

考生編號_____

分數_____

2022 年中華民國生物奧林匹亞國手選拔營實作試題

第 2 試場實作

生物資訊學實作—A

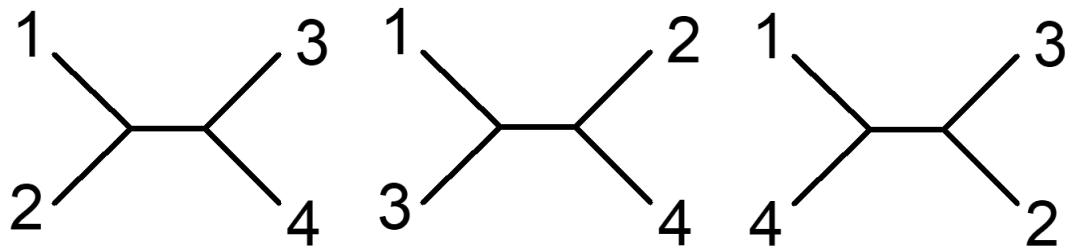
項目	數量
筆電(已安裝 MEGA-11 程式)	1 台
筆電桌面-存放生物資訊學實作 A 資料夾(內含 heat shock protein 20.fas 與 heat shock protein 20.meg 二個檔案)	1 個

※ 請注意：

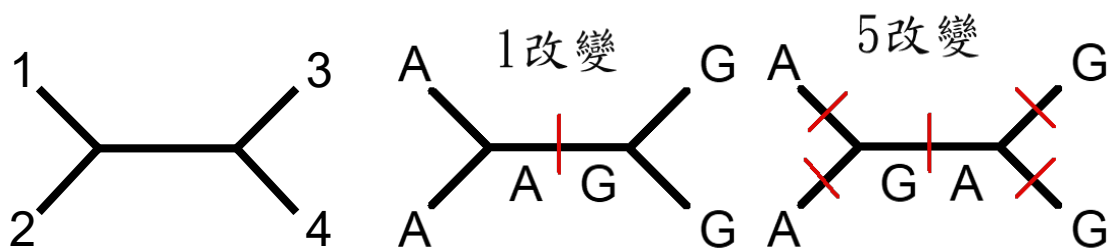
1. 請核對試卷上的考號是否和你的考號相同。
2. 本試卷(含封面、試題卷)共 5 頁，於交卷時全部繳回。
3. 本試場含實作 A 及實作 B，作答時間共為 90 分鐘
4. 請於本卷上作答，試題答案可寫在試卷背面，但請註明並標上題號。

第一題：

親緣關係樹是用來顯示生物物種演化關係的圖形表示，親緣關係樹所建構的物種演化關係是假說，親緣關係樹中的分枝模式顯示出物種或其他群體如何從一系列共同祖先演化而來。現階段，大多數親緣關係樹都是依據分子數據構建的，例如 DNA 或蛋白質序列。一般構建親緣關係樹需要四個不同的步驟：步驟 1、識別和獲取一組同源 DNA 或蛋白質序列；步驟 2、對齊這些序列；步驟 3、從對齊的序列估算樹型；步驟 4、以清楚表達相關信息的方式呈現此親緣關係樹。而在建構親緣關係樹的方法學而言，其中一種會運用到最大簡約原則 (Maximum Parsimony)，此方法是一種基於特徵字符的方法，經由最小化步驟解釋分配在分枝上的給定數據集所需的演化步驟總數來推斷親緣關係樹。若是使用 DNA 序列資料，最大簡約關係樹的建構是基於遺傳密碼變化步驟最少為最佳。若以四物種的 DNA 序列來建構最大簡約關係樹，依據可能的物種間分枝關係推測，可以推測有三種可能的無根樹(unrooted trees)如下：



然後將具有差異的鹼基標示於樹狀圖上，例如物種 1 和 2 的某一特定鹼基為 A，物種 3 和 4 的此鹼基為 G，標於下方的無根樹可能為 1 改變或是 5 改變，依據最大簡約原則，則 1 改變是最符合最大簡約原則的。

