

## 2023 年【科學探究競賽-這樣教我就懂】

普高組 成果報告表單

題目名稱：如何推測新型病毒？

### 一、摘要

過去逾兩年半的時間裡，全球都因為新冠疫情 Covid-19 肆虐，大幅度影響了我們的生活方式、習慣。如此大規模的病毒感染事件直至今日仍在世界各地延燒。除了人民在經過此次大規模疫情事件後應更有防疫的意識外，政府在制定政策時也更該以此為戒，事前做好完善的防疫規劃，超前部屬準備對抗疫情所需的物資，加上公共衛生宣導等等，以因應不可測的未來。

本研究透過 Model3. 的 *JC* 模型 (Jukes and Cantor's)，搭配 20 筆直系母子兩代的病原體傳播進化速度資料，成功推算出該未知病原體的基因序列長度範圍約在 187 ~ 444 個鹼基之間。

### 二、探究題目與動機

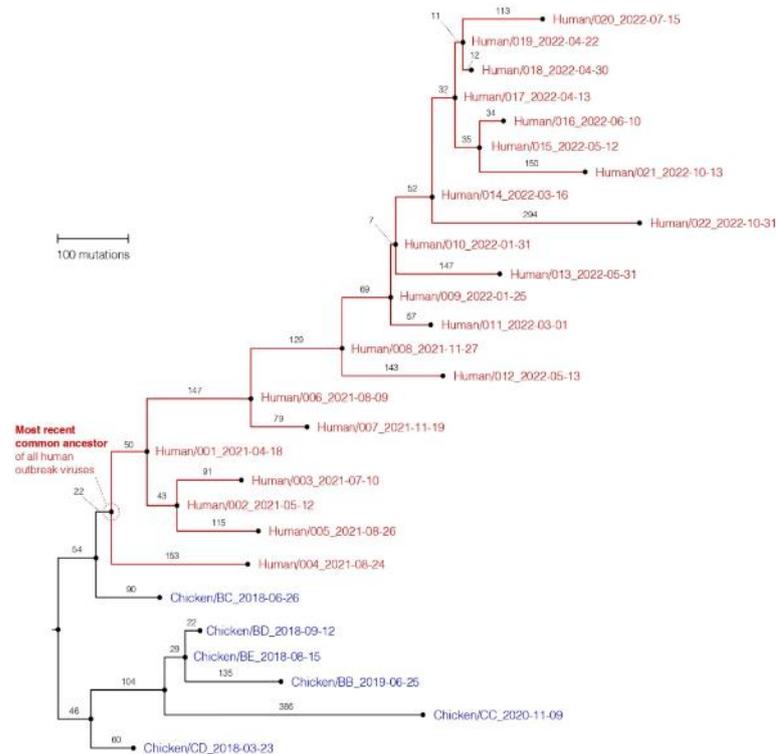


圖 1.人類患者和受感染雞的病毒序列建立的系統發育樹

過去逾三年的時間裡，全球都因為新冠疫情 Covid-19 肆虐，大幅度影響了我們生活的方式與習慣。如此大規模的病毒感染事件直至今日仍在世界各地延燒。除了人民在經過此次大規模疫情事件後應更有防疫的意識外，政府在制定政策時也更該以此為戒，事前做

好完善的防疫規劃，超前部屬準備對抗疫情所需的物資，加上公共衛生宣導等等，以因應不可測的未來。

根據 IMMC 2023 中華賽區秋季賽之 B 題所提供之數據資料 ( 如圖 1. ) 進行研究及延伸。

### 三、探究目的與假設

- 1、 建立數學建模並推算未知病原體在人類群體中傳播的進化速度是多少。(即以每年的突變量為單位的突變率是多少)
- 2、 推算該病原體何時開始在人類中爆發，以及比首個被發現的病例早多少天。
- 3、 探討該病毒可能為何種已知病原體。

