國立交通大學

分子醫學與生物工程研究所

碩士論文

A 型 H1N1 流行性感冒病毒紅血球凝集素之基因與抗原分析並與 H3N2 病毒之比較

Genetic and antigenic analysis to the hemagglutinin of influenza A H1N1 virus and comparisons with H3N2 virus

研 究 生：林韋帆

指導教授：楊進木　教授

中 華 民 國 一百 年 四 月

A 型 H1N1 流行性感冒病毒紅血球凝集素之基因與抗原分析

並與 H3N2 病毒之比較

Genetic and antigenic analysis to the hemagglutinin of influenza A H1N1 virus and comparisons with H3N2 virus

研 究 生：林韋帆　　Student：Wei-Fan Lin

指導教授：楊進木　　Advisor：Jinn-Moon Yang

國 立 交 通 大 學

分 子 醫 學 與 生 物 工 程 研 究 所

碩 士 論 文

A Thesis Submitted to Institute of Bioinformatics and Systems Biology

College of Biological Science and Technology

National Chiao Tung University in partial Fulfillment of the Requirements

for the Degree of Master in

Molecular Medicine and Bioengineering

April 2011

Hsinchu, Taiwan, Republic of China

中華民國一百年四月

A 型 H1N1 流行性感冒病毒紅血球凝集素之基因與抗原分析並與 H3N2 病毒之比較

學生：林韋帆　　指導教授：楊進木

國立交通大學 分子醫學與生物工程研究所

摘　　要

A 型流行性感冒是一種非常重要的人類感染疾病，它對全世界的公共衛生造成極大的威脅。H1N1 為 A 型流感中的一種亞型，在二十世紀曾發生過全球性大流行並且造成約兩千萬人的死亡。最近源自於豬的新型 H1N1 病毒開始感染人類並造成 2009 年的全世界性大流行。紅血球凝集素(Hemagglutinin，HA)為流感病毒表面的抗原性醣蛋白，在流感病毒感染或接種疫苗期間紅血球凝集素會被抗體所中和。紅血球凝集素上之突變累積會造成抗原性漂移(Antigenic drift)的發生，此時疫苗往往需要重新設計來對於下一波疫情提供足夠的保護力。然而直到 2009 年為止大部分對於紅血球凝集素的研究主要針對 H3N2 亞型，目前對於 H1N1 亞型紅血球凝集素的了解仍然不足。針對 H1N1 亞型的紅血球凝集素進行研究對於公共衛生與疫苗發展是一個重要且有高度急迫性的議題。

在本論文中，我們針對 H1N1 亞型之紅血球凝集素進行基因與抗原分析，並且將結果與 H3N2 亞型做比較。在基因層次，我們收集了一千五百二十五株 H1N1 之紅血球凝集素序列，並且利用亂度值(Shannon entropy)分析每個胺基酸位置之改變程度。在抗原層次，我們從近 40 年的流病週報(Weekly Epidemiological Record) 以及相關文獻收集了二百筆 HIN1 之血球凝集抑制試驗值(hemagglutination inhibition，HI)，並且利用統計方法量度每個胺基酸位置之抗原性變化大小。最後並且利用決策樹方法(Decision tree)(C4.5)來預測 H1N1 病毒之抗原性漂變株(antigenic variants)。

透過在 H1N1 上進行基因與抗原分析，我們在紅血球凝集素上挑選了三十個具有重要性的胺基酸位置，觀察這些位置中，有二十六個位於表面，此外有九個位置位於抗原決定位(Epitope)上。根據基因與抗原分析的結果，本論文發現 H1N1 亞型之重要區域主要分為二塊，包括鄰近受體嵌合區(Receptor binding site)以及遠離受體嵌合區之抗原決定位(epitope)。相較之下，H3N2 亞型之重要區域大部分皆分佈於鄰近受體嵌合區。同時透過比較 H1N1 與 H3N2 亞型，發現了許多在 H3N2 亞型抗原決定位上之胺基酸位置也很有可能是 H1N1 亞型的抗原決定位。除此之外，決策樹方法建立之模型可達到 85%的預測率。由以上的結果，顯示我們的方法具有穩健之特性並且有助於了解 H1N1 病毒之基因與抗原性演化，並且對於設計疫苗有極大幫助。

Genetic and antigenic analysis to the hemagglutinin of influenza A H1N1 virus and comparisons with H3N2 virus

Student：Wei-Fan Lin Advisor：Dr. Jinn-Moon Yang

Institute of Molecular Medicine and Bioengineering

National Chiao Tung University

Abstract

Influenza A virus causes significant morbidity and mortality in humans. H1N1 is one of the current circulating influenza A subtypes in human. The H1N1 pandemic occurred in the early 20th century and resulted in approximately 20 million deaths in the world. Recently, the emerged swine-origin H1N1 virus has infected human population and cause the 2009 influenza pandemic. Hemagglutinin (HA), which is an antigenic glycoprotein on the surface of influenza virus, is neutralized by antibodies during infection or vaccination. Accumulation of mutations on HA can lead to antigenic drift. The emergence and spread of antigenic variants often requires a new vaccine strain to be selected before coming epidemic. Most of studies on HA focused on the H3N2 subtype. However, the genetic evolution and antigenic evolution of the HA is poor understood for subtype A (H1N1). To study the genetic and antigenic evolution of subtype A (H1N1) is an emergent issue for public health and vaccine development.

In this thesis, we performed the genetic and antigenic analysis to the HA of A (H1N1) viruses. In the sequence level, we collected 1525 HA sequences and used Shannon entropy to quantify the genetic diversity of each amino acid. In the vaccine efficacy level, we collected 202 pairs of HI assays from weekly epidemiological record (WER) and publications in last 40 years. Based on the collected Hemagglutination Inhibition (HI) assays, we applied a statistical index to quantify the antigenic score of each amino acid on HA. Finally, a decision tree tool (C4.5) was used to build a model for predicting the antigenic variants of H1N1 virus.