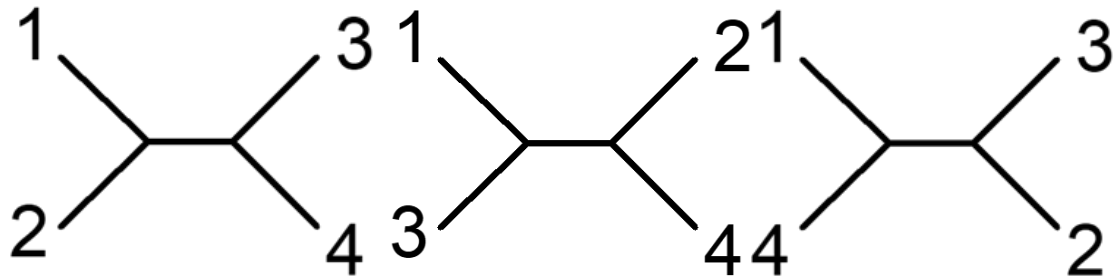


若現今有一組序列資料，共有 4 物種，DNA 序列長度為 10 鹼基如下：

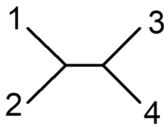
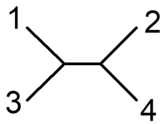
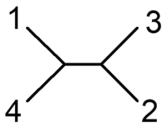
	1	2	3	4	5	6	7	8	9	10
物種 1	A	G	G	G	T	A	A	C	T	G
物種 2	A	C	G	A	T	T	A	T	T	A
物種 3	A	T	A	A	T	T	G	T	C	T
物種 4	A	A	T	G	T	T	G	T	C	G

依據上述資料，回答下列問題：

1. 請依據上述原理與方法，在下方三種可能的親緣關係樹上，一個一個鹼基來逐步建構出可能的最大簡約關係樹。(每個樹型繪製 3 分，共計 9 分)



2. 請依據上述原理與方法，在下方的表格中，填入每一個鹼基在三種可能的親緣關係樹上，所需要的改變次數。(每單格 0.5 分，共計 15 分)

需要步驟	鹼基 1	鹼基 2	鹼基 3	鹼基 4	鹼基 5	鹼基 6	鹼基 7	鹼基 8	鹼基 9	鹼基 10
										
										
										

3. 由上述資料，推測建構的最大簡約關係樹為哪一個無根樹，請在下方繪出樹型 (2 分)，寫出所需要的步驟為多少次鹼基改變(1 分)？

4. 由上述所得到的最大簡約關係樹，推測出哪幾個鹼基的資訊是具有分辨能力的，請寫出鹼基號碼？(2 分)

第二題：

生物面對熱逆境時，會啟動熱休克蛋白基因大量產生熱休克蛋白，可協助維持蛋白質正常摺疊，此種蛋白質在各種生物細胞中多能發現其存在，科學家在人(*Homo sapiens*)、家鼠(*Rattus*)、小家鼠(*Mus musculus*)和黃牛(*Bos taurus*)找到一種分子量為 20kDa 的熱休克蛋白，在經由分子生物學方式，找到牠們的基因片段，長度約 254 鹼基，已知是帶有氨基酸序列的基因片段，但是卻沒有得到全長，此序列檔檔名為 heat shock protein 20.fas 或 heat shock protein 20.meg，請依據下面題目設定，使用 MEGA-11 程式進行下列分析。

1. 在 4 條 20kDa 的熱休克蛋白 DNA 序列，此片段的保守(Conserved)核苷酸位點所占百分比為多少？變異(Variable)核苷酸位點所占百分比為多少？Parsimony information 核苷酸位點所占百分比為多少？Singleton 核苷酸位點所占百分比為多少？(各 1.5 分，共 6 分)

項目	答案
Conserved(C):	
變異(Variable)(V):	
Parsimony information(P):	
Singleton(S):	

2. 在 4 條 20kDa 的熱休克蛋白 DNA 序列，因為是已知是帶有氨基酸序列的基因片段，請使用 MEGA-11 程式的相關功能，將 DNA 序列轉譯可能的氨基酸序列，並將前 10 個氨基酸序列填入下方表格 (每格 0.375 分，共 15 分)