### 四、探究方法與驗證步驟

#### (一) 研究過程

##### 1、 研究目的一

##### (1) Model 1. : 直系母子兩代平均法

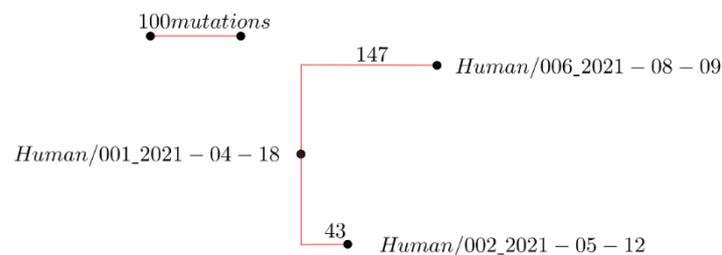


圖 2.Model1. 直系母子兩代數據示意圖

點 Human/001 ( 2021-04-18 ) - Human/002 ( 2021-05-12 )

點 Human/001 ( 2021-04-18 ) - Human/006 ( 2021-08-09 )

以此類推...。其各點求解結果如下表 1。

表 1.Model1.求解

母代	子代	突變數	時間(天數)	突變數/時間
Human/001	Human/002	43	24	1.791667
	Human/006	147	113	1.300885
Human/002	Human/003	91	59	1.542373
	Human/005	115	106	1.084906
Human/006	Human/008	79	102	0.774510
	Human/007	129	110	1.172727
Human/008	Human/009	69	60	1.150000
	Human/012	143	168	0.851190
Human/009	Human/010	7	6	1.166667
	Human/011	57	35	1.628571
Human/010	Human/014	147	120	1.225000
	Human/013	52	44	1.181818
Human/014	Human/017	32	28	1.142857
	Human/022	294	229	1.283843
Human/017	Human/019	35	29	1.206897
	Human/015	11	8	1.375000
Human/015	Human/016	34	29	1.172414
	Human/021	150	154	0.974026
Human/019	Human/020	113	84	1.345238
	Human/018	12	8	1.500000
平均一天改變量:				1.243529

由表 1.可以看到，20 個點的母子兩代間的平均傳播進化速度為每天 1.243529 個突變。如轉換單位為一年之平均，則為 $1.243529 \times 365 = 453.999085$  (突變數/年)

Model 2.：直系母子第  $N$  與  $N + 2$  兩代平均法

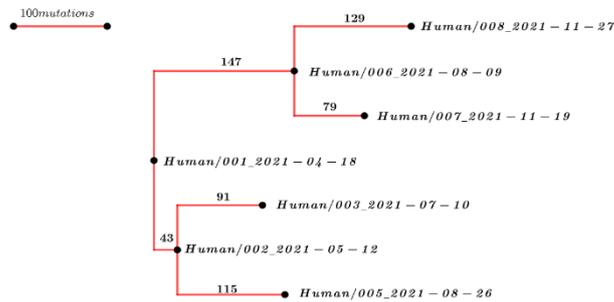


圖 3. Model2.直系母子第  $N$  與  $N + 2$  兩代數據示意圖

點 Human/001 ( 2021-04-18 ) - Human/003 ( 2021-07-10 )

點 Human/001 ( 2021-04-18 ) - Human/005 ( 2021-08-26 )

以此類推...。其各點求解結果如下表 2.。

由表 2.可以看到，18 個點的母代 ( 第  $N$  代 ) 與子代 ( 第  $N + 2$  ) 間的平均傳播進化速度為每天 1.213479 個突變。如轉換單位為一年之平均，則為