We select 30 critical positions of H1N1 hemagglutinin by the genetic and antigenic analysis. There are 26 positions on the surface of the HA and 9 positions on the H1N1 epitopes. Based on the genetic and antigenic analysis on HA, we found that there are two sites with both high genetic diversity and antigenic score in A (H1N1) virus. These two sites include one site around the receptor binding site and the other antigenic site about 45 Å distant from receptor binding site. In contrast, there is only one site, which is around the receptor binding site, have high genetic diversity and high antigenic in A (H3N2) virus. By comparing the HA of two subtypes of influenza A virus, we found that some amino acid positions locating on the antigenic sites of influenza A (H3N2) virus are potential epitope residues for influenza A (H1N1) virus. In addition, the accuracy of our model for predicting antigenic variants was 85% by using HA sequences as input. We believe that our methods are useful for the vaccine development and understanding the genetic and antigenic evolution of influenza A (H1N1) virus.

誌謝

在交大研習碩士學位的期間中，對於曾經在研究及做事態度上給予熱心協助和教誨的老師與學長姊、同學們、學弟妹，使我得以在這個學習階段能順利並成長，對於這些良師益友韋帆十分感激與感謝。

韋帆非常感謝指導教授楊進木老師，在交大學習期間特別用心指導與栽培，更不厭其煩在研究態度與研究技巧上給予教誨，尤其在於態度上，老師帶給學生謹慎力求卓越的態度，更是樹立良好的風範。在研究上，老師嚴謹的求證過程與創新的思考，不斷追求卓越與進步，更讓學生了解人生是要不斷努力學習，並且培養團隊合作的精神。感謝老師提供優質的研究環境與助學金，使學生在交大學習階段能順利無擔憂。最後感謝老師您在學習研究及人生經驗上的提醒，得以讓學生更穩健成長，謝謝老師。

其次學生要感謝博士後研究員黃章維學長，在研究上對於學生給予非常多的幫忙與協助，尤其在實驗方法與分析上的製作細節更是不吝教導，給予許多研究上的建議，並在論文上幫助修訂與校正，對於學長不辭辛勞的幫助，由衷感謝學長所做的一切，韋帆銘記於心。

接著感謝彭慧玲老師與楊昀良老師能參與韋帆的碩士學位口試，給予韋帆指導和建議。同時感謝實驗室的學長姐與各位同學，在研究上給予韋帆討論交流及協助，感謝宇書、怡馨、志達、凱程、伸融、峻宇、一原等學長姐提供研究上的分析技巧和操作給予幫助。超哥、敬立、Gay 哲、怡瑋在實驗室生活中給予幫忙。感謝實驗室的學弟妹幫忙。另外感謝我的好朋友羅志傑、紀志弘、JuJu 在我背後一直給予支持與幫忙，分享許多人生經驗與心得。

最後感謝我的家人，在我背後始終給予最大支持和關心，謝謝老爸和老媽細心的照顧，也感謝我親愛的哥哥和嫂嫂的噓寒問暖讓我能努力向前。

總目錄

圖目錄

表目錄

壹、緒論

一、研究背景

流行性感冒是重要且廣泛之人類感染性疾病，每年流感地區流行造成全世界約五十萬人死亡[1]。此外在過去一百年間發生四次全球性高度死亡率流感的世界大流行。在這四次流感世界大流行中，1918 年 H1N1 流感大流行造成兩千萬人死亡[2]。最近，緣起於豬流感病毒的 2009 年 H1N1 流感大流行為全世界公眾健康造成重大威脅[3]。

A 型流感病毒是一種單股負鏈RNA 病毒，它會感染人類和其他動物包含豬、雪貂、和許多鳥類物種。流感病毒有三種類型(A、B 和 C)流行於人類，A 型流感具有高度的基因變異性並且對人類的致病性最高[4]，常會導致嚴重的疾病。在A 型流感中有八個基因組片段編碼十一個蛋白質[5]，在這十一個蛋白質之中，兩個表面蛋白質紅血球凝集素和神經胺酸酶是人體免疫系統辨識主要的目標。另外 A 型流感病毒可藉由紅血球凝集素主要的差異分成不同亞型。目前有十六種紅血球凝集素和九種神經胺酸酶已經被定義[6]，其中一些不同組合之紅血球凝集素和神經胺酸酶流感病毒帶原於野生水鳥上[7]。

二、 H1N1 和 H3N2 介紹

H1N1 為 A 型流感病毒中的一種亞型，發生過幾次著名的大流行，曾經在 1918 年西班牙流感中造成約兩千萬至五千萬人死亡，可能為人類歷史上最具有死亡性的流感大流行之一，另外在 2009 年爆發新的 H1N1 流感大流行，世界衛生組織統計至 2010 年初，這次的大流行所造成全球的死亡人數為一萬七千人。然而，在 1957 年時，H1N1 流感病毒消失而被新的重組流感病毒 H2N2 所取代，直到 1977 年時，帶有與 1950 年代基因與抗原相似的 H1N1 流感病毒才又再次出現[8]，並且與 A 型流感中亞型 H3N2 和 B 型流感成為目前季節性流感的主要病毒之一。

H3N2 是 A 型流感病毒中的另一個亞型，為人類流感病毒研究中非常重要的病毒，曾經在 1968 年在香港爆發大流行造成，而後此流感疫情傳到美國，這次大流行從 1968 年至 1969 年總計造成一百萬人死亡。另外，在每年季節性 H3N2 的流感病毒造成美國約三萬六千人死亡。