物種	1	2	3	4	5	6	7	8	9	10
人										
家鼠										
小家鼠										
黃牛										

考生編號_____

分數_____

2022 年中華民國生物奧林匹亞國手選拔營實作試題

第 2 試場實作

生物資訊學實作—B

項目	數量
筆電	1 台
筆電桌面-存放生物資訊學實作 B 資料夾 Q1 資料夾：NCBI Vectorbuilder Q1_sequence.txt Q2 資料夾：BioEdit Expasy Reverse Complement Q2-1_sequence.txt Q2-2_sequence.txt Q2-3_sequence.txt	1 個

※ 請注意：

1. 請核對試卷上的考號是否和你的考號相同。
2. 本試卷(含封面、試題卷)共 5 頁，於交卷時全部繳回。
3. 本試場含實作 A 及實作 B，作答時間共為 90 分鐘
4. 請於本卷上作答，試題答案可寫在試卷背面，但請註明並標上題號。
5. 請注意考試中僅能使用資料夾所提供之網站及生物資訊軟體。其他軟體或網站一律禁止，經發現一律扣分。

第一題、NCBI-BLAST (Basic Local Alignment Search Tool)為基因序列比對之工具，將 DNA 或蛋白質序列與 NCBI 資料庫中的序列進行比對，得到相似序列並辨認此序列為哪個物種；鳥嘌呤(G)與胞嘧啶(C)所占全序列之比率稱為 GC 含量，GC 含量會影響 DNA 穩定性與 mRNA 二級結構等，透過 Vectorbuilder 計算能得到序列之 GC 含量與長度。日前，研究人員得到一筆 Q1_sequence.txt 未知 DNA 序列，欲知道此序列是來自什麼物種、其 GC 含量與序列長度。請同學利用 Q1 資料夾中的 NCBI 與 Vectorbuilder 分析 Q1_sequence.txt 並回答下列問題。

問題一、以 NCBI-BLAST 經序列比對辨認其物種，請問 Q1_sequence.txt 的序列是哪個物種 (請寫下學名)? (5 分)

問題二、承上題，此物種是屬於生物分類學上哪一界? (5 分)

問題三、利用 Vectorbuilder 分析 Q1_sequence.txt 序列的 GC 含量和鹼基對(bp)分別是多少(Window size 為 50)? (4 分)

問題四、承上題，(1)最高的 GC 含量在多少鹼基區間；(2)最低的 GC 含量在多少鹼基區間? (以 1000 鹼基區間為單位)(4 分)

第二題、研究人員以 PCR 技術將某物種 DNA 片段進行放大後，經定序後得到 Q2-1_sequence.txt、Q2-2_sequence.txt 與 Q2-3_sequence.txt，Q2-1_sequence.txt 為野生型的編碼序列(coding sequence, CDS)；Q2-2_sequence.txt 與 Q2-3_sequence.txt 為突變型的編碼序列，其中 Q2-1_sequence.txt 與 Q2-3_sequence.txt 序列方向為 5'端到 3'端，Q2-2_sequence.txt 序列方向為 3'端到 5'端。研究人員先利用 Reverse Complement 改變序列方向，再以 BioEdit 進行核酸序列比對並得知序列中哪個位點發生變異，接著以 ExPASy 將 DNA 序列轉譯成胺基酸，最後以 BioEdit 判斷胺基酸序列突變位點是無義突變 (nonsense mutation)、錯義突變 (missense mutation)或是沉默突變 (silent mutation)，請同學將 Q2 資料夾中這三個序列 Q2-1_sequence.txt、Q2-2_sequence.txt 與 Q2-3_sequence.txt，利用 Q2 資料夾中的 Reverse Complement、ExPASy 與 BioEdit 進行分析並回答下列問題。

問題五、三個序列進行核酸序列比對，列出其變異位點 (答案呈現方式：第 10 個位點 A→T)。(16 分)

問題六、承上題，三個序列進行胺基酸序列比對，說明第某個胺基酸突變位點的轉換，是屬於何種突變方式 (答案呈現方式：第 10 個位點 M→L；何種突變方式)? (16 分)