO4): clarity, organization of the exposition and ease of repeatability (10% each) |
| Intended Learning Outcomes measured | ILO1 – software: 30%<br><br>ILO2 – methods: 40%<br><br>ILO4 – communication: 30% |
| Task length | At least 5 pages including figures; around 1500 words |
| Date due | Week 8. |

## Assessment Task 5: Assignment 2

| | |
|---|---|
| Task description | Assignment 2: The second assignment is to design and implement a repeatable bioinformatics workflow in order to answer a set of questions regarding the students' own data (or other data if required).<br><br>Some of these questions will be chosen from the "major classes" of bioinformatics problems, being the most appropriate one(s) for the data set used. The methods used to solve the problems will be thoroughly explained as to a scientific, but not necessarily bioinformatics-focussed, audience.<br><br>The assignment will describe the analysis and provide a critical discussion of the results.<br><br>The workflow that is created must be shared in such a way that it can be run by others on the same data, e.g., using a public bioinformatics platform such as Galaxy. |
| Assessment criteria | The assessment will be graded according to the following criteria:<br><br>Software (ILO1): design of workflow;<br><br>Methods (ILO2): evidence of understanding of the methodologies used;<br><br>Major Problems (ILO3): understanding and communication of this part of the analysis – dependent on the nature of the students' data;<br><br>Communication (ILO4): clarity of the exposition and ease of repeatability (10% each) |
| Intended Learning Outcomes measured | ILO1: 30%<br><br>ILO2: 30%<br><br>ILO3: 20%<br><br>ILO4: 20% |
| Task length | At least 8 pages including figures; around 2500 words |
| Date due | Examination period |

## How your final result is determined*

In order to pass the unit each of the intended learning outcomes must be attained, and as such the performance against each of these will be assessed separately.

For students attaining all of the learning outcomes, the mark for this unit will be calculated as the weighted average of the three assessments, take home quiz and two assignments, as per the assessment schedule.

## Submission of assignments*

Submission of assignments will be on the MyLO website for this unit. Written documents must all be in .pdf format; if there are multiple files for a submission these must be compressed as .zip archives.

### Requests for extensions

Requests for extensions should be made in writing to the Unit Coordinator; all reasonable requests will be considered.

### Penalties*

Work submitted late without prior arrangement or permission for an extension will incur a late penalty of 20% of the assessment's worth per day or part thereof.

### Review of results and appeals

If you wish to appeal a decision of the Unit Coordinator on a grade then you must make your appeal in writing to the Head of Discipline of Mathematics within 7 days of receiving your grade.

## Academic referencing*

In your written work you will need to support your ideas by referring to scholarly literature, works of art and/or inventions. It is important that you understand how to correctly refer to the work of others, and how to maintain academic integrity.

Failure to appropriately acknowledge the ideas of others constitutes academic dishonesty (plagiarism), a matter considered by the University of Tasmania as a serious offence.

The appropriate referencing style for this unit is the "abbrv" BibTeX style, available from the Comprehensive TeX Archive Network (CTAN). This style has citations in brackets and numbered in the order in which they appear in the text; the bibliography has authors names displayed as initials-surname. If three or more authors are given just put the first one followed by "*et al.*". If submitting work using a word processor such as Word, the equivalent citation style should be used if possible.

The University library provides information on presentation of assignments, including referencing styles and should be referred to when completing tasks in this unit.

Please read the following statement on plagiarism. Should you require clarification please see your unit coordinator or lecturer.

### Plagiarism

Plagiarism is a form of cheating. It is taking and using someone else's thoughts, writings or inventions and representing them as your own; for example, using an author's words without putting them in quotation marks and citing the source, using an author's ideas without proper acknowledgment and citation, copying another student's work.

If you have any doubts about how to refer to the work of others in your assignments, please consult your lecturer or tutor for relevant referencing guidelines. You may also find the Academic Honesty site on MyLO of assistance.

The intentional copying of someone else's work as one's own is a serious offence punishable by penalties that may range from a fine or deduction/cancellation of marks and, in the most serious of cases, to exclusion from a unit, a course or the University.

**The University and any persons authorised by the University may submit your assessable works to a plagiarism checking service, to obtain a report on possible instances of plagiarism. Assessable works may also be included in a reference database. It is a condition of this arrangement that the original author's permission is required before a work within the database can be viewed.**

For further information on this statement and general referencing guidelines, see the Plagiarism and Academic Integrity page on the University web site or the Academic Honesty site on MyLO.

## Academic misconduct*

Academic misconduct includes cheating, plagiarism, allowing another student to copy work for an assignment or an examination, and any other conduct by which a student:

a. seeks to gain, for themselves or for any other person, any academic advantage or advancement to which they or that other person are not entitled; or

b. improperly disadvantages any other student.