三、研究動機與目的

H1N1 為 A 型流感中的一種亞型，在二十世紀曾發生過全球性大流行並且造成約四千萬人的死亡。最近源自於豬的新型 H1N1 病毒開始感染人類並造成 2009 年的全世界性大流行，因此研究 H1N1 病毒亞型為現今十分重要的流感議題。因此我們針對 H1N1 進行研究，並將結果與 H3N2 做比較，透過基因層次及抗原層次的分析方式，想要找出 H1N1 亞型在抗原性漂移上的機制與規則，首先在紅血球凝集素基因層次，我們收集了一千多株 H1N1 之紅血球凝集素序列，並且利用亂度值(Shannon entropy)分析每個胺基酸位置之改變程度。在紅血球凝集素抗原層次，我們從近 40 年的流病週報(Weekly Epidemiological Record)以及相關文獻收集了二百多筆 HIN1 之血球凝集抑制試驗值(hemagglutination inhibition，HI)，並且利用統計方法量度每個胺基酸位置之抗原性變化大小，最後透過決策樹預測 H1N1 之抗原變異株。觀察 H1N1 流感病毒在紅血球凝集素上對於抗原性漂移扮演關鍵的胺基酸位置、抗原變異株之規則同時透過基因序列資訊預測抗原變異株、重要的基因及抗原上之行為，並與 H3N2 比較這些分析結果。

四、流感病毒基因與抗原的演化

流感病毒之紅血球凝集素具有高度的基因變異性，來自於易出錯的 RNA 聚合酶，高複製機率和基因片段的重組[9]。突變(取代、缺失和插入)是一種產生流感病毒基因變異最重要的機制。

五、抗原性漂移(Antigenic drift)

儘管紅血球凝集素(hemagglutinin, HA)和神經胺酸酶(neuraminidase, NA)兩者是抗體目標，然而紅血球凝集素包含最高比例的抗原作用位可被人類免疫系統辨識[10-12]。在流感基因組上頻繁的累積突變會導致紅血球凝集素結構發生改變。由於人類免疫系統不能完全交叉保護對抗病毒感染[13]，紅血球凝集素上新的突變可能導致抗體不再辨識變異後的病毒而讓病毒逃離免疫系統的辨識。這個隨著時間抗原結構上逐漸改變稱之為抗原性漂移[14]。另外全球流感監測網絡透過紅血球凝集素抑制測試定期檢測新興的抗原變異株[15, 16]。疫苗效價測試是一種接合實驗，代表一種病毒(例如流行病毒)株和另一個動物抗血清(疫苗)株之間的接合能力，此外紅血球凝集素是目前流感疫苗的主要成分[16]。

六、抗原性轉換(Antigenic shift)

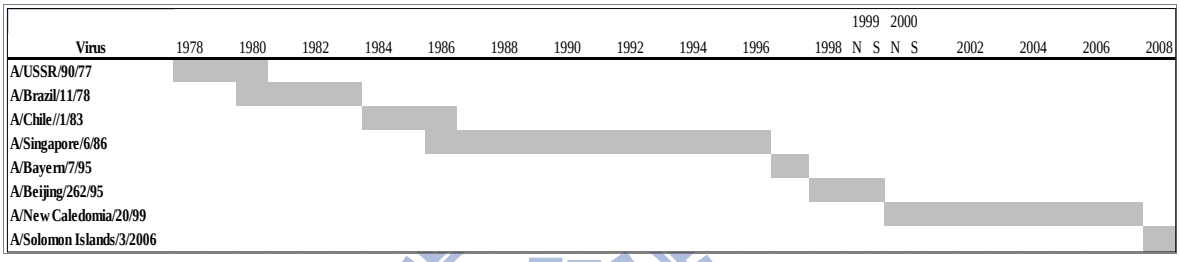
當一個宿主被兩種或更多不同亞型流感病毒共同感染時，在兩者流感病毒之間片段的基因組可能會重新組合產生一種新的流感亞型具有新穎的紅血球凝集素和神經胺酸胺的混合。這個結果稱之為抗原性轉換[17, 18]，新的流感亞型通常會對人類造成嚴重傷害因為人類對新病毒未具抗體。在過去一百年中，有四次流感大流行，分別源自於在紅血球凝集素和神經胺酸胺的重組：1918 年(H1N1 亞型)[19]，1957 年(H2N1 亞型)[20]，1968 年(H3N2 亞型)[21]和最近 2009 年(H1N1 亞型)[7]源自於豬流感病毒基因組重組的流感大流行[3, 22]。此外，新的流感病毒通常會成為接續幾年流行的流感病毒祖先[23]。

七、每年疫苗審查和疫苗株選擇

目前疫苗接種是預防流感最主要的防治措施[24, 25]，當疫苗株和流行株有高度相似抗原特性的紅血球凝集素時，疫苗可提供有效的保護力[26]。人類免疫系統對抗單一感染侵入的流感病毒株可提供終生免疫[27]；然而，不同的流感病毒經歷抗原性漂移後可能在接下來幾年感染人類。為了確保疫苗效力能有效對抗

流行病毒株，世界衛生組織建立了一個全球監控網路來監測新型流感病毒的出現[28]。每個流感季節，一群專家小組舉行會議從最近流行的病毒株內共同挑選一株合適的病毒株，在隔年冬天作為疫苗病毒株(圖一)[15]。這個方法引發了一個問題，今年的病毒株是否可能成為接下來冬天流感疫情的病毒株[29]。

