

University of Macau
Faculty of Science and Technology
Department of Computer and Information Science
CISB366 Bioinformatics
Syllabus
2nd semester 2016/2017
Part A – Course Outline

Elective course in Computer Science

Course description:

(D2H1.5) 3 credits. Bioinformatics is the study of biological information through computer modeling or analysis. Its goal is to reveal relationships between sequences, structures, and functions of molecules. In this course, we intend to give a broad introduction of the algorithmic techniques used in bioinformatics. Topics which will be covered include sequence similarity analysis, biological database search, phylogenetic analysis, and machine learning.

This course is designed for undergraduate CS students and assumes no prior knowledge of molecular biology beyond the high school level. Basic concepts of molecular biology will be given in the first lecture, as well as in subsequent lectures before the computational problems are defined. The ultimate goal of this course is to prepare students with the knowledge and skills to conduct research in the area of bioinformatics. Assignments will include both solving theoretical problems and programming.

Course type:

Theoretical

Prerequisites:

CISB120 Algorithms and data structures I

Programming skill of at least one programming language (such as C, Java, Perl or Python)

Textbook(s) and other required material:

- Wing-Kin Sung, *Algorithms in Bioinformatics: A Practical Introduction*, Chapman and Hall/CRC, 2009.

References:

- Ingvar Eidhammer, Inge Jonassen, and William R. Taylor, *Protein Bioinformatics*, Wiley, 2004.
- Caroline St. Clair, Jonathan Visick, *Exploring Bioinformatics – A project-based approach*, Jones and Bartlett, 2010.

Major prerequisites by topic:

- Fundamental data structures and algorithms

Course objectives:

- Understand the fundamental biological concepts and acquire the vocabulary to analyze bioinformatics problems [a]
- Learn the key bioinformatics algorithms and their implementations [a, c]
- Become familiar with existing tools and resources for computational analysis of biological data [c]

Topics covered:

- **Introduction to molecular biology (3 hours):** In this lecture, we introduce the basic concepts in molecular biology, which include the building blocks of cells (DNA, RNA, and protein), the central dogma, the genetic code and some basic biotechnologies.
- **Sequence similarity (9 hours):** Comparing sequences is the most fundamental operation in bioinformatics. Here, we will discuss various alignment methods, including global alignment, local alignment, and semi-global alignment. We also study gap penalty and scoring function. Two important scoring matrices BLOSUM and PAM will be introduced.

- **Database search (6 hours):** Due to the advance of biotechnology, sequence data increases exponentially. Hence it is important to have methods which allow efficient database searching. In this lecture, we discuss various biological database searching methods including FASTA, BLAST, and variations of BLAST algorithm.
- **Multiple sequence alignment (3 hours):** Multiple sequence alignment is a natural extension of pairwise alignment. Comparing several sequences can reveal what is common for a whole family and thus provide a basis for the phylogenetic analysis. Several approaches for making multiple alignments including dynamic programming, center star, progressive and iterative methods will be described in this lecture.
- **Phylogeny reconstruction and comparison (9 hours):** This lecture describes a phylogenetic tree and its applications. We discuss how to construct a phylogenetic tree given a character-based dataset or a distance-based dataset. Then, we discuss methods for computing similarity and distance, and finding consensus of a set of trees.
- **Immunoinformatics and machine learning (6 hours):** Machine learning methods are essential tools for many bioinformatics applications. In this lecture, we introduce artificial neural network (ANN) as a method to solve an important classification problem in vaccine design, namely, to classify peptides which bind or not bind to Major Histocompatibility Complex (MHC) molecules of the human immune system.
- **Computer-aided drug design (6 hours):** The field of structure-based drug design explicitly exploits the concept of 3D binding sites that interact with ligands, computational methods and models play herein a critical role. In this lecture, we describe the main components of a protein-ligand docking algorithm (i.e. the search algorithm and the scoring function). The most popular docking software AutoDock will be discussed and demonstrated its use. Since this is the only *Structure* chapter in this course, introduction to protein structure (secondary, tertiary and quaternary) and methods for structure comparison (RMSD calculation) will be given.

Class/laboratory schedule:

Timetabled work in hours per week			No of teaching weeks	Total hours	Total credits	No/Duration of exam papers
Lecture	Tutorial	Practice				
2	2	0	14	42	3	2/2&3 hours

Student study effort required:

Class contact:	
Lecture	28 hours
Tutorial	26 hours
Mid-term Exam	2 hours
Other study effort	
Self-study	42 hours
Assignment	9 hours
Project	10 hours
Total student study effort	117 hours

Student assessment:

Final assessment will be determined on the basis of:

Assignment	10%
Project	30%
Mid-term exam	20%
Final exam	40%

Course assessment:

The assessment of course objectives will be determined on the basis of:

- Homework assignments, projects and exams
- Course evaluation

Course outline:

Week	Topic	Course work
1	Introduction to Molecular Biology DNA, RNA, proteins, the central dogma, genetic code	

Week	Topic	Course work
2-4	Sequence Similarity Global alignment, local alignment, gap penalty, scoring function, PAM, BLOSUM matrices	Assignment 1
5-6	Database Search Biological databases, FASTA, BLAST, E-value, p-value, bit score, PSI-BLAST, PSSM	Project 1
7	Multiple Sequence Alignment Dynamic programming method, center star method, progressive alignment method (ClustalW, profile-profile alignment), iterative method	Assignment 2
8-10	Phylogeny Reconstruction Evolution, constant molecular clock assumption, phylogeny, tree-rooting method, character-based methods (Fitch's and Sankoff's algorithms), distance-based methods (UPGMA, Neighbor Joining), searching the tree space	Mid-term
11-12	Immunobioinformatics and Machine Learning Immune system, vaccine design, weight matrix method, Artificial Neural Network (ANN), MHC-peptide classification	Project 2
13-14	Computer-aided Drug Design Disease and drug, drug development process, Structure-based Drug Design (SBDD), protein-ligand docking algorithm, scoring function, RMSD calculation, virtual screening	Assignment 3

Contribution of course to meet the professional component:

This course prepares students with fundamental knowledge and experiences in bioinformatics.

