

---

UNIVERSITI SAINS MALAYSIA

First Semester Examination  
2015/2016 Academic Session

December 2015/January 2016

**CIT562 – Bioinformatics Computing**  
*[Perkomputeran Bioinformatik]*

Duration : 2 hours  
*[Masa : 2 jam]*

---

**INSTRUCTIONS TO CANDIDATE:**

***[ARAHAN KEPADA CALON:]***

- Please ensure that this examination paper contains **SIX** questions in **FIVE** printed pages before you begin the examination.

*[Sila pastikan bahawa kertas peperiksaan ini mengandungi **ENAM** soalan di dalam **LIMA** muka surat yang bercetak sebelum anda memulakan peperiksaan ini.]*

- Answer **ALL** questions.

*[Jawab **SEMUA** soalan.]*

- You may answer the questions either in English or in Bahasa Malaysia.

*[Anda dibenarkan menjawab soalan sama ada dalam bahasa Inggeris atau bahasa Malaysia.]*

- In the event of any discrepancies, the English version shall be used.

*[Sekiranya terdapat sebarang percanggahan pada soalan peperiksaan, versi bahasa Inggeris hendaklah diguna pakai.]*

1. (a) What is “bioinformatics” as described by Luscombe, Greenbaum and Gerstien in their review paper “What is bioinformatics? An Introduction and overview”, Yearbook of Medical Informatics, 2001.  
(5/100)
- (b) What are other definitions of Bioinformatics and what are the main differences between these definitions?  
(5/100)
- (c) Discuss the main sub-research areas in Bioinformatics and what are the type of data involved in the sub-research areas.  
(10/100)
2. (a) Compare the different data types in biological databases and name the main public databases that keep them.  
(8/100)
- (b) The central dogma in molecular biology is an important concept and theory for many researchers in Bioinformatics. Discuss the concept and name the processes or functions involved in each of the stage of the central dogma.  
(7/100)
3. Sequence Alignment is the basic and important operation in Bioinformatics. Substitution matrix is an important substitution cost matrix when doing an alignment.
  - (a) Discuss different types of Substitution Matrices and explain their respective usage.  
(6/100)
  - (b) Distinguish **three (3)** characteristics of BLAST, FASTA and Dynamic Programming Based sequence alignment algorithms. Compare them on the basis of:
    - the main focus of one algorithm,
    - how comparison operation is done,
    - type of parameters involved in the alignment.  
(9/100)

4. Multiple Sequence Alignment (MSA) cannot be efficiently handled using pure dynamic programming.
- (a) Choose **two (2)** existing MSA methods and compare them on the basis of parameters involved and the alignment process. (10/100)
- (b) Discuss the **five (5)** processes in the automatic proteins analysis pipeline. (10/100)
5. The purpose of phylogeny is to reconstruct the history of life and explain the present diversity of living creatures. Phylogenetics is a special kind of phylogeny that relies on the comparison of equivalent genes coming from several species for reconstructing the genealogic tree of these species.
- (a) Discuss the **three (3)** major reason of why you want to use Phylogenetics. (6/100)
- (b) What are the **four (4)** major ingredients to be consider when a distance based methods are used to compute a phylogenetic tree? (9/100)
- 6 Perl is a popular programming language that is extensively used in areas such as Bioinformatics and web programming. Perl has become popular with biologists because it is so well-suited to several bioinformatics tasks.
- (a) What are the special features of Perl that makes it a popular choice for bioinformatics applications? (5/100)
- (b) Write a Perl program to concatenate two DNA sequences. The output of the program is given below:
- Here are the original two DNA fragments:  
ACGGGAGGACGGGAAAATTACTACGGCATTAGC  
ATAGTGCCGTGAGAGTGATGTAGTA  
Here is the concatenation of the first two fragments  
(version 1):  
ACGGGAGGACGGGAAAATTACTACGGCATTAGCATAGTGCCGTGAGAGTGATGTAGTA  
Here is the concatenation of the first two fragments  
(version 2):  
ACGGGAGGACGGGAAAATTACTACGGCATTAGCATAGTGCCGTGAGAGTGATGTAGTA  
Here is the concatenation of the first two fragments  
(version 3):  
ACGGGAGGACGGGAAAATTACTACGGCATTAGCATAGTGCCGTGAGAGTGATGTAGTA
- (10/100)

