2014 年臺灣國際科學展覽會
優勝作品專輯

作品編號　110007

參展科別　電腦科學

作品名稱　由基因序列的親疏設計流感疫苗探討

得獎獎項　大會獎：一等獎

候補作品：1

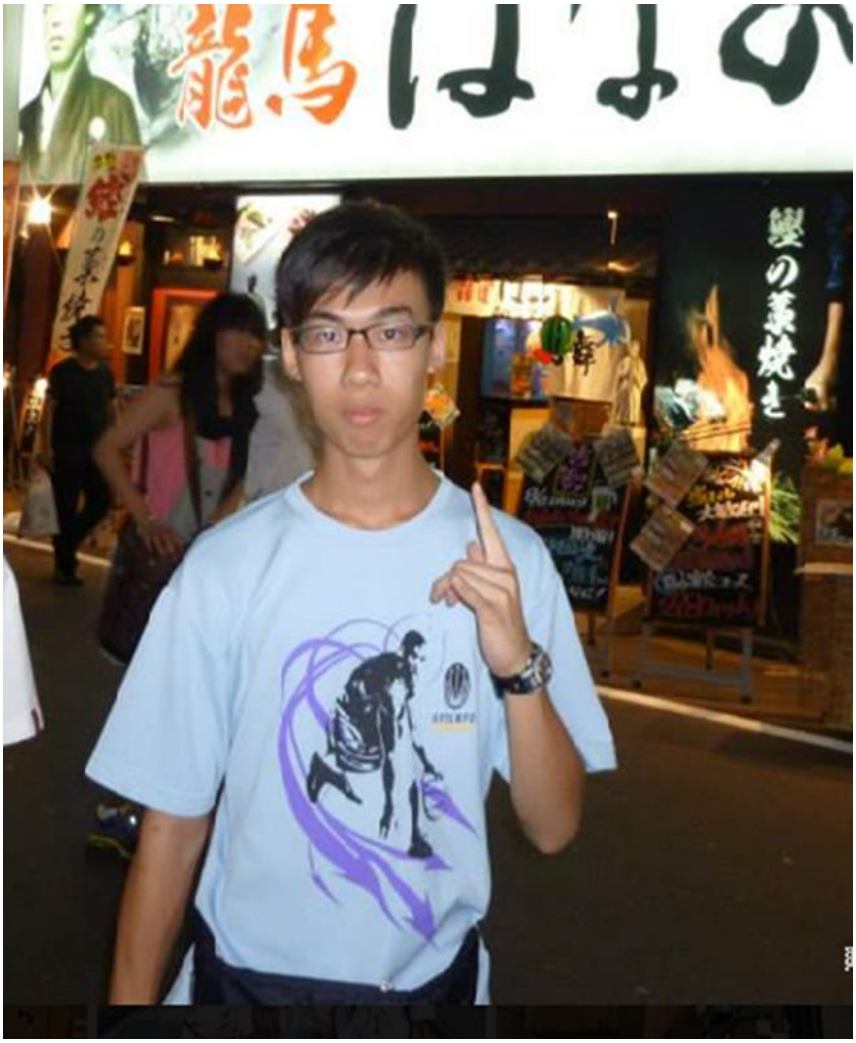
就讀學校　國立中科實驗高級中學

指導教師　胡裕仁、蔡明致

作者姓名　王翰霖、林義傑、溫英華

關鍵字　生物序列、疫苗研發、序列比對

作者簡介



我是王翰霖，目前就讀中科實驗高級中學三年級數理實驗班。家庭有四位成員，相處得非常融洽。爸爸是一位醫生，所以從小就讓我培養了做生物實驗的熱忱與態度。在實驗的過程需要細心與耐心， 因為隨時會有許多奇妙的可能發現，同時也使我認知知識是沒有邊界的。雖然有時會面臨到挫折，但我堅定的意志告訴自己，就算失敗了一百次、一千次，持之以恆，總有一天一定會成功！

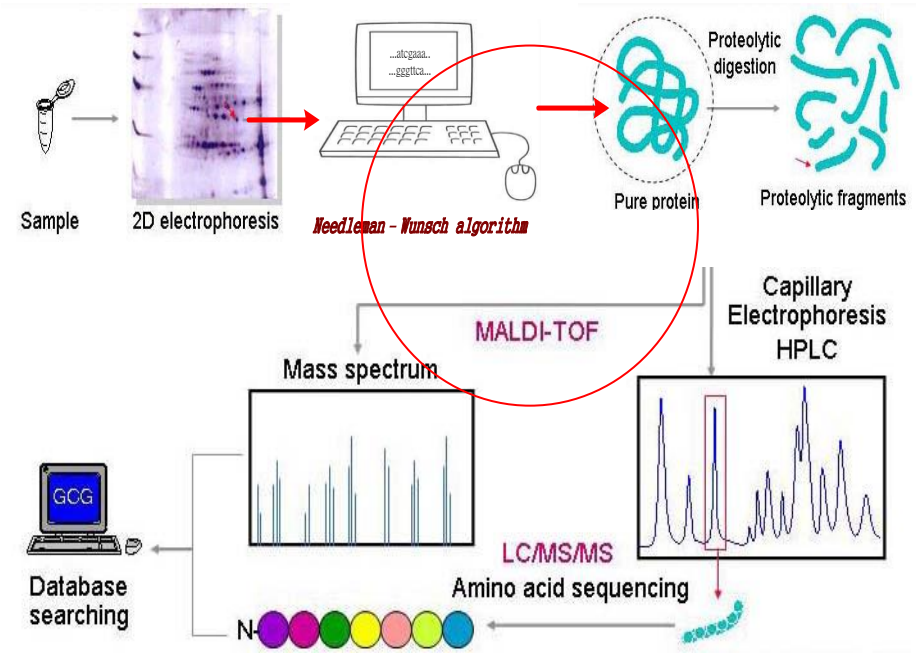
(左 1 我媽　圖中 我本人　圖右 國中老師)

大家好，我叫林義傑，今年剛滿十七歲，從小到大我都非常謝謝爺爺幫我取這個名字，我常常因為名字而在第一印象中使人印象深刻，我也以我的名字為榮。目前我就讀國立中科實驗高中，今年才剛升二年級。我因為非常喜歡科學，所以在剛進學校時，就來參加科展研究了。去年我們小組參加馬來西亞發明展 I-EVENX 榮獲國際組銀牌，這次比賽中我學習了很多，也成長了很多，我也知道我有很多能力還不及別人，所以經由這次比賽使我更加奮發圖強。而在 2013 年的暑假，我跟學長獲選為唯一台灣代表隊可在日本國際研究科學展覽會發表，而我在這次比賽中也認識了許多國家的學生，像是中國、韓國、泰國、新加坡、德國等等，我可以從他們身上學習不同文化，不同語言，也同時擴展了我的國際觀。未來我希望能到世界各國認識不同的朋友，學習更廣泛的知識，因此，我要為未來的我做更多的努力。

我是溫英華，十六歲，就讀國立中科實驗高級中學二年級，我從小就喜歡那些機械裝置，覺得金屬閃亮亮又會動，真是太帥了，於是崇拜工程師等職業，喜歡收集壞掉的電子零件，拼拼湊湊，把玩於五指之間，自得其樂。後來又發現我也喜歡文學、風景，那種內心的感動。於是我平常白日夢的內容，從一片高科技的未來世界，變成光影動人的自然風光，後來中和成人文與自然平衡的和諧畫面。國中時老師們總告訴我們說，高中生活很有趣，有社團，有小論文，甚至做科展，升上高中後，成績顯示喜歡文學並不代表文科成績會比較好，理科給我更多的自信，自然也是興趣所向，喜歡研究、探索，於是自然科學數理是我想走的路。

摘要

本研究主要利用最大概似法及動態規劃演算法來嘗試縮短生醫領域在疫苗研發的時程。透過序列比較的計算方式加速找出病毒序列具有專一性的有效區段。使科學家可以減少盲目測試的實驗。我們期望找出經過電泳之後，可以判斷具有可製造疫苗的最佳生物序列區段。藉由已知流感病毒的基因序列來分析現有流感病毒的演化親緣關係。嘗試由已知流感病毒疫苗來設計未知的流感病毒疫苗之建議(結果如下圖)。



由研究所提方法來研發疫苗流程圖

Abstract

The research mainly applies dynamic programming and MLE method to shorten the antibody R & D timeframe in biomedical field. Through the sequence mapping algorithm to speed up the specificity effective segment in the viral sequence.