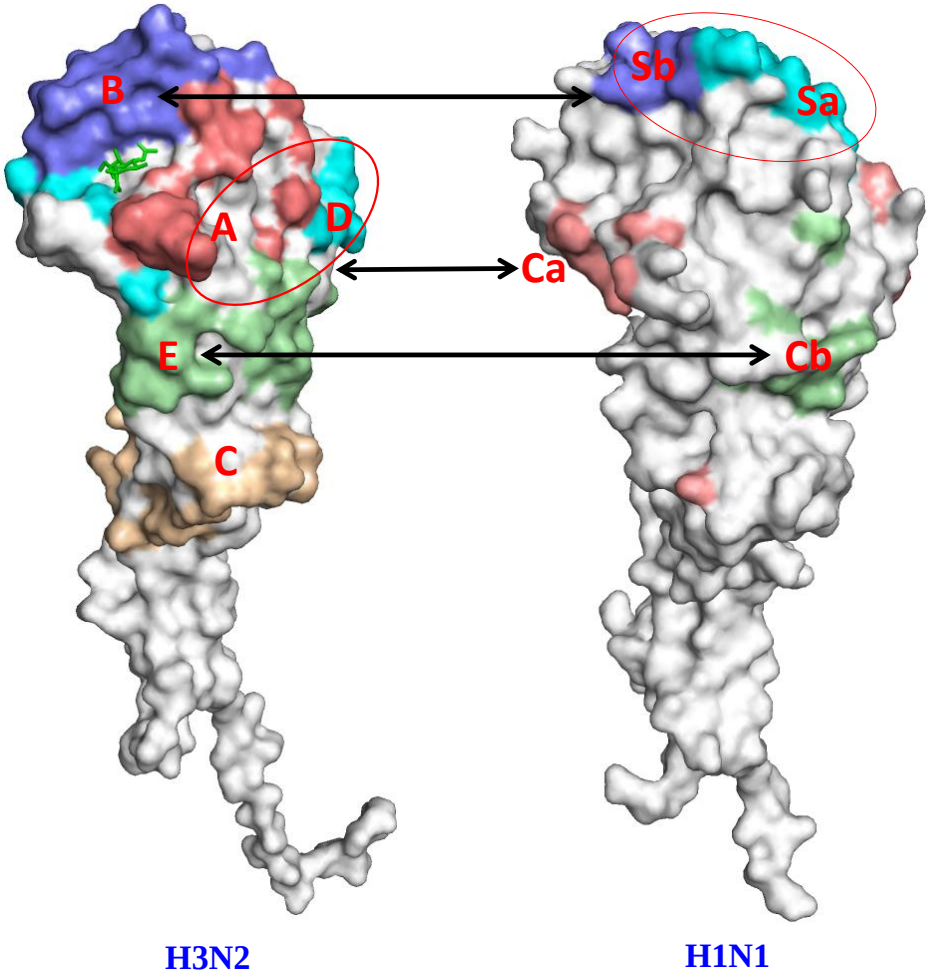
由於生產流感疫苗需要六個月以上的時間[30]，建議的疫苗株在疫苗使用季節前約九至十二個月就開始被製造[25]。因為疫苗的製造時間長，當新型變異株未能及早發現，會造成疫苗株與流行株錯誤配對。一個很好的例子就是發生在2009 年疫苗株與流行株(H1N1 亞型)錯誤配對。



圖一、A 型流感 H1N1 疫苗所包含的病毒(1977-2008)。

八、抗原決定位(Epitope)

觀察 H3N2 及 H1N1 兩者的抗原決定位對應關係，在過去研究中對於 H3N2 和 H1N1 抗原決定位已經從蛋白質結構上所定義，H3N2 病毒的紅血球凝集素可分為 A、B、C、D、E 五個抗原決定位(圖二)，佔紅血球凝集素 1 全部 329 個胺基酸中共 131 個胺基酸，其中抗原決定位 A 和 B 接近宿主之受體接合的位置，並且和抗體中和反應的效果呈現高度相關。H1N1 病毒的紅血球凝集素可分為 Sa、Sb、Ca、Cb 四個抗原決定位(圖二)，佔紅血球凝集素 1 全部 329 個胺基酸中共 30 個胺基酸，其中抗原決定位 Sa 和 Sb 為接近宿主之受體接合的位置，Ca 是交互作用介面上的組成次單位，可細分為 Ca1 和 Ca2，Cb 則坐落於退化酯酶的功能區塊上。透過蛋白質結構疊合及過去研究提供相關對應資訊，於下(圖二)中看到 H3N2 和 H1N1 抗原決定位對應情況，可以觀察到在 H3 中抗原決定位 B 對應到 H1 上抗原決定位 Sa 和 Sb，H3 之抗原決定位 A 和 D 對應 H1 抗原決定位 Ca，H3 之抗原決定位 E 對應 H1 抗原決定位 Cb。



圖二、H3N2 對應 H1N1 紅血球凝集素抗原決定位結構對應圖。H3N2(PDB: 4HMG)抗原決定位依序為 A 至 E。H1N1(PDB: 1RU7)抗原決定位為 Sa、Sb、Ca、Cb。

九、過去研究

對於流感病毒最急迫的議題就是如何選擇合適的疫苗株。一個適合的疫苗株可以有效提供疫苗效力抵抗流行病毒株[31]。根據這個議題很多方法被提出來研究紅血球凝集素的演化和疫苗的發展[16, 26, 32, 33]。我們將這個議題根據資料分成不同構面加以分析。

(一)、演化式方法(Phylogenic methods)

許多過去的研究著重於流感的基因演化因為從公共資料庫有大量的序列資訊可獲得。有研究提出第一個方法藉由紅血球凝集素序列去預測流感病毒的演化[32]。他們收集了從 1983 年到 1997 年間三百五十七種紅血球凝集素序列，並且建構一個系統發生樹。藉由系統發生樹，它們透過正選擇定義了十八個密碼子[34]。根據回朔測試，他們的研究顯示在系統發生樹上，在最近十一個流感季節中有九個流感季節，病毒譜系在正選擇密碼子上有大量的突變為未來紅血球凝集素的流感祖先譜系[33]。他們的研究證實了解紅血球凝集素的基因演化有助於疫苗株的選擇。

(二)、考慮基因資訊為主的分群方法(Clustering methods based on genetic data)

有研究提出一個分群方法去預測未來主導的紅血球凝集素序列，並且探討其潛在相關疫苗株的選擇[35]。基於紅血球凝集素的序列分群，它們的方法挑選目前季節中最具主導性的序列作為未來的疫苗株。此外，他們研究空間和時間的病毒族群分布並且比較它們的方法挑選目前季節中最具主導性的分群和世界衛生組織所推薦的流感疫苗。他們的研究證實了將紅血球凝集素序列結構分群的分析有助於疫苗株的選擇[35]。

(三)、考慮抗原資訊為主的分群方法(Clustering methods based on antigenic data)