Students engaging in any form of academic misconduct may be dealt with under the Ordinance of Student Discipline, and this can include imposition of penalties that range from a deduction/cancellation of marks to exclusion from a unit or the University. Details of penalties that can be imposed are available in Ordnance 9: Student Discipline – Part 3 Academic Misconduct.

# WHAT LEARNING OPPORTUNITIES ARE THERE?

## MyLO

MyLO is the online learning environment at the University of Tasmania. This is the system that will host the online learning materials and activities for this unit.

### Getting help with MyLO

It is important that you are able to access and use MyLO as part of your study in this unit. To find out more about the features and functions of MyLO, and to practice using them, visit the Getting Started in MyLO unit. For access to information about MyLO and a range of step-by-step guides in pdf, word and video format, visit the MyLO Student Support page on the University website.
If something is not working as it should, first check the current issues page in case this is a known problem; otherwise, contact the service desk (http://www.utas.edu.au/service-desk, e-mail service-desk@utas.edu.au, or phone 6226 1818), or request IT help online.

## Resources

### Required readings

There are no required readings for this course.

### Recommended readings

Recommended readings will be provided as new publications appear in this rapidly changing field. In general, the journal Briefings in Bioinformatics is a good starting point for most topics covered.

### Reading Lists

Reading Lists provide direct access to all material on unit reading lists in one place. This includes eReadings and items in Reserve. You can access the Reading List for this unit from the link in MyLO, or by going to  on the University Library website.

### Equipment, materials, software, accounts

All necessary equipment, software and accounts will be provided.

## Activities

### Learning expectations

The University is committed to high standards of professional conduct in all activities, and holds its commitment and responsibilities to its students as being of paramount importance. Likewise, it holds expectations about the responsibilities students have as they pursue their studies within the special environment the University offers.

The University's Code of Conduct for Teaching and Learning states:

**Students are expected to participate actively and positively in the teaching/learning environment. They must attend classes when and as required, strive to maintain steady progress within the subject or unit framework, comply with workload expectations, and submit required work on time.**

## Details of teaching arrangements*

The unit will be delivered as a blended unit, which can be mostly completed online. There are some assessments that require students to be present: the mid-semester quiz, and at least a majority of the lab exercises.

MyLO will be used as a focus for online discussions until the end of the course.

## Specific attendance/performance requirements*

Students must attend the mid-semester quiz and a majority of the lab practical exercises.

## Teaching and learning strategies

The *blended*, *flexible* delivery of this course, coupled with continuing support through online discussions and tutorials, is designed to fit in with requirements of research scientists and practitioners who are unable to commit long periods to a study course.

The intensive course is also intended to deliver the skills and knowledge to students as quickly as possible as they are often in demand for current research challenges.

## Work Health and Safety (WHS)

The University is committed to providing a safe and secure teaching and learning environment. In addition to specific requirements of this unit you should refer to the University's  and policy.

## Communication

News and announcements will all be via MyLO; students will be expected to be aware of the content of such posts within 48 hours of them being posted.

## Further information and assistance

If you are experiencing difficulties with your studies or assignments, have personal or life-planning issues, disability or illness which may affect your course of study, you are advised to raise these with the unit coordinator in the first instance.

There is a range of University-wide support services available to you including Student Learning Support, Student Advisers, Disability Services, and more which can be found on the Student Support and Development page of the University website.

Should you require assistance in accessing the Library, visit the Library website for more information.

## Unit schedule

| WEEK | DATE BEGINNING | TOPIC/ MODULE/ FOCUS AREA | ACTIVITIES |
|---|---|---|---|
| 0 | February 22 | Orientation | Self-assessment survey |
| * | Continuing | | online lectures; weekly online practice questions |
| 1 | February 27 | Introduction | Ongoing online quiz begins |
| 2 | March 6 | Command-line and Cloud | Practical exercise |
| 3 | March 13 | Fundamental mathematics & statistics | |
| 4 | March 20 | Pairwise sequence alignment | Practical exercise |
| 5 | March 27 | Multiple sequence alignment | |
| 6 | April 3 | Likelihood and Models | Mid-semester quiz |
| 7 | April 10 | Phylogenetic Inference | Practical exercise |
| 8 | April 24 | Finding perfect matches: hashing | Assignment 1 |
| 9 | May 1 | Genomics and Sequencing | Practical exercise |
| 10 | May 8 | Read mapping | |
| 11 | May 15 | Genome assembly | |
| 12 | May 22 | RNA-Seq | Practical exercise |
| 13 | May 29 | CHiP-Seq | |
| exam period | | | Assignment 2 |