To reduce the blind test experiment for the scientist. We hope to identify the optimal biological sequence for antibody production. In traditional, the researchers randomly find out a specific segments. We find that the new antibody manufacturing methods(In figure 1).

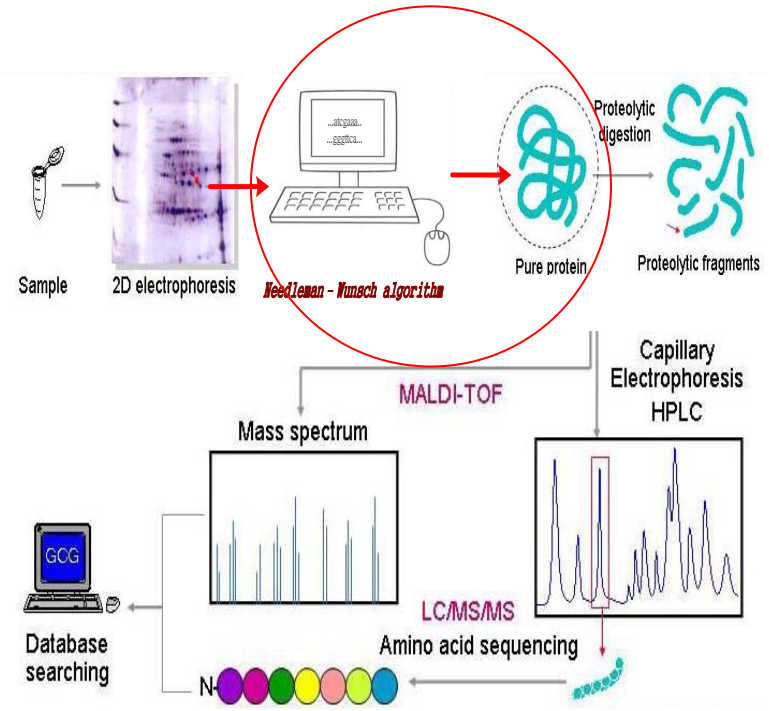


Figure 1 : The Antibody R&D Process Flow Diagram with Proposed Methodology.

In this study, we first use of statistical calculations in the antibody research. We can effectively reduce to randomly seeking might with specificity segments time. Finally, this study greatly enhance the success of the R&D of antibody.

一、　前言(Introduction)

在現代的疫苗研發製造流程中要進行電泳分析及純化蛋白實驗時。如果想要順利找出可當作疫苗的位置序列[1][13]。就像是買樂透期望會中獎一樣無法預期一定會有好結果。雖然存在有固定實驗步驟流程卻沒有系統性的實驗操作模式。因此在實驗室進行疫苗研發製造的實驗過程就如同數學中的機率問題：運氣好的科學家一次實驗就成功，運氣不好的科學家可能要重複好多年仍然是難以成功。

本研究提出如何處理數量龐大的基因體或蛋白質體的計算實驗，並透過「流感病毒」來測試此方法。我們希望找出可以快速且正確地在實驗中來檢定一個病毒是否具有可發展為疫苗的生物序列區段。並且判斷它是否具有發展為疫苗的關鍵位置之可能。而如何順利地找出此區段，正是疫苗研究實驗是否可以成功的一個關鍵步驟[1]。

根據 2013 年 8 月「自然」期刊刊登的研究顯示[19]，科學家在中國活禽市場的雞隻身上,發現可能造成動物感染與死亡的新種禽流感病毒株H7N7,一旦散佈，將出現人類感染案例，可能引發比 H7N9 還嚴重的疫情。

相關研究是由香港大學微生物學系教授管軼領導的國際團隊發現，他們檢驗雞、鴨、鵝等家禽感染 H7N9 病毒狀況時，意外發現變種病毒 H7N7。團隊發現，候鳥把 H7 病毒帶進中國東南部傳染給家鴨，鴨子成了 H7 型禽流感病毒的「大熔爐」,接著再傳染給雞隻，H7 病毒與雞隻病毒相結合之後，產生 H7N9 與 H7N7 病毒，可能在活禽市場傳染給人類[19]。

他們由實驗室的生化分析檢驗發現，這種前所未見的 H7N7 病毒會造成哺乳類動物，例如:雪貂感染嚴重肺炎，由於雪貂是研究人類感染流感病毒的模式生物之一[19]，因此這種新型病毒株可能威脅人類。例如:2003 年荷蘭曾傳出另一種

H7N7 病毒流行，當時造成超過八十起人類感染病例，其中一人死亡，其他人症狀都不嚴重。

另外根據中國大陸國家禽流感參考實驗室陳化蘭研究團隊 2013 年的最新研究顯示[20]，活禽市場是 H7N9 禽流感病毒的源頭，應加強活禽市場管理，防止病毒向養殖場傳播。陳化蘭科研團隊的研究顯示，人感染 H7N9 病毒與同期存在活禽市場上的 H7N9 禽流感病毒高度同源，這是國際間首次從病原學角度，揭示新型 H7N9 流感病毒的來源[20]。

且由 2013 年「科學」期刊報導[20]，新型 H7N9 流感病毒是不同來源病毒的重組病毒，它的 8 個內部基因(PB1、PB2、HA 等...)來源於 H9N2 禽流感病毒，而 HA 和 NA 基因來源尚不明確。另外陳化蘭團隊研究發現，H7N9 病毒的受體結合位點，獲得部分人流感病毒特徵的突變；病毒在感染人後獲得關鍵氨基酸位點的適應性突變，這些可能與病毒對人的感染和致死能力有關[23]。

目前大陸的國家禽流感參考實驗室儲備有 H7 亞型禽流感病毒的疫苗，這個疫苗對新型 H7N9 禽流感病毒的免疫保護效果正在評估，參考實驗室研究人員正在加緊研製，專門針對新型 H7N9 禽流感病毒的疫苗[23]。而本研究率先嘗試以生物序列計算的生物資訊分析方式破解流感病毒序列結構嘗試提出新的疫苗製程。

二、 研究方法(Materials and Methods)

隨著近年來生物學的蓬勃發展，除了生物基因庫的資料已漸趨完備外，解讀基因序列的軟體更是比以往發達，因此我們希望能善用這資源來進一步分析流感病毒間的變異性。本研究主要利用最大概似法及動態規劃演算法來嘗試縮短生醫領域在疫苗研發的時程。透過序列比較的計算方式加速找出病毒序列具有專一性的有效區段。使科學家可以減少盲目測試的實驗。我們期望找出經過電泳之後，可以判斷具有可製造疫苗的最佳生物序列區段。本研究希望由已知流感病毒的基因序列來計算現有流感病毒的演化親緣關係。最後提出由已知流感病毒疫苗來設計未知的流感病毒疫苗之方式。

概似函數(Likelihood Function，L(θ))[3]

$L(\theta) = f(X_1,\cdots,X_n;\theta) = f(x_1;\theta)\cdot f(x_2;\theta)\cdot\cdots\cdot f(x_n;\theta)$，本函數將在參數$\theta$的所有可能取值上，使這個函數最大化。這個使$L(\theta)$可能性最大的$\hat{\theta}$值即被稱為$\theta$的最大概似估計。

最大概似估計

由於生物序列中基因呈現離散資料型態，即假設目前有三組生物基因序列

$A=\{atcatgctacgtcag\}$、$B=\{ctgatcgatgcatgc\}$、$C=\{agtctggatcgagtc\}$、$\theta = \dfrac{f}{n}$。字母f代表在不同基因字串中被觀察到相同字元的次數，n代表各基因字串的總數。

$\theta_{AB} = \dfrac{3}{15}$(A 和 B 序列)、$\theta_{BC} = \dfrac{6}{15}$(B 和 C 序列)、$\theta_{AC} = \dfrac{5}{15}$(A 和 C 序列)。因此可以令$\hat{\theta}=\theta_{BC} = \dfrac{6}{15}$，為概似函數取得最大值，即為$\hat{\theta}$的最大概似估計

最大概似估計法原理

A. 概似函數(Likelihood Function)： $L(\theta)=L(x_1,...,x_n;\theta)=f(X_1,\cdots,X_n;\theta)$

$=f(x_1;\theta)\cdot f(x_2;\theta)\cdot\cdots\cdot f(x_n;\theta)=\prod_{i=1}^{n}f(x_i;\theta)$，$\theta$ 為未知參數的估計對象。