全球流感監測網絡定期透過紅血球凝集素抑制測試分析流行病毒株抗原特性[15, 16]。雖然抗原資料是疫苗株選擇的關鍵指標，抗原資料大量未探討導致難以定量解釋。過去的研究提出一個 A 型流感抗原圖像展示抗原的演化如何對應基因的演化[16]。紅血球凝集素抗原演化在自然情況中產生間斷可透過抗原圖像上視覺化看到[16]，它們的方法量化從 1968 年到 2003 年在疫苗株與流行株之

間抗原性距離，因此可幫助疫苗株的選擇。在過去研究中最重要的發現之一，抗原演化相較於基因演化更具有間斷性，基因改變有時候會不成比例地大量對抗原的影響[16]。他們的研究證實抗原資訊與基因資訊兩者對流感病毒的演化提供了有價值的見解。

(四)、同時考慮基因和抗原資訊的混合性方法(Hybrid method considering genetic and antigenic data)

目前，紅血球凝集素抑制測試是表示流行株抗原特性的主要方法。然而在公共資料庫紅血球凝集素抑制測試資訊的量化數據遠小於序列資訊[36-38]。過去的研究提出第一個方法基於紅血球凝集素序列去預測抗原性變異[26]。所使用的數據集包含一百八十一對紅血球凝集素序列，結果顯示基於五個抗原作用位上的模型具有最佳的預測變異準確性。

十、決策樹(Decision tree)

決策樹為一種輔助決策的工具，它使用一種樹狀圖或者決策模型表示可能的結果，包括效用、資源成本、偶發事件結果等。它是一種用來呈現一個演算法的方法。決策樹常被用於運作研究，特別是決策分析，可幫助找到一個最有可能達到目標的策略。在機器學習中，決策樹為一種預測的模型，它代表對應值和對應屬性之間的一種映射關係，樹中各個節點表特定對象，每個分支的路徑表特定對象可能的屬性，而每個葉的結點對應從根的節點至特定的節點的路徑表示特定對象的屬性值，透過數據產生進而產生決策樹的機器學習稱之為決策樹學習，也稱作決策樹。

十一、挑戰

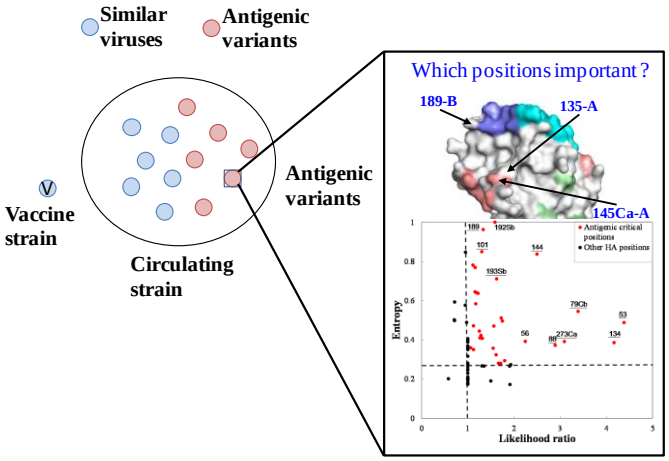
在發展流感疫苗的一個重要議題是改善疫苗株的選擇準確性。在今天所選擇的疫苗株很可能成為接下來一年主導的流感疫情[29]。此外，提前於監測系統未

能在早期偵測變異株，一個更全面的了解基因與抗原之間的演化關係是有助於預測流感病毒的演化。另外，有數以千計收錄於公共資料庫的紅血球凝集素序列缺乏抗原資訊。如果一個方法可以連結基因演化(序列資料)與抗原演化(抗原資料)，將可提供了解抗原性漂移和疫苗發展有價值的見解。

十二、論文總覽

在本篇論文中，我們研究著重於結構空間上基因與抗原演化的關係。論文編排結構如下。在之後我們透過一個方法用來定義抗原變異株上重要的胺基酸位置和規則(圖三)。這個規則描述為當一株(例如：流行株)的紅血球凝集素無法被抗體辨識對抗其他株(例如：疫苗株)。在紅血球凝集素結構上重要位置分布很廣；然而紅血球凝集素被抗體辨識的區域對於抗原作用位上結構改變是高度相關。我們透過以抗原作用位為基礎的方法應用抗原決定位上結構改變來定義 A 型流感的抗原性漂移。我們根據這個章節的兩個議題：第一議題，如何在發生改變的抗原性決定位上量化分析結構改變的程度；第二議題，找出發生改變的抗原決定位和抗原性漂移之間的關係。