Relationship to CS program objectives and outcomes:

This course primarily contributes to the Computer Science program outcomes that develop student abilities to:

- (a) An ability to apply knowledge of computing and mathematics to solve complex computing problems in computer science discipline;
- (c) An ability to analyse a problem, and identify and define the computing requirements appropriate to its solution.

Relationship to CS program criteria:

Criterion	DS	PF	AL	AR	OS	NC	PL	HC	GV	IS	IM	SP	SE	CN
Scale: 1 (highest) to 4 (lowest)			1											2

Discrete Structures (DS), Programming Fundamentals (PF), Algorithms and Complexity (AL), Architecture and Organization (AR), Operating Systems (OS), Net-Centric Computing (NC), Programming Languages (PL), Human-Computer Interaction (HC), Graphics and Visual Computing (GV), Intelligent Systems (IS), Information Management (IM), Social and Professional Issues (SP), Software Engineering (SE), Computational Science (CN).

Course content distribution:

Percentage content for			
Mathematics	Science and engineering subjects	Complementary electives	Total
0%	100%	0%	100%

Persons who prepared this description:

Dr. W. I. Siu

Part B – General Course Information and Policies

2nd Semester 2016/2017

Instructor: Dr. W. I. Siu

Office: E11-4025

Office hour: Tue, Fri 3:00 – 5:00 pm, or by appointment

Phone: 8822 4452

Email: shirleysiu@umac.mo

Time/Venue:

Grading distribution:

Percentage Grade	Final Grade	Percentage Grade	Final Grade
100 - 93	A	92 - 88	A–
87 - 83	B+	82 - 78	B
77 - 73	B–	72 - 68	C+
67 - 63	C	62 - 58	C–
57 - 53	D+	52 - 50	D
below 50	F		

Comment:

Objectives of the lectures are to explain and to supplement the text material. Students are responsible for the assigned material whether or not it is covered in the lecture. Students who wish to succeed in this course should read the textbook prior to the lecture and should work all homework assignments. You are encouraged to look at other sources (other texts, etc.) to complement the lectures and text.

Homework policy:

- There will be 3 homework assignments and 2 programming projects.
- Assignments and projects can be worked in a group of at most 2; deadline will be announced when the assignment is released. Late penalty will be applied weighted by the number of days delayed.
- Copying homework is not tolerated. Zero mark will be given regardless of which is the original copy.

Exams:

There will be two examinations. One in the middle of the semester (mid-term) and another one at the end of the semester (final exam). Both exams are closed-book and closed-notes.

Note:

- Check UMMoodle (<https://ummoodle.umac.mo/>) for announcement, homework, lab/tutorial arrangements and lectures. Report any mistake on your grades within one week after posting.
- No make-up exam is given except for CLEAR medical proof.
- Cheating is absolutely prohibited by the university.

STUDENT DISABILITIES SUPPORT SERVICE

The University of Macau is committed to providing an equal opportunity in education to persons with disabilities. If you are a student with a physical, visual, hearing, speech, learning or psychological impairment(s) which substantially limit your learning and/or activities of daily living, you are encouraged to communicate with your instructors about your impairment(s) and the accommodations you need in your studies. You are also encouraged to contact the Student Disability Support Service of the Student Counselling and Development Section (SCD) in Student Affairs Office, which provides appropriate resources and accommodations to allow each student with a disability to have an equal opportunity in education, university life activities and services at the University of Macau. To learn more about the service, please contact SCD at scd.disability@umac.mo, or 8822 4901 or visit the following website: http://www.umac.mo/sao/scd/sds/aboutus/en/scd_mission.php.

Appendix:

Rubric for Program Outcomes

(a) An ability to apply knowledge of computing and mathematics appropriate to the programme outcomes and to the discipline

Measurement Dimension	Excellent (80-100%)	Average (60-79%)	Poor (<60%)
1. An ability to apply knowledge of computing to the solution of complex computing problems.	Students understand the computing principles, and their limitations in the respective applications. Use the computing principles to formulate and solve complex computing problems.	Students understand the computing principles, and their limitations in the respective applications. But they have trouble in applying these computing principles to formulate and solve complex computing problems.	Students do not understand the computing principles, and their limitations in the respective applications. Do not know how to apply the appropriate computing principles to formulate and solve complex computing problems.

(c) An ability to analyse a problem, and identify and define the computing requirements appropriate to its solution

Measurement Dimension	Excellent (80-100%)	Average (60-79%)	Poor (<60%)
1. An ability to understand problem and identify the fundamental formulation	Students understand problem correctly and can identify the fundamental formulation	Student understand problem correctly, but have trouble in identifying the fundamental formulation	Students cannot understand problem correctly, and they do not know how to identify the fundamental formulation
2. An ability to choose and properly apply the correct techniques	Students know how to choose and properly apply the correct techniques to solve problem.	Students can choose correct techniques but have trouble in applying these techniques to solve problem.	Students have trouble in choosing the correct techniques to solve problem.