B. 生物序列的最大概似估計量推定計算:

I. 先計算連續函數的 $L(\theta)=\prod_{i=1}^{n}f(x_i;\theta)$，則以 $\frac{d\ln L(\theta)}{d\theta}=0$ 及 $\frac{d^2\ln L(\theta)}{d\theta^2}<0$ 來判斷[87][97]

II. 由於生物序列中討論的核酸序列或胺基酸序列均是為離散的資料型態，因此在判斷最大概似參數時是以雙面逼近法 $\frac{L(\theta)}{L(\theta-1)}\geq 1$；$\frac{L(\theta+1)}{L(\theta)}\leq 1$ 來找出最佳 θ 的估計結果[97]。

最大概似估計(Maximum Likelihood Estimation，MLE)，它是常用的系統生物學計算方法，主要是用來求一個樣本資料集的相關機率密度函數的參數估計，目前在系統生物學的研究中，主要運用在計算生物物種分類的工具，而 ***Likelihood*** 的意思係用已知的資料來作出可能影響所要研究結果的參數的函數，藉以估計參數可能的數值[21][26]。所以在未確定生物物種的親緣關係時，要利用現有物種資料作為參數，以概似函數的生成機率密度函數 $L(\theta)$ 來計算推論[21]。因此以最大概似估計應用於生物序列分析，主要是估計找出使 $L(\theta)$ 的可能性最大的 $\hat{\theta}$ 值，故 $L(\theta)$ 被稱為 θ 的最大概似估計。

演化樹個數計算及其生物涵意[26]

A. 有根樹個數： $N_R=\frac{(2n-3)!}{2^{(n-2)}\cdot(n-2)!}=(2n-3)!!$，$n\geq 3$ 有意義。有根樹的樹根表示演化上相關序列的共同源頭，並且與其它分類都有聯繫，樹枝則判斷可能的演化距離及時間。

B. 無根樹個數：$N_U = \dfrac{(2n-3)!}{2^{(n-3)} \cdot (n-3)!} = (2n-5)!!$，$n \geq 4$ 有意義。無根樹的樹根表示無法分析演化源頭，但樹枝可判斷演化距離及時間。

最大概似估計建構生物序列演化計算

利用最大概似法建構演化樹，係由待分析的生物序列本身的所有字元序列來探討所有可能組合出代表序列演化的樹。在有根樹圖中的每一個樹結點可以視為一個父(母)代，並且利用序列所有的可能字元的替換組合方式(點突變)，再進一步選擇可能性最大的一種當做計算結果[21][26]。

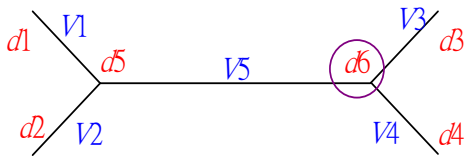


圖 2-1:最大概似法計算演化樹的樹枝長及其結點-無根樹圖

由圖 2-1 假設演化是由樹的一個內部結點 $d6$ 開始，經序列比對之後第 x 條的序列**概似函數** $l_x = f_{d5} \times P_{d5d1}(V1) \times P_{d5d2}(V2) \times P_{d5d6}(V5) \times P_{d6d3}(V3) \times P_{d6d4}(V4)$，$l_x$ 表示概似函數；f_{d5} 表示結點 $d5$ 為某一個字元時的機率值；$P_{didj}(Vx)$ 表示結點 di 字元替換發生到 dj 的機率值、例如:$P_{d1d5}(V1)$ 表示結點 $d1$ 字元替換發生到 $d5$ 的頻率值，Vx 表示兩序列的演化距離。

因為 Jukes-Cantor 模型是單一參數模型[17]，且假設所有序列的核酸鹼基字元是彼此獨立，並且在固定單位時間內的核酸鹼基字元，係由機率值 α (本研究 α 假定為 1/4，即在核酸序列各有 1/4 的平均機率值)演變至其它三種核酸字元的一種。因為在生物序列的點突變過程中，由於序列的核酸字元數目往往很大，因此由二項分佈的伯努利試驗中，若試驗次數樣本很大時且二項分佈的機率很小時，可由參數乘積 $\lambda = np$ 表示之，並使事件出現的次數的機率用卜瓦松(Poission)分佈來逼近。事實上，在生物應用時以 Jukes-Cantor 模型中的核酸字元突變替代率來服從 λt

的卜瓦松機率分布結果，更可滿足 Jukes-Cantor 模型假設單一替代率且有不均等的核酸鹼基字元出現的機率結果、如式 2-1[17][21][26]。

$$P_{didj}(Vx)=\begin{cases}\frac{1}{4}(1+3\exp(-\frac{4}{3}d_E)), di=dj\\ \frac{1}{4}(1-\exp(-\frac{4}{3}d_E)), di\neq dj\end{cases} \text{---式 2-1}$$

由於核酸字元的突變替代率是服從 λt 的卜瓦松機率分布，且任一核酸演變為其它三種的機率為 3α 、即 $\lambda=3\alpha$ 。又在 Jukes-Cantor 模型中，任兩生物序列的進化程序，皆是由其中一個子序列到父序列，再由父序列回到另一子序列，因此可將 λt 的卜瓦松機率的 t 乘於 2。所以每一條序列的核酸字元替代期望值為 $d_E=6\alpha t$，所以式 2-1 可改寫為

$$P_{didj}(Vx)=\begin{cases}\frac{1}{4}(1+3\exp(-8\alpha t)), di=dj\\ \frac{1}{4}(1-\exp(-8\alpha t)), di\neq dj\end{cases} \text{---式 2-2}$$

最大概似法求 Jukes-Cantor 的物種演化樹[21]

設 $d1$ 、$d2$ 是兩隻病毒的序列且 $d1$:AGC、$d2$:ATC，且假設單位時間為 1。由於兩序列只有一個根結點 $d3$、如圖 2-2 所示。

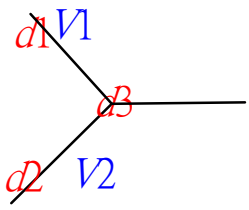


圖 2-2：最大概似法求 Jukes-Cantor 模型的物種演化樹圖

得出演化機率 $P_{didj}(Vx)=\begin{cases}\frac{1}{4}(1+3\exp(-8\alpha t)), di=dj\\ \frac{1}{4}(1-\exp(-8\alpha t)), di\neq dj\end{cases}$，在此 $0\leq\alpha\leq1$ 、$t\geq0$ 故

$$P_{AA}(V)=P_{TT}(V)=P_{CC}(V)=P_{GG}(V)=\frac{1}{4}(1+3\exp(-8\alpha t))$$

$$P_{AT}(V)=P_{AC}(V)=P_{AG}(V)=P_{CT}(V)=P_{CA}(V)=P_{CG}(V)=P_{GA}(V)=P_{GT}(V)=P_{GC}(V)$$

$$=P_{TA}(V)=P_{TC}(V)=P_{TG}(V)=\frac{1}{4}(1-\exp(-8\alpha t))$$

經序列比對結果、如圖 2-4。

$d1$:	A	G	C
$d2$:	A	T	C

圖 2-4:兩序列比對圖

第一行 概似函數計算，概似函數為 $l_{1d3}=f_{d3}\times P_{d3A}(V1)\times P_{d3A}(V2)$

$$l_{1A}=f_A\times P_{AA}(V1)\times P_{AA}(V2)=\frac{1}{3}\times\frac{1}{4}(1+3\exp(-8\alpha t))\times\frac{1}{4}(1+3\exp(-8\alpha t))$$

$$=\frac{1}{48}(1+3\exp(-8\alpha t))^2$$

$$l_{1C}=f_C\times P_{CA}(V1)\times P_{CA}(V2)=\frac{1}{3}\times\frac{1}{4}(1-\exp(-8\alpha t))\times\frac{1}{4}(1-\exp(-8\alpha t))$$

$$=\frac{1}{48}(1-\exp(-8\alpha t))^2$$

$$l_{1G}=f_G\times P_{GA}(V1)\times P_{GA}(V2)=\frac{1}{6}\times\frac{1}{4}(1-\exp(-8\alpha t))\times\frac{1}{4}(1-\exp(-8\alpha t))$$

$$=\frac{1}{96}(1-\exp(-8\alpha t))^2$$

$$l_{1T}=f_T\times P_{TA}(V1)\times P_{TA}(V2)=\frac{1}{6}\times\frac{1}{4}(1-\exp(-8\alpha t))\times\frac{1}{4}(1-\exp(-8\alpha t))$$

$$=\frac{1}{96}(1-\exp(-8\alpha t))^2$$

$$l_{1sum}=l_{1A}+l_{1T}+l_{1C}+l_{1G}=\frac{1}{48}(3+2\exp(-8\alpha t)+11\exp(-16\alpha t))$$