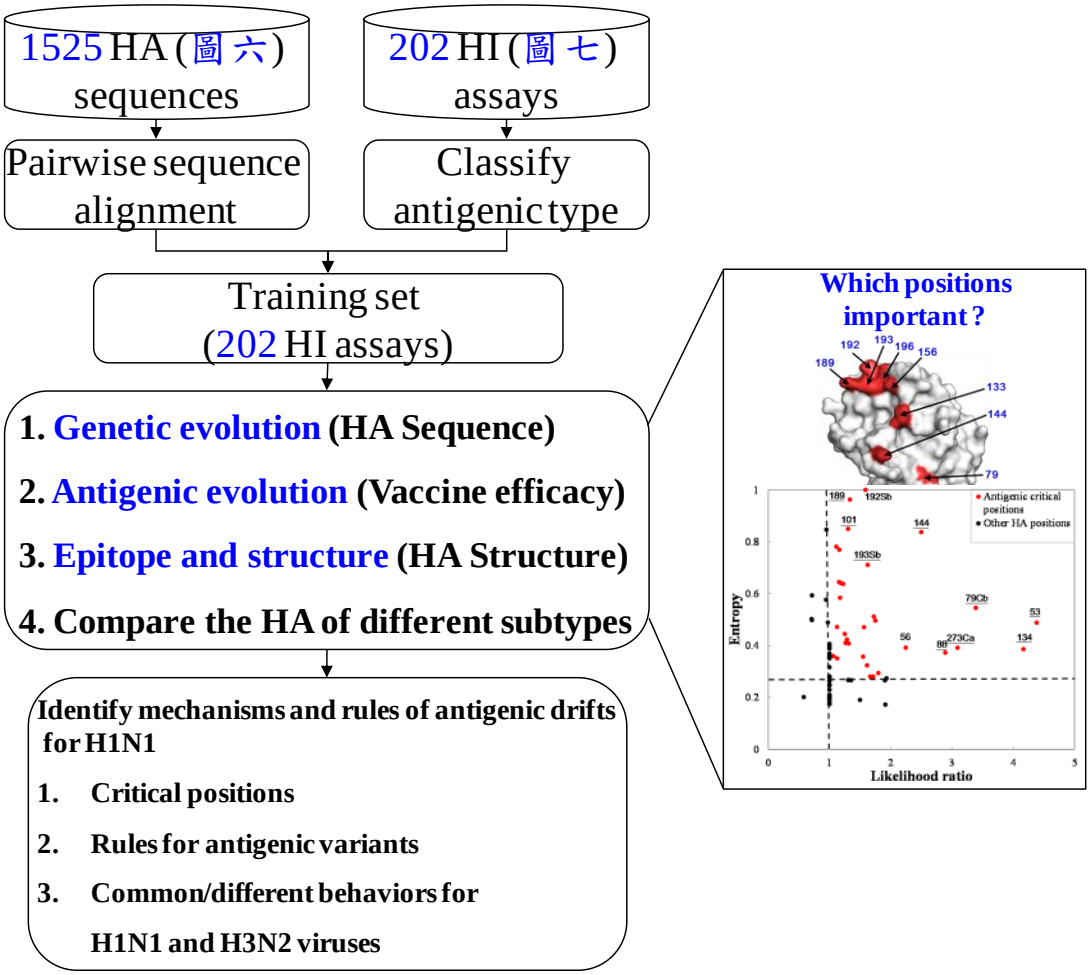
從上述這兩個構面，我們觀察到一些胺基酸的突變能導致抗原變異株而一些其他的突變卻對抗原變異株產生影響極小。此外，我們也注意到在抗原決定位 A 和抗原決定位 B 上的突變似乎可能較容易導致抗原變異株的產生。上述這些觀察引發一個問題是否胺基酸位置的突變在抗原決定位上影響相同與否。



圖三、基因與抗原演化關係示意圖

貳、研究材料與方法

研究的流程圖如下(圖四)。在過去的研究中，在分析多種不同亞型流感病毒之紅血球凝集素胺基酸位置時，透過具有代表性的 H3N2 病毒株之血球凝集素胺基酸位置編號為參考基準依據。首先我們計算從代表性的資料集內計算三百二十九個紅血球凝集素胺基酸位置，評估資訊獲得量代表在基因和抗原演化之間的適用性。



圖四、A 型流感 H1N1 研究方法流程圖。

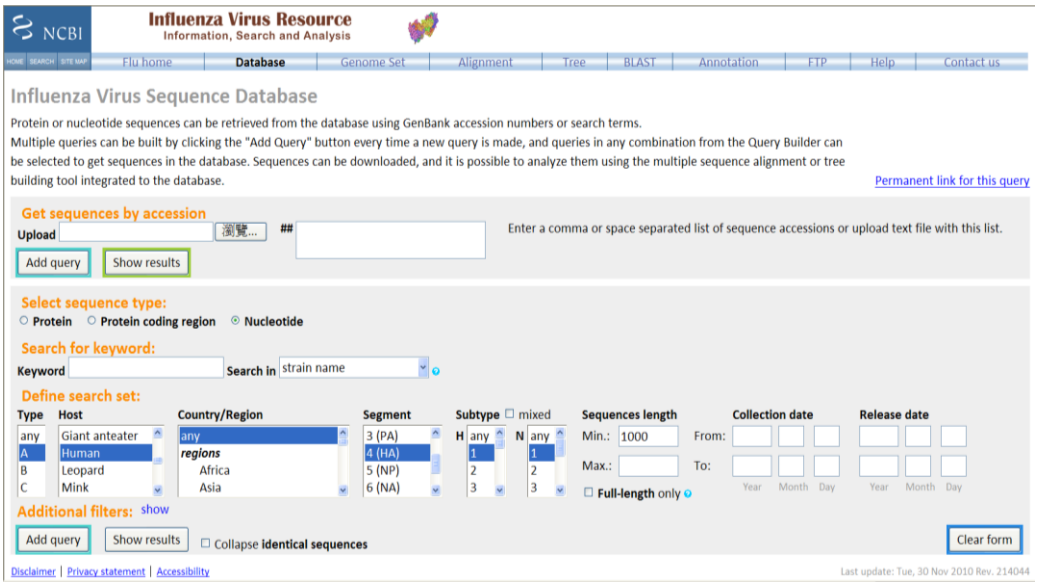
我們首先挑選一個被使用在已發表的研究上的代表性訓練集[26]。然後我們從紅血球凝集素序列和蛋白質結構析取特徵。紅血球凝集素抑制力價是由抗原性漂移透過一系列稀釋倍數而來。抗原性距離較遠的意義是指病毒兩者之間具有較大的抗原差異。當有兩個變數(基因特徵和抗原距離)後，我們可以計算紅血球凝集素上三百二十九個位置各自的資訊獲得量。透過基於資訊獲得量的已知方法(Decision Tree C4.5)我們能選擇一些重要位置上的分群並且得到一個預測抗原變異株的對應模型。在我們找到這些重要位置後，探討基因與抗原演化間的資訊獲得量合適性代表的相關程度。這些被挑選的位置將會被使用於預測疫苗變異株和比較相關研究的預測準確率。

