



Second Semester Examination
2024/2025 Academic Session

July/August 2025

BOE202 – Introduction to Bioinformatic
[Pengenalan kepada Bioinformatik]

Duration: 2 hours
[Masa: 2 jam]

Please ensure that this examination paper contains SIX (6) printed pages before you begin the examination.

[Sila pastikan bahawa kertas peperiksaan ini mengandungi ENAM (6) muka surat yang bercetak sebelum anda memulakan peperiksaan ini.]

Instructions: Answer **FOUR (4)** out of **FIVE (5)** questions, in English or Bahasa Malaysia. Each question carries 25 marks.

[Arahan: Jawab **EMPAT (4)** daripada **LIMA (5)** soalan yang diberikan dalam Bahasa Inggeris atau Bahasa Malaysia. *Tiap-tiap soalan bernilai 25 markah.]*

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].

Should any candidate be caught cheating or in possession of materials not authorised to be brought into the Examination Hall during the examination, appropriate disciplinary action will be taken against the candidate concerned. In the event a candidate is found guilty of cheating, he/she can be expelled from the University.

[Sekiranya seseorang calon didapati meniru/menipu ataupun mempunyai bahan-bahan yang tidak dibenarkan dibawa ke dalam Dewan Peperiksaan semasa peperiksaan diadakan, tindakan disiplin yang sewajarnya akan dikenakan terhadap calon berkenaan. Jika seseorang calon didapati bersalah meniru/menipu beliau boleh disingkir dari Universiti.]

1. [a] Explain bioinformatics and its significance in biology, focusing on its applications and contributions to genomics and proteomics.
[Terangkan bioinformatik dan kepentingannya dalam biologi, dengan memberi tumpuan kepada aplikasi dan sumbangannya terhadap genomik dan proteomik.]

(10 marks / 10 markah)

- [b] Develop the comprehensive methodology for identifying and classifying newly discovered insect species during scientific expeditions in Carry Island, emphasizing the integration of data retrieval and analysis from various online biological databases, specifically utilizing the NCBI database for genomic and taxonomic information.

[Bangunkan metodologi komprehensif untuk mengenal pasti dan pengkelasan spesies serangga baharu yang ditemui semasa ekspedisi saintifik di Pulau Carry, dengan menekankan integrasi pengambilan dan analisis data daripada pelbagai pangkalan data biologi dalam talian, khususnya menggunakan pangkalan data NCBI untuk maklumat genomik dan taksonomi.]

(15 marks / 15 markah)

2. [a] Discuss the role of biological databases in evolutionary biology research. Include examples of how these databases facilitate studies on phylogenetics, species diversity, and evolutionary patterns.

[Bincangkan peranan pangkalan data biologi dalam penyelidikan biologi evolusi. Sertakan contoh bagaimana pangkalan data ini memudahkan kajian mengenai filogenetik, kepelbagaian spesies, dan corak evolusi.]

(10 marks / 10 markah)

- [b] Given the following distance matrix for **FIVE (5)** species (A, B, C, D, E) in Table 1, construct a phylogenetic tree using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) method. Show all your calculations and label the branch lengths.

*[Diberikan matriks jarak bagi **LIMA (5)** spesies (A, B, C, D, E) dalam Jadual 1, bina pokok filogenetik menggunakan Kaedah Kelompok Berpasangan Tak Berpemberat Dengan Min Aritmetik (UPGMA). Tunjukkan semua pengiraan anda dan labelkan panjang cabang.]*

(15 marks / 15 markah)

Table 1
[Jadual 1]

	A	B	C	D	E
A	0	4	6	8	10
B	4	0	6	8	10
C	6	6	0	8	10
D	8	8	8	0	10
E	10	10	10	10	0

3. [a] Elaborate **TWO (2)** types of sequence alignment algorithms and illustrate sequence alignments to show the function for each type.
 [Huraikan **DUA (2)** jenis algoritma penjajaran jujukan dan lakarkan penjajaran jujukan untuk menunjukkan fungsi bagi setiap jenis.]