第二行 概似函數計算，概似函數為 $l_{2d3}=f_{d3}\times P_{d3G}(V1)\times P_{d3T}(V2)$

$$l_{2A}=f_A\times P_{AG}(V1)\times P_{AT}(V2)=\frac{1}{3}\times\frac{1}{4}(1-\exp(-8\alpha t))\times\frac{1}{4}(1-\exp(-8\alpha t))$$

$$=\frac{1}{48}(1-\exp(-8\alpha t))^2$$

$$l_{2C} = f_C \times P_{CG}(V1) \times P_{CT}(V2) = \frac{1}{3} \times \frac{1}{4}(1-\exp(-8\alpha t)) \times \frac{1}{4}(1-\exp(-8\alpha t))$$

$$= \frac{1}{48}(1-\exp(-8\alpha t))^2$$

$$l_{2G} = f_G \times P_{GG}(V1) \times P_{GT}(V2) = \frac{1}{6} \times \frac{1}{4}(1+3\exp(-8\alpha t)) \times \frac{1}{4}(1-\exp(-8\alpha t))$$

$$= \frac{1}{96}(1+2\exp(-8\alpha t)-3\exp(-16\alpha t))$$

$$l_{2T} = f_T \times P_{TG}(V1) \times P_{TT}(V2) = \frac{1}{6} \times \frac{1}{4}(1-\exp(-8\alpha t)) \times \frac{1}{4}(1+3\exp(-8\alpha t))$$

$$= \frac{1}{96}(1+2\exp(-8\alpha t)-3\exp(-16\alpha t))$$

$$l_{2sum} = l_{2A} + l_{2T} + l_{2C} + l_{2G} = \frac{1}{48}(3-2\exp(-8\alpha t)-\exp(-16\alpha t))$$

第三行 概似函數計算，概似函數為 $l_{3d3} = f_{d3} \times P_{d3C}(V1) \times P_{d3C}(V2)$

$$l_{3A} = f_A \times P_{AC}(V1) \times P_{AC}(V2) = \frac{1}{3} \times \frac{1}{4}(1-\exp(-8\alpha t)) \times \frac{1}{4}(1-\exp(-8\alpha t))$$

$$= \frac{1}{48}(1-\exp(-8\alpha t))^2$$

$$l_{3C} = f_C \times P_{CC}(V1) \times P_{CC}(V2) = \frac{1}{3} \times \frac{1}{4}(1+3\exp(-8\alpha t)) \times \frac{1}{4}(1+3\exp(-8\alpha t))$$

$$= \frac{1}{48}(1+3\exp(-8\alpha t))^2$$

$$l_{3G} = f_G \times P_{GC}(V1) \times P_{GC}(V2) = \frac{1}{6} \times \frac{1}{4}(1-\exp(-8\alpha t)) \times \frac{1}{4}(1-\exp(-8\alpha t))$$

$$= \frac{1}{96}(1-\exp(-8\alpha t))^2$$

$$l_{3T} = f_T \times P_{TC}(V1) \times P_{TC}(V2) = \frac{1}{6} \times \frac{1}{4}(1-\exp(-8\alpha t)) \times \frac{1}{4}(1-\exp(-8\alpha t))$$

$$= \frac{1}{96}(1-\exp(-8\alpha t))^2$$

$$l_{3sum} = l_{3A} + l_{3T} + l_{3C} + l_{3G} = \frac{1}{48}(3+2\exp(-8\alpha t)+11\exp(-16\alpha t))$$

演化樹的概似度比

$$y = l_{1sum} \times l_{2sum} \times l_{3sum}$$

$$= \frac{1}{48}(3+2\exp(-8\alpha t)+11\exp(-16\alpha t)) \times \frac{1}{48}(3-2\exp(-8\alpha t)-\exp(-16\alpha t)) \times$$

$$\frac{1}{48}(3+2\exp(-8\alpha t)+11\exp(-16\alpha t))$$

$$\Rightarrow \ln y = -3\ln 48 + 2\ln(3+2\exp(-8\alpha t)+11\exp(-16\alpha t)) + \ln(3-2\exp(-8\alpha t)-\exp(-16\alpha t))$$

$$\Rightarrow \frac{\partial \ln y}{\partial t} = \left\{ \frac{2(\text{-}16\alpha\exp(\text{-}8\alpha\text{t})\text{-}176\alpha\exp(\text{-}16\alpha\text{t}))}{3+2\exp(-8\alpha t)+11\exp(-16\alpha t)} + \frac{16\alpha\exp(-8\alpha t)+16\alpha\exp(-16\alpha t)}{3-2\exp(-8\alpha t)-\exp(-16\alpha t)} \right\} \times$$

$$\frac{1}{48}(3+2\exp(-8\alpha t)+11\exp(-16\alpha t)) \times \frac{1}{48}(3-2\exp(-8\alpha t)-\exp(-16\alpha t)) \times$$

$$\frac{1}{48}(3+2\exp(-8\alpha t)+11\exp(-16\alpha t))$$

$$= \frac{\alpha}{6912}\exp(-8\alpha t)[11\exp(-16\alpha t)+2\exp(-8\alpha t)+3] \times$$

$$[33\exp(-24\alpha t)+59\exp(-16\alpha t)-57\exp(-8\alpha t)-3] \ \#$$

同理可求

$$\frac{\partial \ln y}{\partial \alpha} = \frac{t}{6912}\exp(-8\alpha t)[11\exp(-16\alpha t)+2\exp(-8\alpha t)+3] \times$$

$$[33\exp(-24\alpha t)+59\exp(-16\alpha t)-57\exp(-8\alpha t)-3] \ \#$$

令 $\frac{\partial \ln y}{\partial \alpha}$=0、$\frac{\partial \ln y}{\partial t}$=0，再以牛頓逼近法（Newton's Method）求出滿足 $0 \le \alpha \le 1$、$t \ge 0$ 的限止式時的 $\frac{\partial \ln y}{\partial \alpha}$=0、$\frac{\partial \ln y}{\partial t}$=0 最大值。即為最大概似法以 Jukes-Cantor 模型求出之最佳演化樹圖。由於三條以內的生物序列只有無根演化樹[21]，因此若有四條及其以上的生物序列同時計算時，則會有多種情形的無根樹及有根樹發生，因此必須把每種演化樹的最大概似值求出，以決定樹的拓樸結構，並由式 4-3-1 估計出演化樹的分枝長度。以本例而言，如果假設單位時間 t=1、$\alpha = \frac{1}{4}$ 時，因此可

計算各序列字元出現的頻率 f_i 和演化機率 $P_{ij}(Vx)$，$f_A=\frac{2}{6}=\frac{1}{3}$、$f_C=\frac{2}{6}=\frac{1}{3}$、$f_G=\frac{1}{6}$、$f_T=\frac{1}{6}$，另外)令，代入式 2-2，則得出演化機率：

$$P_{didj}(Vx)=\begin{cases}\frac{1}{4}(1+3\exp(-2)),\ di=dj\\ \frac{1}{4}(1-\exp(-2)),\ di\neq dj\end{cases}$$，故

$$P_{AA}(V)=P_{TT}(V)=P_{CC}(V)=P_{GG}(V)=\frac{1}{4}(1+3\exp(-2))=0.351501$$

$$P_{AT}(V)=P_{AC}(V)=P_{AG}(V)=P_{CT}(V)=P_{CA}(V)=P_{CG}(V)=P_{GA}(V)=P_{GT}(V)=P_{GC}(V)$$

$$=P_{TA}(V)=P_{TC}(V)=P_{TG}(V)=\frac{1}{4}(1-\exp(-2))=0.216166$$

由圖 2-2 的序列比對結果可得出下列計算數值。

第一行 概似函數計算，概似函數為 $l_{1d3}=f_{d3}\times P_{d3A}(V1)\times P_{d3A}(V2)$

$$l_{1A}=f_A\times P_{AA}(V1)\times P_{AA}(V2)=\frac{1}{3}\times 0.351501\times 0.351501=0.0411843$$