## KERTAS SOALAN DALAM VERSI BAHASA MALAYSIA

[CIT562]

- 4 -

1. (a) Apa itu "bioinformatics" seperti yang dihuraikan oleh Luscombe, Greenbaum dan Gerstien dalam kertas ulasan "What is bioinformatics? An Introduction and overview", Yearbook of Medical Informatics, 2001.  
(5/100)
- (b) Apakah definisi-definisi Bioinformatik dan apakah perbezaan antara definisi-definisi ini?  
(5/100)
- (c) Bincangkan sub-sub bidang penyelidikan dalam Bioinformatik dan apakah jenis data yang terlibat dalam setiap sub-bidang penyelidikan ini.  
(10/100)
2. (a) Bandingkan jenis-jenis data yang berbeza dalam pangkalan data biologi dan namakan pangkalan data umum yang utama yang menyimpan data-data ini.  
(8/100)
- (b) Dogma pusat dalam biologi molekul adalah konsep dan teori penting bagi kebanyakan penyelidik dalam Bioinformatik. Bincangkan konsep ini dan namakan proses-proses atau fungsi-fungsi yang terlibat pada setiap peringkat dogma pusat ini.  
(7/100)
3. Penjajaran Jujukan adalah operasi asas dan penting dalam Bioinformatik. Matrik penggantian merupakan satu matrik kos penggantian yang penting semasa proses penjajaran.
  - (a) Bincangkan jenis-jenis matrix penggantian yang berbeza dan terangkan penggunaan masing-masing.  
(6/100)
  - (b) Bezakan **tiga (3)** sifat BLAST, FASTA dan algoritma penjajaran berasaskan pengaturcaraan dinamik. Bandingkan mereka berdasarkan kepada:
    - fokus utama algoritma di atas,
    - bagaimana perbandingan dilakukan oleh algoritma,
    - jenis parameter yang terlibat dalam penjajaran.  
(9/100)

4. Penjajaran Jujukan Berbilang (MSA) tidak boleh dikendalikan secara cekap menggunakan pengaturcaraan dinamik.
- (a) Pilih **dua (2)** kaedah MSA yang sedia ada dan bandingkan keduanya dari segi parameter yang terlibat dan proses penjajaran. (10/100)
- (b) Bincangkan **lima (5)** proses dalam talian paip analisis protein automatik. (10/100)
5. Tujuan filogeni adalah untuk membina semula sejarah kehidupan dan menerangkan kepelbagaian makhluk yang hidup. Filogenetik adalah sejenis filogeni yang bergantung kepada perbandingan gen sama yang datang dari beberapa spesies untuk membina semula pepohon genealogik spesies ini.
- (a) Bincangkan **tiga (3)** sebab utama kenapa anda menggunakan filogenetik. (6/100)
- (b) Apakah **empat (4)** bahan utama yang akan dipertimbangkan apabila kaedah berasaskan jarak digunakan untuk mengira pepohon filogenetik. (9/100)
6. Perl adalah bahasa pengaturcaraan yang popular yang banyak digunakan dalam bidang seperti Bioinformatik dan pengaturcaraan web. Perl telah menjadi popular dalam kalangan ahli-ahli biologi kerana ia sesuai untuk beberapa tugas bioinformatik.
- (a) Apakah ciri-ciri istimewa Perl yang menjadikan ia pilihan popular bagi aplikasi bioinformatik? (5/100)
- (b) Tulis program Perl untuk mencantum dua urutan DNA. Hasil program ini adalah seperti berikut:
- Here are the original two DNA fragments:  
ACGGGAGGACGGGAAAATTACTACGGCATTAGC  
ATAGTGCCGTGAGAGTGATGTAGTA  
Here is the concatenation of the first two fragments  
(version 1):  
ACGGGAGGACGGGAAAATTACTACGGCATTAGCATAGTGCCGTGAGAGTGATGTAGTA  
Here is the concatenation of the first two fragments  
(version 2):  
ACGGGAGGACGGGAAAATTACTACGGCATTAGCATAGTGCCGTGAGAGTGATGTAGTA  
Here is the concatenation of the first two fragments  
(version 3):  
ACGGGAGGACGGGAAAATTACTACGGCATTAGCATAGTGCCGTGAGAGTGATGTAGTA
- (10/100)