在接下來的部分我們會先展示如何準備資料及方法論的細節。

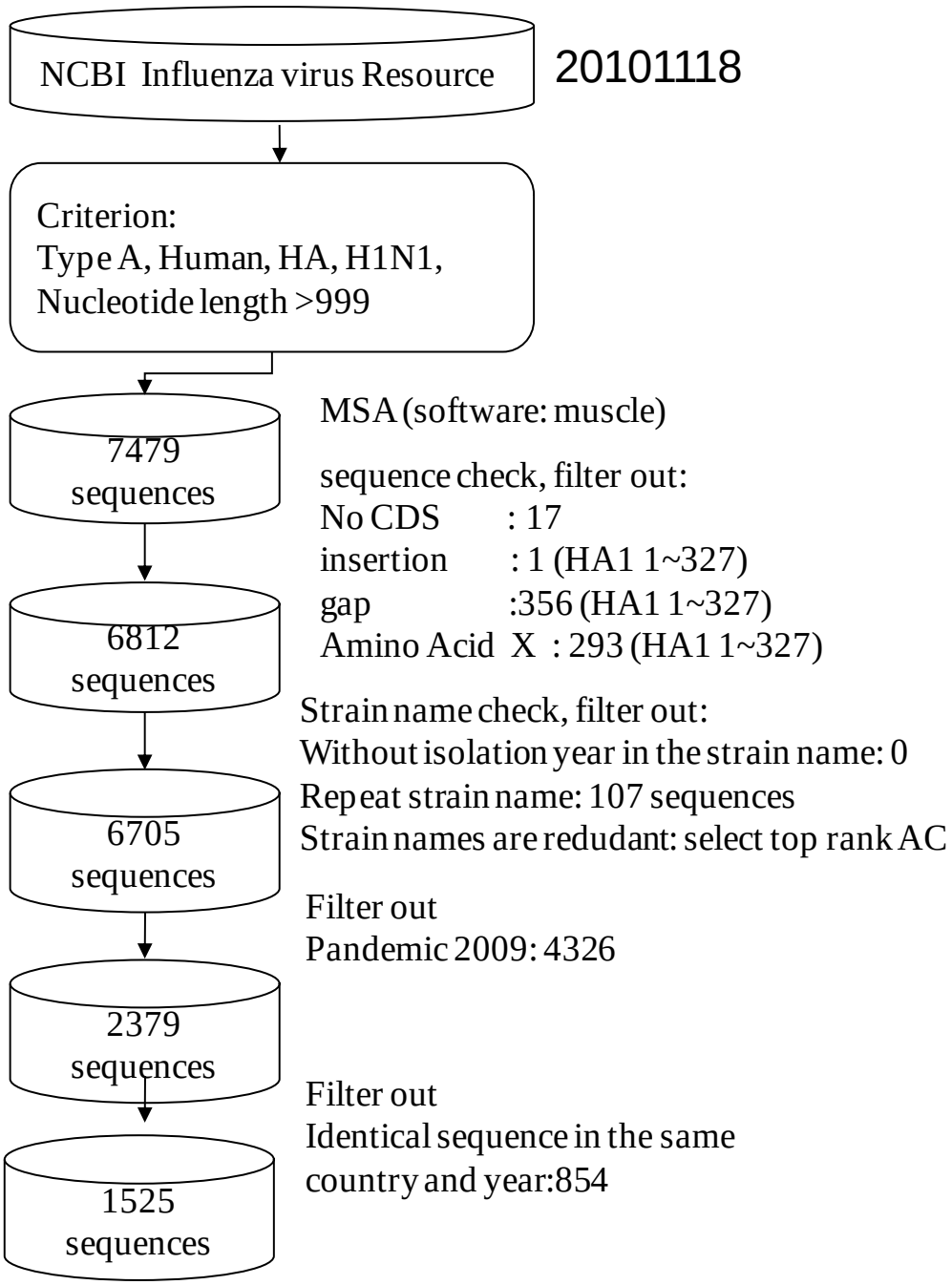
一、流感序列資料集

流行感冒病毒的資料，在 NCBI 網站下有專門收錄流感病毒的蛋白質及核苷酸資訊的資料庫 Influenza Virus Sequence Database[37]，它包含了在 GenBank 內全部三種類型(Type A,B,C)的流行感冒病毒核苷酸序列和八個蛋白質序列，以及由核苷酸序列而來的編譯區(圖五)。

為了探討 H1N1 的紅血球凝集素抗原性漂移的相關機制，我們根據 NCBI 所提供的 Influenza Virus Sequence Database 資料庫所收錄的相關 A 型流感病毒之紅血球凝集素核苷酸資訊，首先針對 H1N1 紅血球凝集素上，由以下條件，(1)Type：A (2)Host： human (3)Country/Region：any (4)Subtype：H1N1 (5)Sequence length：Min：999 以上，總共蒐集了 7479 條 H1N1 的核苷酸序列(至 2010 年 10 月 20)，再透過 GenBank 上所註記的索引值找出對應胺基酸序列，接著進行多重序列比對，經多重序列比對後的序列篩除以下有問題之序列，(1)沒有編碼序列 (2)序列發生插入 (3)序列含有間隙 (4)序列中含有任意胺基酸，再檢查病毒株名稱，過濾掉(1)病毒株名稱中沒有註記分離年份、(2)病毒株名稱重複病毒株，最後篩除下列序列，(1)2009 新型 H1N1 流感病毒 (2)在相同國家和年份的相同序列。(圖六)，最終經過這些步驟篩選，得到 1525 條 H1N1 的紅血球凝集素蛋白質序列。