$$l_{1C}=f_C\times P_{CA}(V1)\times P_{CA}(V2)=\frac{1}{3}\times 0.216166\times 0.216166=0.0155759$$

$$l_{1G}=f_G\times P_{GA}(V1)\times P_{GA}(V2)=\frac{1}{6}\times 0.216166\times 0.216166=0.0077880$$

$$l_{1T}=f_T\times P_{TA}(V1)\times P_{TA}(V2)=\frac{1}{6}\times 0.216166\times 0.216166=0.0077880$$

$$\boxed{l_{1sum}=l_{1A}+l_{1T}+l_{1C}+l_{1G}=0.072337743}$$

第二行 概似函數計算，概似函數為 $l_{2d3}=f_{d3}\times P_{d3G}(V1)\times P_{d3T}(V2)$

$$l_{2A}=f_A\times P_{AG}(V1)\times P_{AT}(V2)=\frac{1}{3}\times 0.216166\times 0.216166=0.0155759$$

$$l_{2C}=f_C\times P_{CG}(V1)\times P_{CT}(V2)=\frac{1}{3}\times 0.216166\times 0.216166=0.0155759$$

$$l_{2G} = f_G \times P_{GG}(V1) \times P_{GT}(V2) = \frac{1}{6} \times 0.351501 \times 0.216166 = 0.0077880$$

$$l_{2T} = f_T \times P_{TG}(V1) \times P_{TT}(V2) = \frac{1}{6} \times 0.216166 \times 0.351501 = 0.01266376$$

$$\boxed{l_{2sum} = l_{2A} + l_{2T} + l_{2C} + l_{2G} = 0.05160356}$$

第三行 概似函數計算，概似函數為 $l_{3d3} = f_{d3} \times P_{d3C}(V1) \times P_{d3C}(V2)$

$$l_{3A} = f_A \times P_{AC}(V1) \times P_{AC}(V2) = \frac{1}{3} \times 0.216166 \times 0.216166 = 0.0155759$$

$$l_{3C} = f_C \times P_{CC}(V1) \times P_{CC}(V2) = \frac{1}{3} \times 0.351501 \times 0.351501 = 0.0411843$$

$$l_{3G} = f_G \times P_{GC}(V1) \times P_{GC}(V2) = \frac{1}{6} \times 0.216166 \times 0.216166 = 0.0077880$$

$$l_{3T} = f_T \times P_{TC}(V1) \times P_{TC}(V2) = \frac{1}{6} \times 0.216166 \times 0.216166 = 0.0077880$$

$$\boxed{l_{3sum} = l_{3A} + l_{3T} + l_{3C} + l_{3G} = 0.072337743}$$

因為演化樹的概似度比率 $y = l_{1sum} \times l_{2sum} \times l_{3sum} = 0.00027003$

所以 $\ln y = -3.569$，由於本研究的示例中，物種的演化樹只有一種所以可直接取圖 2-2 為本例的最佳樹圖結構。

若有二種及其以上的計算結果，則取最大的 y 值所代表的演化樹圖作為計算完成之分類結果。以本例而言 $\ln y = -3.569$，$\Rightarrow y = \exp(-3.659) = 0.028184$，但是無其它值、因此可以直接使用圖 2-2 之演化樹樹圖結果。另外樹枝長度以 $\alpha = p = \frac{1}{4}$ 代入式 4-3-1，可得 $d = -\frac{3}{4}\ln(1 - \frac{4}{3}p) = -\frac{3}{4}\ln\frac{2}{3} \approx -1.365 \Rightarrow \exp(-1.365) \approx 0.3413$ (分枝長度單位)。

若以本例直觀分析來看兩序列都只有一個字元的不同，且序列長度為三字元、即 $\frac{1}{3} \approx 0.3333$ 結果近似。由此估計出最佳樹圖時，我們可以判斷序列的病毒物種相似性，此時再以 Needleman-Wunsch 動態規劃演算法來進行比較最相鄰物種的全序列比對。

由於計算出最佳樹圖時，我們可以判斷序列的物種相似距離，並判斷最接近物種，以本研究流感病毒疫苗計算設計時，我們一但確定要研發的新病毒，只要找出與它最近且已開發的病毒疫苗，再利用 Needleman-Wunsch 動態規劃演算法進行全序列比對搜尋計算出兩序列的共同區及保留區，以判斷出新病毒應由那一段序列進行專一性生化實驗、如圖 2-4 所示。其中紅色 P 表示序列的差異性區段，其餘為一致性區段[22]。

```
Query  1    ATGGAGAGAATAAAAGAATTAAGAGATTTAATGTCGCAGTCCCGCACTCGCGAGATACTA  60
            ||||||||||||||||||||||||||P||||||||||||||P| ||||||||||||||||
Sbjct  1    ATGGAGAGAATAAAAGAATTAAGAGATCTAATGTCGCAGTCTCCCACTCGCGAGATACTA  60
```

圖 2-4：兩序列比對結果示意圖

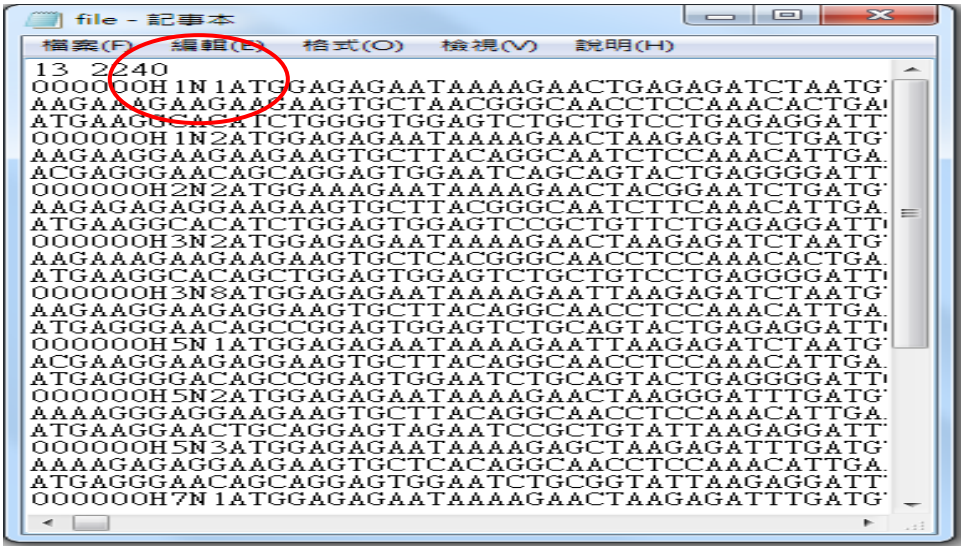
三、　實驗討論(Discussions)

為了比對流感病毒基因序列之間的親疏關係，我們從 NCBI (National Center for Biotechnology Information)網站的資料庫中擷取了 H1N1、H1N2、H2N2、H3N2、H3N8、H5N1、H5N2、H5N3、H7N1、H7N2、H7N3、H7N7、H10N7 共 13 個流感病毒的基因序列，然後利用 PHYLIB3.69 的演化分析軟體 dnaml[6]來計算。

程式操作說明:

在軟體 dnaml 的輸入檔中，因為格式限定，每個染色體序列的長度都要等長，但事實上每個流感的基因序列的長度都不盡相同。以本例而言最短的是 H3N8，共有 2271 個字元。最長的是 H10N7，有 2301 個字元，為了能降低到最小誤差折衷擷取了 2240 個字元並將其整理排列後貼上記事本。(請見下圖 3-1，以 13 組 RNA 序列為例說明)