$1.213479 \times 365 = 442.919835$  (突變數 / 年)。

(2) Model 3.：JC 模型

表 2.Model2.求解

母代	子代	突變數	時間(天數)	突變數/時間
Human/001	Human/003	134	83	1.614458
	Human/005	158	130	1.215385
	Human/008	276	223	1.237668
	Human/007	226	215	1.051163
Human/008	Human/010	76	65	1.169231
	Human/011	126	94	1.340426
Human/010	Human/022	346	273	1.267399
	Human/017	84	72	1.166667
Human/017	Human/016	69	58	1.189655
	Human/021	185	183	1.010929
	Human/020	124	93	1.333333
	Human/018	23	17	1.352941
Human/006	Human/009	198	169	1.171598
	Human/012	272	277	0.981949
Human/009	Human/014	59	50	1.180000
	Human/013	154	126	1.222222
Human/014	Human/015	67	57	1.175439
	Human/019	43	37	1.162162
平均一天改變量：				1.213479

利用 JC ( Jukes and cantor's ) 模型公式，計算每個點的傳播進化速度 ( 單位是突變數/天 )，計算完畢後將 21 個點取算術平均數，再乘以 365 ( 天 ) 用以轉換單位，即得到平均傳播進化速度 ( 突變數/年 )。

1、JC 模型公式：

$$K = -\frac{3}{4} \ln\left(1 - \frac{4}{3}D\right)$$

$$R = \frac{K}{2T(\text{年})}$$

2、*JC* 模型的計算：

透過直系母子兩代的突變數與時間數據，本研究經過 Model1. 的計算得到直系母子兩代平均傳播進化速度（如表 1.）。最後由於 *D* 的計算方式為（突變個數）/（基因序列全長），透過圖 1.數據將突變個數代入後，即可推算出每個點的基因序列全長。此也是本研究的另外一個發現。其各點的基因序列全長反推計算結果如表 3.。

下表 3.中的最後一行是本研究團隊透過 *JC* 模型方法透過各點的突變數與時間數據，反推計算出各點代表的病原體基因序列長度。表 3.中的數據顯示，本團隊估算其基因序列長度約在 187 ~ 444 個鹼基範圍內。唯基因序列中鹼基通常為 3 個一組，意即基因序列長度通常為 3 的倍數，本團隊僅是透過反推估算得出其可能的範圍。

表 3. *JC* 模型計算基因序列長度

母代	子代	突變數	時間(天數)	突變數/時間	<i>JC</i> 模型計算基因序列長度
Human/001	Human/002	43	24	1.791667	212
	Human/006	147	113	1.300885	297
Human/002	Human/003	91	59	1.542373	250
	Human/005	115	106	1.084906	270
Human/006	Human/008	79	102	0.774510	240
	Human/007	129	110	1.172727	229
Human/008	Human/009	69	60	1.150000	190
	Human/012	143	168	0.851190	294
Human/009	Human/010	7	6	1.166667	187
	Human/011	57	35	1.628571	223
Human/010	Human/014	147	120	1.225000	298
	Human/013	52	44	1.181818	219
Human/014	Human/017	32	28	1.142857	205
	Human/022	294	229	1.283843	444
Human/017	Human/019	35	29	1.206897	207
	Human/015	11	8	1.375000	190
Human/015	Human/016	34	29	1.172414	206
	Human/021	150	154	0.974026	224
Human/019	Human/020	113	84	1.345238	268
	Human/018	12	8	1.500000	191
平均一天改變量:				1.243529	

3、研究目的二

(1) 下圖 4.可知，Most recent common ancestor（此稱為第 0 號感染者）可由

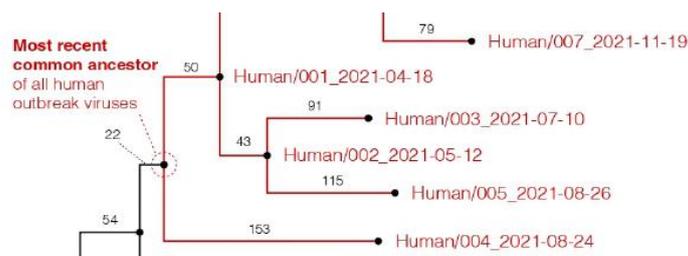


圖 4.推算第 0 號感染者的感染時間

Human/001 的感染時間 2021-04-18 往前回溯及可由 Human/004 的感染時間 2021-08-24 往前回溯。

(2) 依據 Model 1.的計算計算出了未知病原體的平均傳播進化速度分別為每天 1.243529 個突變，由圖 4.知第 0 號感染者到 Human/001 間歷經了 50 個突變數。

$$\frac{50}{1.243529} = 40.2081495 \text{ (天)} \text{，即約 40 天左右。從發現 Human/001 的 2021-04-18}$$

