



UNIVERSITI SAINS MALAYSIA

First Semester Examination
2016/2017 Academic Session

December 2016 / January 2017

CIT562 – Bioinformatics Computing
[Perkomputeran Bioinformatik]

Duration : 2 hours
[Masa : 2 jam]

INSTRUCTIONS TO CANDIDATE:
[ARAHAN KEPADA CALON:]

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

*[Sila pastikan bahawa kertas peperiksaan ini mengandungi **EMPAT** soalan di dalam **EMPAT** 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) Differentiate between Genetics and Genomics.

Bezakan antara Genetik dan Genomic.

(6/100)

- (b) Protein exists in **four (4)** forms. Describe these forms and the processes that produced them.

Protein wujud dalam empat (4) bentuk. Huraikan bentuk-bentuk ini serta proses-proses yang menghasilkan mereka.

(8/100)

- (c) Name **three (3)** databases and their respective continents that collaborate to form the International Nucleotide Sequence Database Collaboration.

Namakan tiga (3) pangkalan data dan benua masing-masing yang bekerjasama untuk membentuk Kolaborasi Pangkalan Data Jujukan Nukleotida Antarabangsa.

(6/100)

2. (a) Differentiate between Needleman-Wunsch and Smith-Waterman algorithms.

Bezakan antara algoritma Needleman-Wunsch and Smith-Waterman.

(6/100)

- (b) Given two sequences $S = \{C, A, T, T, C, A, C\}$ and $T = \{C, T, C, G, C, A, G, C\}$, show how

Diberikan dua jujukan $S = \{C, A, T, T, C, A, C\}$ dan $T = \{C, T, C, G, C, A, G, C\}$, tunjukkan bagaimana

- (i) the Needleman-Wunsch Algorithm works to align them if the scoring conditions are +10 for match, -2 for mismatch and -5 for space/gap.

Algoritma Needleman-Wunsch bekerja untuk menjajarkan mereka jika syarat pemarkahan adalah +10 untuk padanan, -2 untuk ketidakpadanan dan -5 untuk jarak/jurang.

(10/100)

- (ii) the Smith-Waterman Algorithm works to align them if the scoring conditions are +1 for match, -1 for mismatch and -1 for space/gap.

Algoritma Smith-Waterman bekerja untuk menjajarkan mereka jika syarat pemarkahan adalah +1 untuk padanan, -1 untuk ketidakpadanan dan -1 untuk jarak/jurang.

(10/100)

- (c) A simple alignment scoring scheme is given as follows:

Diberi skema pemarkahan mudah untuk penjajaran seperti berikut:

Match: +1

Padanan: +1

Mismatch: 0

Ketidakpadanan: 0

Gap Creation Penalty: -2

Penalti penciptaan jurang:-2

Gap Extension Penalty: -2

Penalti lanjutan jurang:-2

Calculate the alignment scores for the three alignments below.

Kira markah penjajaran untuk ketiga-tiga penjajaran di bawah.

(i) A A T C T A T A
A A G - A T - A

(2/100)

(ii) A A T C T A T A
A A - G - A T A

(2/100)

(iii) A A T C T A T A
A A - - G A T A

(2/100)

3. (a) What is Multiple Sequence Alignment (MSA) and why is it needed?

Apakah Penjajaran Jujukan Berganda (MSA) dan kenapa ia diperlukan?

(6/100)

- (b) Describe how MSA is done by CUSTALW.

Huraikan bagaimana MSA dilakukan oleh CLUSTALW.

(6/100)

- (c) Discuss progressive alignment by stating its definition, problems, consequences and solutions.

Bincangkan penjajaran progresif dengan menyatakan definisi, masalah, akibat dan penyelesaiannya.

(14/100)

4. (a) Describe phylogenetic analysis (Hint: evolutionary, tree, branch, nodes).

Huraikan analisis filogenetik (Petunjuk: evolusi, pohon, cabang, nod).

(6/100)

- (b) There are **three (3)** categories of phylogenetic tree construction methods. Name these categories and state conditions that will lead to the use of methods from each of these categories.

*Terdapat **tiga (3)** kategori kaedah pembinaan pohon filogenetik. Namakan kategori-kategori ini dan nyatakan keadaan yang akan membawa kepada penggunaan kaedah-kaedah dari setiap satu kategori ini.*

(8/100)

- (c) Perl and Python are two programming languages that are widely used in Bioinformatics.

Perl dan Python adalah dua bahasa pengaturcaraan yang digunakan secara meluas dalam Bioinformatik.

- (i) State the advantage of Python over Perl.

Nyatakan kelebihan Python berbanding Perl.

(4/100)

- (ii) Python has APIs that were developed specifically for Bioinformatics application. Name **four (4)** of these APIs.

*Python mempunyai API yang telah dibina khas untuk aplikasi Bioinformatik. Namakan **empat (4)** API tersebut.*

(4/100)