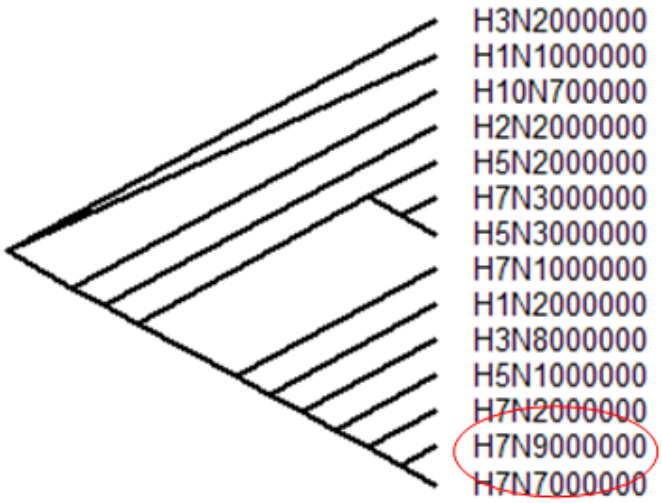
1. 首先將此檔名取名為 file。
2. 再來，將圖 3-1 視窗內左上角的「13」表示有 13 個序列，當中的「2240」表示每個序列共 2240 個字元，而 13 和 2240 中間的空格則是為了符合程式的設計，如圖 3-1 的紅色圈圈內所示。
3. 其中「000000H1N1」、「000000H1N2」等等…是序列名稱。程式的運用是統一將流感病毒基因變數放在各序列的最前端，且預設為十個字元，而不足的部分則予以補零。
4. 再者，各序列的字元均須連續排列，如圖 3-1 所示。



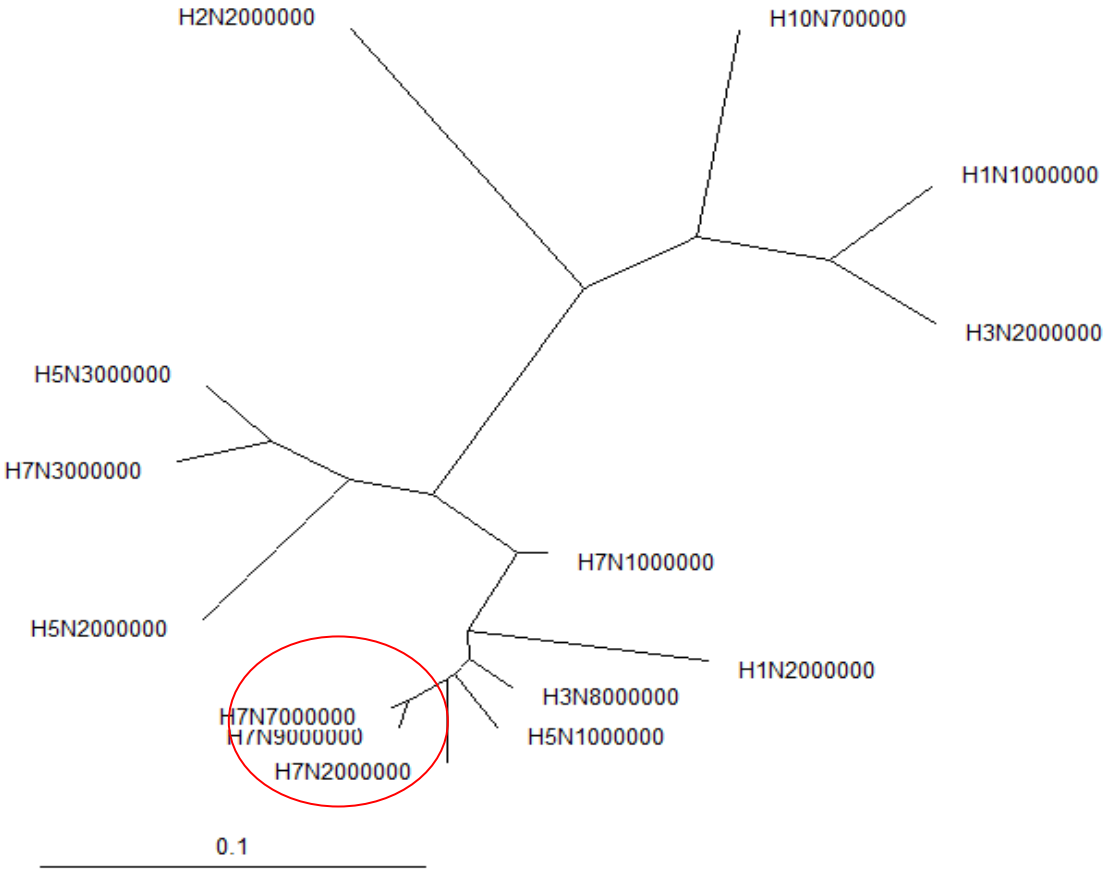
(圖 3-1：在記事本中檔案的處理)

使用演化分析軟體 PHYLIB 3.69 軟體來計算最大概似法 dnaml [7]使用過程如下：開啟 dnaml 的畫面。輸入檔名「file.txt」後按下 enter 鍵，電腦便會從資料夾中，尋找我們選定的檔案。接著我們在 PHYLIB 3.69 軟體中進行 dnaml 各一次，然後我們按下「y」表示同意此程式依照預設程式對我們輸入的資料夾進行計算。

再來，有待此程式執行完工作後。接著我們按下 enter 關閉視窗後，即可從 dnaml 均獲得「outtree」及「outfile」兩個檔案。接著把資料夾中剛修改檔名的 tree 拖曳進另一軟體 treeview，同樣的步驟我們也對檔案 tree ml 執行了一次，即可得到如圖 3-2 及圖 3-3 所示，利用這些圖形我們可推測出種種病毒間彼此之間的演化關連性。



(圖 3-2：tree ml 由最大概似法求出的流感病毒基因序列的演化樹關係-有根樹)



(圖 3-3：tree ml 由最大概似法求出的流感病毒基因序列的演化樹關係-無根樹)

四、 討論(Discussions)

生醫資訊研究若單獨使用 BLAST 方法，可以讓研究者在其中尋找與其感興趣的序列或類似的序列。但是在進行疫苗研發時卻又面臨一個高度不確定的機率問題。

因此本研究提出在嘗試對未知病毒基因進行電泳分析之後，再針對現有病毒及疫苗序列以 Needleman -Wunsch 動態規劃演算法[15]進行序列比對計算，找尋最佳實驗起始疫苗區段，以降低疫苗研發的不確定性。相關流程、如圖 4-1 所示。

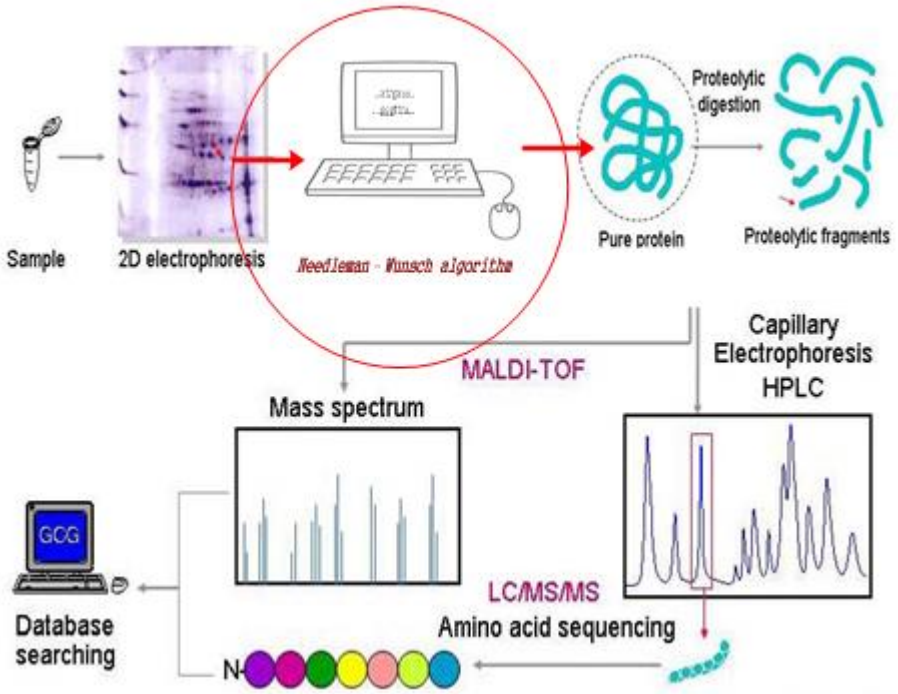


圖 4-1：由研究所提方法來研發疫苗流程圖

因為 Needleman-Wunsch 動態規劃演算法已被證明，可以進行任兩序列的完整比對[11]。因此本研究先計算出已知流感的病毒序列和未研發出的流感病毒序列以 Needleman-Wunsch 動態規劃演算法逐條做完整的序列比對計算。

透過 NCBI 中建置了 Needleman-Wunsch 動態規劃演算法的 Needle 計算程式軟體，我們使用全序列計算比對功能。因為生物學上定義，當蛋白質中超過 25%的