起往前回溯 40 天，即為 2021-03-09。

第 0 號感染者到 Human/004 間歷經了 153 個突變數。

$$\frac{153}{1.243529} = 123.036938 \text{ (天)} \text{，即約 123 天左右。從發現 Human/004 的 2021-08-24}$$

起往前回溯 123 天，即為 2021-04-24。

經過上述計算可推論，新病原體約三月初應就在人類中爆發。而第 0 號感染者比首個被發現的病例早了大約 40 天。

(3) 依據 Model 2.的計算計算出了未知病原體的平均傳播進化速度分別為每天 1.213479 個突變，由圖 4.知第 0 號感染者到 Human/001 間歷經了 50 個突變數。

$$\frac{50}{1.213479} = 41.2038445 \text{ (天)} \text{，即約 41 天左右。從發現 Human/001 的 2021-}$$

04-18 起往前回溯 41 天，即為 2021-03-08。

第 0 號感染者到 Human/004 間歷經了 153 個突變數。

$$\frac{153}{1.213479} = 126.083764 \text{ (天)} \text{，即約 126 天左右。從發現 Human/004 的}$$

2021-08-24 起往前回溯 126 天，即為 2021-04-21。

此數據同樣看來較不合理，產生誤差的理由不再贅述。但經過上述計算仍可推論，新病原體約三月初應就在人類中爆發。而第 0 號感染者比首個被發現的病例早了大約 41 天。

## 五、結論與生活應用

(一) 在只考慮直系母子兩代的 Model1.中，計算出了未知病原體的平均傳播進化速度分別為每天 1.243529 個突變 (或每年 453.999085 個突變)。在考慮直系母子第  $N$  與  $N+2$  兩代的 Model2.中，計算出了未知病原體的平均傳播進化速度分別為每天 1.213479 個突變 (或每年 442.919835 個突變)。

另外，透過 Model3.的  $JC$  模型，搭配 20 筆直系母子兩代的病原體傳播進化速度資料，成功推算出該未知病原體的基因序列長度範圍約在 187 ~ 444 個鹼基之間。

(二) 新病原體約三月初應就在人類中爆發。而第 0 號感染者比首個被發現的病例早了大約 41 天。

(三) 由研究目的一所算出的平均突變速率約 1.24，再從文獻及資料中盡可能找出具有本研究所猜測條件的病毒，並且基因序列長度在 187 到 444 之間，然後將查到的資料經由刪

減之後找到禽流感中的 H7N9 及 H5N1 病毒，基因序列長度分別是 225 及 252 個鹼基，並利用模型所計算出的突變速率與本研究計算出來的最為接近，而 H7N9 及 H5N1 病毒突變速率為 1.03 及 1.24，得到的結果可發現 1.24 最為接近，故猜測該病毒為 H5N1 病毒。

#### 參考資料

[1] 顾红雅-核酸演化的速率(上)(第 8 講)【線上論壇】。 <https://reurl.cc/7j8611>

[2] 顾红雅-核酸演化的速率(下)+KA/KS 比值的意义(第 8 講)【線上論壇】。  
<https://reurl.cc/jREbA1>

[3] 國立科學工藝博物館·生物科技面面觀【電子郵件清單訊息】。取自 <https://reurl.cc/DX2Wze>

[4] 感染的禽流感(H5N1, H7N9 及 H10N8)：過去與現在(中華民國 104 年 4 月第二十五卷二期)。 <https://ssur.cc/qYmFNnw>