(10 marks / 10 markah)

- [b] In a basic gene prediction task using a Hidden Markov Model (HMM), a DNA sequence is analysed to find out if each nucleotide comes from an exon (E) or an intron (I). The sequence starts in the exon state. The transition probabilities between states are shown in Figure 1. The emission probability for both A and G is 0.25 in exons, and 0.4 in introns. Given the observed sequence A G G, identify all possible state paths. Then, calculate the total probability for each path using the initial state, transition, and emission values. Finally, choose the most likely path and explain your answer.

[Dalam satu tugas ramalan gen menggunakan Model Tersembunyi Markov (HMM), satu jujukan DNA dianalisis untuk menentukan sama ada setiap nukleotida berasal daripada kawasan ekson (E) atau intron (I). Jujukan ini bermula dari keadaan ekson. Kebarangkalian peralihan antara keadaan ditunjukkan dalam Rajah 1. Kebarangkalian pengeluaran bagi A dan G ialah 0.25 dalam ekson, dan 0.4 dalam intron. Diberikan jujukan yang diperhati A G G, kenal pasti semua laluan keadaan yang mungkin. Seterusnya, kira jumlah kebarangkalian bagi setiap laluan menggunakan nilai keadaan awal, peralihan, dan pengeluaran. Akhir sekali, pilih laluan yang paling mungkin dan jelaskan jawapan anda.]

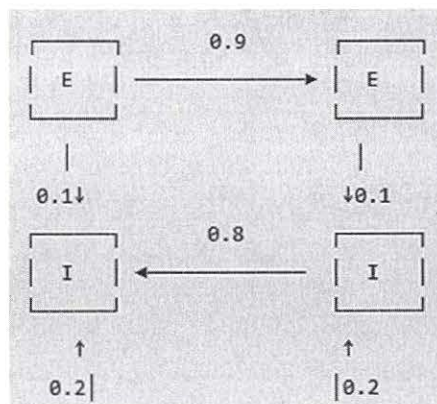


Figure 1
[Rajah 1]

(15 marks / 15 markah)

4. [a] Unweighted Pair Group Method with Arithmetic Mean (UPGMA) and Neighbour Joining (NJ) are two methods for constructing phylogenetic trees. Discuss their differences in terms of algorithmic approach, assumptions, computational complexity, and suitability for different types of data.

[Kaedah Kelompok Berpasangan Tak Berpemberat Dengan Min Aritmetik (UPGMA) dan Penyambungan Bersebelahan (NJ) adalah dua kaedah untuk membina pokok filogenetik. Bincangkan perbezaannya dari segi pendekatan algoritma, andaian, kompleksiti komputasi, dan kesesuaian untuk jenis data yang berbeza.]

(10 marks / 10 markah)

- [b] Protein alignment typically provides better evolutionary insights than DNA alignment, however, DNA alignment offers advantages in specific contexts. Identify **THREE (3)** reasons why protein alignment is generally preferred and provide **TWO (2)** scenarios where DNA alignment proves more beneficial.

*[Penjajaran protein biasanya memberikan pemahaman evolusi yang lebih baik berbanding penjajaran DNA, walau bagaimanapun, penjajaran DNA menawarkan kelebihan dalam konteks tertentu. Kenal pasti **TIGA (3)** sebab mengapa penjajaran protein umumnya lebih diutamakan dan berikan **DUA (2)** senario di mana penjajaran DNA terbukti lebih bermanfaat.]*

(15 marks / 15 markah)

5. [a] Explain the role of neutral mutations and natural selection on molecular evolution.
[Terangkan peranan mutasi neutral dan pemilihan semula jadi terhadap evolusi molekul.]

(10 marks / 10 markah)

- [b] Write **THREE (3)** main important reasons of ligand binding prediction. Explain your answers. Provide **ONE (1)** example of a ligand binding prediction web server and describe its function.
*[Tuliskan **TIGA (3)** sebab utama kepentingan ramalan pengikatan ligan. Terangkan jawapan anda. Berikan **SATU (1)** contoh pelayan web ramalan pengikatan ligan dan terangkan fungsinya.]*

(15 marks / 15 markah)