

考生編號\_\_\_\_\_

分數\_\_\_\_\_

2020 年中華民國生物奧林匹亞國手選拔營實作試題

第 2 試場

生物資訊學實作

| 項目                    | 數量  |
|-----------------------|-----|
| 筆電(已安裝 MEGA-X 程式)     | 1 台 |
| 筆電桌面上有 COVID-2019 序列檔 | 1 個 |
| 筆電桌面上有學生考號的資料夾        | 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 4 頁，於交卷時全部繳回。
3. 本試場作答時間共為 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。

新冠肺炎 COVID-2019 在 2019 年至-2020 年全球大爆發，已經導致全球 300 萬人次感染，在 COVID-2019.fas 檔案中，包含 17 條由病人中獲得的檢體中定序結果和 2 條蝙蝠的病毒序列片段，使用 MEGA-X 程式進行下列分析。

1. 在 19 條病毒的序列資料經由 Clustal W 進行序列排列(Alignment)後，所呈現的結果，你判斷遭遇到甚麼問題導致此現象產生？(10 分)

2. 去除掉 2 條蝙蝠的病毒序列後，剩下的病人檢體獲得的 17 條新冠肺炎 COVID-2019 序列，重新使用 Clustal W 進行序列排列(Alignment)後，去除掉前端和後端不等長的區域後，核苷酸鏈合長度為多少？Conserved 核苷酸位點為多少？變異(Variable) 核苷酸位點為多少？Parsimony information 核苷酸位點為多少？Singleton 核苷酸位點為多少？(提示:COVID-2019 序列需另存 MEGA 及 FAST 檔，請存至桌面考號的資料夾，檔名「考號-COVID-2019-1.MEG 及考號-COVID-2019-1.FAS」。)(各 2 分，共 10 分)

| 項目                         | 答案 |
|----------------------------|----|
| 核苷酸鏈合長度:                   |    |
| Conserved(C):              |    |
| 變異(Variable)(V):           |    |
| Parsimony information(Pi): |    |
| Singleton(S):              |    |

3. 在獲得的 17 條新冠肺炎 COVID-2019 序列，找到”ATGGAGAGC”到”AAACCTTT”的區域，去除掉其他區域，此片段為新冠病毒的 ORF1a polypeptide 蛋白質的基因片段，此片段的核苷酸鏈合長度為多少？Conserved 核苷酸位點為多少？變異(Variable) 核苷酸位點為多少？Parsimony information 核苷酸位點為多少？Singleton 核苷酸位點為多少？(提示:ORF1a-2019 序列需另存 MEGA 及 FAST 檔，請存至桌面考號的資料夾，檔名「考號-ORF1a-2019-1.MEG 及考號-ORF1a-2019-1.FAS」。)(各 2 分，共 10 分)

| 項目                         | 答案 |
|----------------------------|----|
| 核苷酸鏈合長度:                   |    |
| Conserved(C):              |    |
| 變異(Variable)(V):           |    |
| Parsimony information(Pi): |    |
| Singleton(S):              |    |

4. 新冠病毒的 ORF1a polyprotein 蛋白質的基因片段，經由轉譯後，此片段的胺基酸鏈合長度為多少？Conserved 胺基酸位點為多少？變異(Variable) 胺基酸位點為多少？Parsimony information 胺基酸位點為多少？Singleton 胺基酸位點為多少？(各 2 分，共 10 分)

| 項目                         | 答案 |
|----------------------------|----|
| 胺基酸鏈合長度:                   |    |
| Conserved(C):              |    |
| 變異(Variable)(V):           |    |
| Parsimony information(Pi): |    |
| Singleton(S):              |    |

5. 在獲得的 17 條新冠肺炎 COVID-2019 序列和 ORF1a polyprotein 蛋白質的基因片段的序列的平均 A、G、C、T 百分比各為多少？(各 2 分，共 16 分)

|            | A | G | C | T |
|------------|---|---|---|---|
| COVID-2019 |   |   |   |   |
| ORF1a      |   |   |   |   |

6. 獲得的 17 條新冠肺炎 COVID-2019 序列和 ORF1a polyprotein 蛋白質的基因片段，使用 DNA 序列，依據 BIC (Bayesian Information Criterion)分數結果，選取出最符合的核苷酸取代模型(nucleotide substitution models)前三名各為何種模型？是否有差異？(各 2 分，共 12 分)

|            | 前一名模型 | 前二名模型 | 前三名模型 |
|------------|-------|-------|-------|
| COVID-2019 |       |       |       |
| ORF1a      |       |       |       |

是否有差異？(6 分)

7. 以 ORF1a polyprotein 蛋白質的基因片段，使用胺基酸序列(Amino acid)方式，依據 BIC (Bayesian Information Criterion)分數結果，選取出最符合的核苷酸取代模型(nucleotide substitution models)前三名為何種模型？(各 2 分，共 6 分)

|       | 前一名模型 | 前二名模型 | 前三名模型 |
|-------|-------|-------|-------|
| ORF1a |       |       |       |

8. 以 ORF1a polyprotein 蛋白質的基因片段，分別使用 DNA 序列(Nucleotide)和胺基酸(Amino acid) 序列方式，依據所獲得的最佳核苷酸取代模型(nucleotide substitution models)和最佳胺基酸取代模型(amino acid substitution models)，分別使用最大似然度法(Maximum Likelihood method)方式進行樹狀圖分析，以 Bootstrap 方法重覆分析 50 次，建構出 DNA 序列(Nucleotide)和胺基酸序列(Amino acid)的樹狀圖，請將 DNA 樹型存成 PDF 檔案並用括號方式將 DNA 序列的分群樣本列出？(共 5 分) 請將胺基酸樹型存成 PDF 檔案並用括號方式將胺基酸序列的分群樣本列出？(共 5 分) 請依據上述二種樹狀圖之結果進行推論？(10 分)