胺基酸序列相同或DNA中超過75%的核酸序列相同，幾乎可以確定蛋白質或DNA序列具有同質性，可作為親緣判定的參考[12]。同時推測出未知流感病毒疫苗可由何種已知流感病毒疫苗來研發出來的可能。

上述推論成立的理由是因為基因組裡常見DNA 突變(稱為差錯型突變) [7]，最常出現是在單核苷酸的替換。科學家稱為「單核苷酸多型性」(single nucleotide polymorphism)。Jukes-Cantor 距離(Jukes-Cantor distance)計算假設是四種核苷酸對核苷酸的替換率是一樣的(突變發生的機率皆相等)。該模型亦是以任兩個序列之間的核苷酸替換的數量進行的最大概似估計(MLE)計算。再以 Jukes 和 Cantor 模型分枝距離公式 $d=-\frac{3}{4}\log_e(1-\frac{4}{3}p)$ 計算相對距離。本研究以 Jukes 和 Cantor 分枝樹距離公式 [2][17]計算了十四種的核苷酸序列相似性矩陣(結果如式 1)。再代入鄰近相接法進行二次分析。

$$\begin{bmatrix} & D1 & D2 & \cdots & D14 \\ D1 & 0 & \cdots & \cdots & \vdots \\ D2 & \vdots & 0 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ D14 & \cdots & \cdots & \cdots & 0 \end{bmatrix} \text{...式1}$$

Needleman-Wunsch 動態規劃演算法

若 A、B 分別為二序列，即

STEP1: $A=(a_1,a_2,...,a_n)$ 、 $B=(b_1,b_2,...,b_m)$ ，

由 A、B 二序列構成的二維表如下：

表 4-1：*A*、*B* 序列的二維表

	a_0	a_1	a_2	$\cdots$	a_n
b_0	$s(0,0)$	$s(1,0)$	$s(2,0)$	$\cdots$	$s(n,0)$
b_1	$s(0,1)$	$s(1,1)$	$s(2,1)$	$\cdots$	$s(n,1)$
b_2	$s(0,2)$	$s(1,2)$	$s(2,2)$	$\cdots$	$s(n,2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
b_m	$s(0,m)$	$s(1,m)$	$s(2,m)$	$\cdots$	$s(n,m)$

STEP2：計算表 4-1 中的每一個元素 $s(i,j)=$

$\max\{s(i-1,j-1)+s(a_i,b_j), s(i-1,j)-d, s(i,j-1)-d\}$ ---(式 2.1)，其中

$s(0,0)=0$、$s(i,0)=-i\cdot d$、$s(0,j)=-j\cdot d$；d :表示序列的連續空位分數，$\{x,y,z\}\in\mathbb{Z}$，

"–" 表示序列比對時的空格，$s(a_i,b_j)=\begin{cases} x & ,a_i=b_j \\ y & ,a_i\neq b_j \\ z & ,\ \text{"–"} \end{cases}$ 。

STEP3：由 STEP2 最後一次算出的 $s(n,m)$ 來回溯計算得出最優比對路徑，並記下相對應的位置若回溯的位置是水平向左、則在縱列插空格並記為 $(a_i,-)$，反之若回溯的位置是水平向上、則在橫列插空格並記為 $(-,b_i)$。

STEP4：統計由 STEP3 的所有回溯記錄計算出兩序列的最佳完整序列比對結果。

五、 結論(Results)

藉由 dnaml 利用最大概似法繪出的圖 3-2 及圖 3-3。我們可看到現有的流感病毒進化的相關性及演化上的相對距離。本實驗分析了來自 NCBI 已公開的十四隻流感病毒的 mRNA 序列。而生醫工程在進行病毒疫苗研發時，多半是直接上 NCBI 進行 BLAST(basic local alignment search tool)的 PCR 引子設計，比對是否已有被研究出來的基因疫苗。雖然利用 BLAST 搜索比對未知及已知的序列，可以由已知序列功能來猜測未知序列功能，且 BLAST 會提供序列比對的實驗建議環境假設、如：建議合成溫度等。

然而 BLAST 是採用 Smith-Waterman 的局部最佳化比對計算[7]，因此對於一條序列到底它的最佳疫苗區段要從那裡開始找起則未提供最好的建議。雖然 Smith-Waterman 的局部比對在進行序列比對檢查時速度快，但卻失去了全盤比對的精確度。透過現代資訊科技發展的快速成長，當科學計算問題與病毒疫苗研發實驗兩者相比較；前者可能只是多花一點時間，但是後者卻可能如同大海撈針般只剩下細微的成功機率。且由於電腦設備的不斷的進步，計算能力的持續提昇。因此未來研究人員亦可以透過以 I-PAD 等各種具有多核心之平板電腦，以及其它類型的智慧型手持裝置來進行相關疫苗的研發所需的序列計算。因此面對計算耗時及未來存有不確定實驗結果的機率，研究人員要考量如何取捨，結果應該不證自明了。

由參考文獻資料[8]，可以發現由已知病毒的現有分子結構可以協助來研發新疫苗[9]，因此我們以本次的實驗結果及文獻資料的引用為基礎做出一項有助於研發疫苗的建議：利用已繪出親緣關係圖上的流感病毒中最相近的兩種或是可以代表其演化的父代病毒，嘗試推論其子代的關聯性。若一者已研發出有效的疫苗則

對另一者(未研發出的疫苗)，理論上應該會存在著緊密的相關性，由此一來，便能增加科學家們在研發疫苗時，成功擷取到有效的基因序列的機率。

本研究由傳統的疫苗研發製造流程加入了最大概似估計及序列比對的 Needleman-Wunsch 動態規劃計算，並使用序列比對軟體來針對流感病毒序列進行序列的全局比對。另外在程式的模擬部分我們利用華盛頓大學 Joseph Felsenstein 教授開發的 PHYLIP-3.69 版進行計算比較分析[7]。

此外計算十四種流感的 mRNA 的相似距離矩陣得到式 1 結果。我們發現進化樹結果如圖 3-2、圖 3-3 所示。研究發現在 H7N3 和 H5N3 兩者在演化的相似度比以 H7N3 和 H7N1 及 H7N2 還近，因此如果我們進行 H7N3 的新藥研發，若能先以 H5N3 的已知疫苗結果來設計可能為 H7N3 的疫苗區段應可加速新疫苗研發成功的機率。並充當實驗分析的初始序列區段，順利找出 H7N3 的疫苗序列的疫苗可能關鍵位置區段實驗，期望協助生醫工程快速找到具有高專一性的目標疫苗序列區段[26]。

本研究提出的實驗猜想方法，在實驗中進行了數值比對計算，並且得到了一些和一般純粹進行生物實驗的生醫學者不同的想法。例如:圖 3-2 及圖 3-3 的 H7N3 和 H5N3 兩類病毒結構由此次分析計算出的結果是最接近的。可是在生醫科學家的直覺卻不那麼想，因為此兩類病毒從分子結構組成的名稱就覺得關係比較遠。本研究的結論未來將進一步地提供生醫科學家在實驗中先以電腦計算。他們可以先檢驗研究進行的推測，再進行生物實驗分析以加速找出具有專一性的疫苗。以提高實驗成功的機率。且相關研究成果也已分別申請國內外發明專利[24]。

本研究也順利完成如何在生醫工程快速找到具有高專一性的目標疫苗序列區段[25]。而由 WHO 資料中推論本研究結果減少現有疫苗製程時間三至六週[24]。因此以本方法所分析之結果亦可提供充分資訊給藥廠或生技公司等專門實驗室進行相關生物技術實驗使用以加速新藥開發。提出如何以數學計算的方法，在流感

病毒序列計算實驗中求出可以簡化疫苗序列專一性找尋的結果[22]。且亦透過文獻比較的方式驗證了所提出的方法的可行[19]。研究內容具體可行[25]，改善了 WHO 自 1930 年代至今的疫苗製程方式[24]。且具體可推論應用於日後的新藥研發流程，減少疫苗研發成本、縮短疫苗開發時程，加速找出可以判斷具有可製造疫苗的最佳生物序列的專一性區段。本研究內容亦可以給醫療專業人員，在面臨病患遭到絕症醫療時提供最後的施藥決策分析，給予患病人員增加一次最後救護的機會。

參考文獻(References)

[1] 莊榮輝、吳裕仁，”***蛋白質體學與單株疫苗應用***”，蛋白質體學，取自台大醫學院生化與分生研究所 http://juang.bst.ntu.edu.tw /Protein/proteomics&mab.htm，2003。

[2] 楊晶、胡剛、王奎、沈世鎰，***生物計算***，科學出版社，2010。

[3] Iyer LM, Balaji S, Koonin EV, Aravind L. "***Evolutionary genomics of nucleo-cytoplasmic large DNA viruses.***" Virus Res., 117 (1): 156–84, 2006.

[4] Villarreal, Luis P. "***Viruses and the Evolution of Life***", ASM Press, 2005.

[5] Liu Y, Nickle DC, Shriner D, et al., "***Molecular clock-like evolution of human immunodeficiency virus type 1***" Virology. 10;329(1):101–8, 2004.

[6] Shors, Teri. **"*Understanding Viruses.*"** Jones and Bartlett Publishers., 2008.

[7] Collier, Leslie; Balows, Albert; ***"Sussman, Max, Topley and Wilson's Microbiology and Microbial Infections ninth edition."***, Volume 1, Virology, volume editors: Mahy, Brian and Collier, Leslie. Arnold., 1998.

[8] Norrby E., ***"Nobel Prizes and the emerging virus concept***" Arch. Virol., 153 (6): 1109–23, 2008.

[9] Dimmock, N.J; Easton, Andrew J; Leppard, Keith, "***Introduction to Modern Virology sixth edition,***" Blackwell Publishing, 2007.

[10] M. H. V. Van Regenmortel, C. M. Fauquet, D. H. L. Bishop, "***Classification and Nomenclature of Viruses : Seventh Report of the International Committee on Taxonomy of Viruses***." Academic Press, 2000.

[11] Carlson GA, Hsiao K, Oesch B, Westaway D, Prusiner SB., "***Genetics of prion infections.***", 7. 1991.

[12] M. Blalock., "***A beginner's guide to microarrays***." Kluwer Academic Publishers, Norwell, Massachusetts USA, 2003.

[13] Wu YJ, Chen HM*, Wu DJ*, Wu JS*, Chu RM, Juang RH, "***Preparation of monoclonal antibody bank against whole water-soluble proteins from rapid-growing bamboo shoots.***" Proteomics 6(22): 5898~5902, 2006.

[14] Leppard, Keith; Nigel Dimmock; Easton, Andrew., "***Introduction to Modern Virology***." Blackwell Publishing Limited, 2007.

[15] Needleman, Saul B., Wunsch, Christian D., "***A general method applicable to the search for similarities in the amino acid sequence of two proteins.***" Journal of Molecular Biology 48 (3): 443–53.,1970.

[16] P. W. Lisette, P. David, "***Noninvasive Genetic Sampling Tools for Wildlife Biologists: A Review of Applications and Recommendations For Accurate Aata Collection***", Journal of Wildl. Manage., 1419-1433, vol 69, 2005.

[17] Jukes T.H. & Cantor C.R., "***Evolution of protein molecules.***" In Munro HN, editor, Mammalian Protein Metabolism, pp. 21-132, Academic Press, New York, 1969.

[18] Saitou N, Nei M. "***The neighbor-joining method: a new method for reconstructing phylogenetic trees.***" Molecular Biology and Evolution, volume 4, issue 4, pp. 406-425, 1987.

[19] T. Y. L. Tommy, W. Jia, S. Yongyi, et al., ***The genesis and source of the H7N9 influenza viruses causing human infections in China***, Nature doi:10.1038/ nature12515, 2013.

[20] Qianyi Zhang et al, "***H7N9 Influenza Viruses Are Transmissible in Ferrets by Respiratory Droplet.***" Science 341, 410. 2013.

[21] 許忠能，***生物信息學***，北京清華大學出版社，2010。

[22] 胡裕仁等，***以計算理論探討病毒疫苗設計之可行性評估***，第 7 屆資訊科技國際研討會，朝陽科技大學，2013。

[23] ***專家揭示 H7N9 來源-國家禽流感參考實驗室主任***-陳化蘭，新聞引據：中央社 2013/4/23.

[24] 全球預警和應對(GAR)，***大流行性流感疫苗的生產程序和時間表***，世界衛生組織 2009 年，H1N1 流感大流行情況說明 7，http://www.who.int/csr/disease/swineflu/ notes/h1n1_vaccine_20090806/zh/ ，2009。

[25] 中興大學智財中心，***運用生物序列程式計算以加快病毒疫苗開發過程之方法***，中華民國發明專利申請中，申請案號: 102134134，2013。

[26] 胡裕仁，系統生物學計算於疫苗藥物設計之研究探討，中興大學應用數學系博士論文，未出版。

附件:14 隻流感病毒的鹼基序列 (下載自 NCBI 資料庫)

1. H1N1:(A/Sydney/DD3-50/2010) polymerase PB2 (PB2) gene, complete cds
2. GenBank: CY092709.1
3. http://www.ncbi.nlm.nih.gov/nuccore/CY092709.1
4. H1N2:(A/duck/Zhejiang/0224-6/2011) segment 1 polymerase PB2 (PB2) gene, complete cds
5. GenBank: JN605400.1
6. http://www.ncbi.nlm.nih.gov/nuccore/345539585
7. H2N2:(A/El Salvador/2/1957) segment 1 polymerase PB2 (PB2) gene, complete cds
8. GenBank: HM204774.1
9. http://www.ncbi.nlm.nih.gov/nuccore/HM204774.1
10. H3N2:(A/Pennsylvania/09/2011) segment 1 polymerase PB2 (PB2) gene, complete cds
11. GenBank: JN655532.1
12. http://www.ncbi.nlm.nih.gov/nuccore/JN655532.1
13. H3N8:(A/environment/Korea/SH11-10/2009) segment 1 polymerase PB2 (PB2) gene, partial cds
14. GenBank: JN087357.1
15. http://www.ncbi.nlm.nih.gov/nuccore/336287088

16. H5N1:(A/domestic green-winged teal/Hunan/67/2005(H5N1)) segment 1 polymerase PB2 (PB2) gene, complete cds
17. GenBank: EU430499.1
18. http://www.ncbi.nlm.nih.gov/nuccore/EU430499.1
19. H5N2:(A/chicken/Belgium/150VB/1999(H5N2)) segment 1 polymerase PB2 (PB2) gene, partial cds
20. GenBank: FJ750564.1
21. http://www.ncbi.nlm.nih.gov/nuccore/FJ750564.1
22. H5N3 (A/Muscovy duck/France/070090b/2007) segment 1 sequence
23. GenBank: CY046163.1
24. http://www.ncbi.nlm.nih.gov/nuccore/CY046163.1
25. H7N1:(A/Baer's pochard/HuNan/414/2010) segment 1 polymerase PB2 (PB2) gene, complete cds
26. GenBank: JQ973647.1
27. http://www.ncbi.nlm.nih.gov/nuccore/JQ973647.1
28. H7N2:(A/duck/Korea/A349/2009) segment 1 polymerase PB2 (PB2) gene, complete cds
29. GenBank: JN244108.1
30. http://www.ncbi.nlm.nih.gov/nuccore/JN244108.1
31. H7N3:(A/wild bird/Korea/A4/2011) segment 1 polymerase PB2 (PB2) gene,

partial cds

32. GenBank: JN244106.1

33. http://www.ncbi.nlm.nih.gov/nuccore/341610036

34. H7N7 (A/wild bird/Korea/A72/2010) segment 1 polymerase PB2 (PB2) gene, complete cds

35. GenBank: JN244116.1

36. http://www.ncbi.nlm.nih.gov/nuccore/341610056

37. H7N9：(A/wild bird/Korea/A9/2011) segment 1 polymerase PB2 (PB2) gene, partial cds

38. GenBank: JN244118.1

39. http://www.ncbi.nlm.nih.gov/nuccore/341610060

40. H10N7：(A/blue-winged teal/Texas/Sg-00086/2007) segment 1, complete sequence

41. GenBank: CY077977.1

42. http://www.ncbi.nlm.nih.gov/nuccore/315111617

致謝

本研究承蒙中興大學生物醫學研究所周寬基教授及中國醫藥大學學士後中醫學系王舜德博士提供技術及實驗協助指導。

評語

結合新興疾病的議題，利用資訊科技來降低病毒變異搜尋的空間，以快速找出有效疫苗的開發時機，整體作法有很好的創新性，具有申請專利的優勢與必要性，類似的方法，也可以應用在新藥的開發上，以了解藥物有效性與基因的關係。