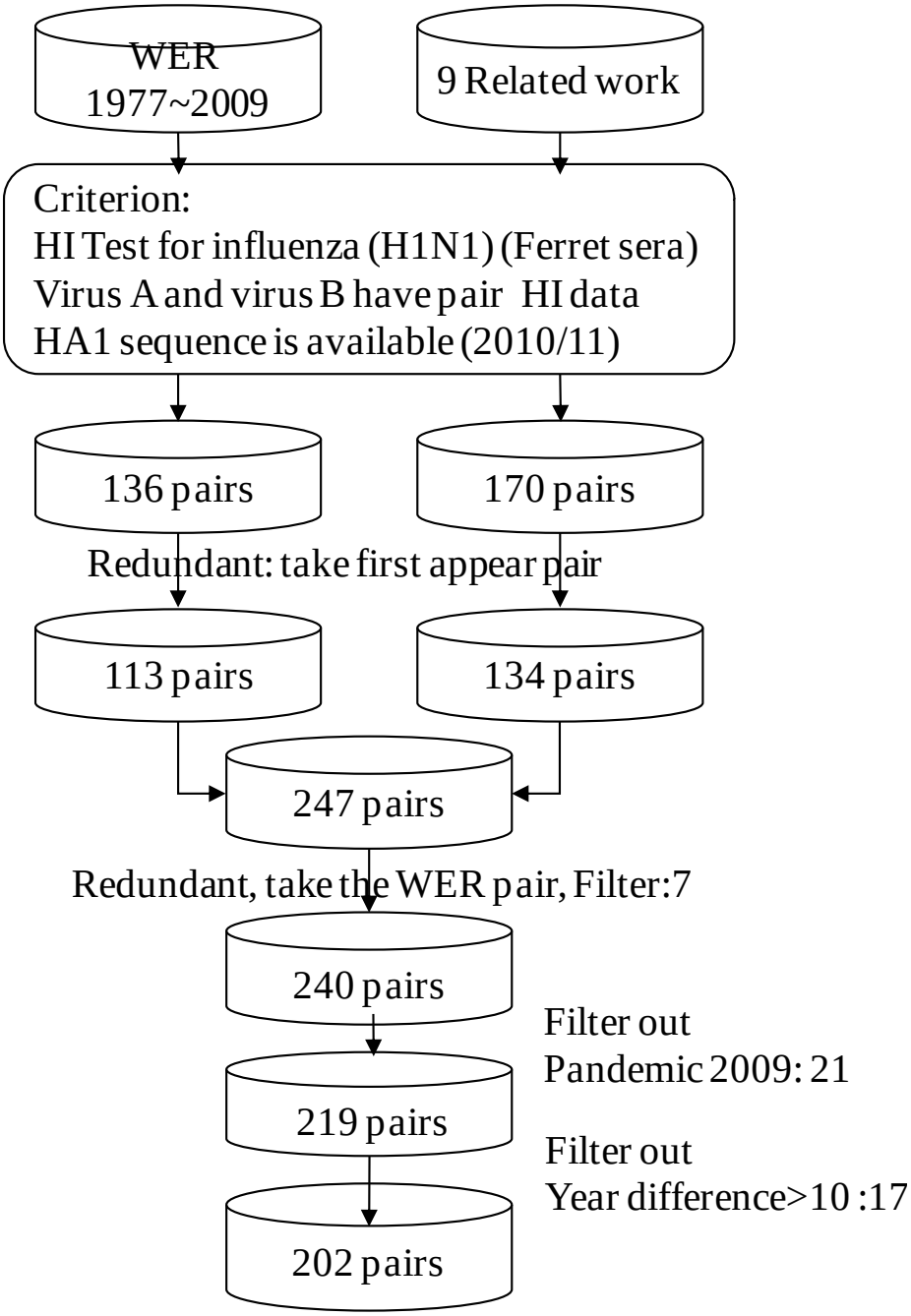
圖五、Influenza Virus Sequence Database 網頁示意圖，流感病毒紅血球凝集素核苷酸序列資料集下載資訊頁面。



圖六、H1N1 紅血球凝集素蛋白質序列資料集流程圖。

二、流感紅血球凝集素抑制測試資料集

為了蒐集 H1N1 的紅血球凝集素抑制測試的資料，我們根據世界衛生組織(WHO)所提供的 The Weekly Epidemiological Record (WER)週報上面所公布的病毒株之間成對的雪貂之紅血球凝集素抑制測試數據，針對 H1N1 收集從 1977 年至 2007 年的 WER 有關 H1N1 紅血球凝集素抑制測試數據，另外蒐集九個相關研究使用的 Centers for Disease Control and Prevention(CDC)收錄流感病毒之紅血球凝集素資料集，透過以下條件留下 WER 疫苗週報和九個相關研究的紅血球凝集素抑制測試對資料集，(1)病毒株和疫苗株配對的紅血球凝集素抑制測試 (2)成對的紅血球凝集素抑制測試各自具有紅血球凝集素 1 的蛋白質序列，再將 WER 疫苗週報和九個相關研究使用的資料集各自刪除重複紅血球凝集素抑制測試對，相同紅血球凝集素抑制測試對僅留下第一次出現的資料，再將兩者剩下紅血球凝集素抑制測試對資料集合併刪除重複，重複出現者留下 WER 疫苗週報紅血球凝集素抑制測試對為主，接著刪除 2009 年新型 H1N1 的紅血球凝集素抑制測試對，最後刪除年份有差異的 H1N1 紅血球凝集素抑制測試對，得到 202 對 H1N1 之紅血球凝集素抑制測試對(圖七)。



圖七、H1N1 紅血球凝集素抑制測試資料集流程圖。

三、從紅血球凝集素的序列和三維蛋白質結構擷取特徵

這個問題的輸入為兩個流感病毒株的紅血球凝集素之蛋白質序列，然後產生兩者的比較。最常用來比較兩個流感病毒株的方法為 hamming distance (HD)，它是計算改變胺基酸的總數[26]。但是 HD 方法不能對於改變抗原特徵解釋每一個位置上不同的重要性。我們這邊應用 position-specific change (PSC) coding，每一個位置改變為獨立紀錄作為特徵。舉例來說，在 A/Panama/2007/99 和 A/Fujian/411/2002 改變的胺基酸數目為十三個位置，所以 HD 為十三。但是位置改變的方法是獨立被記錄，記錄這十三個位置。

由於紅血球凝集素的蛋白質結構已經被結晶，並且收錄於蛋白質資料庫[39]，我們能進一步應用結構環境的資訊去找到紅血球凝集素結構上重要的區域。這邊我們應用接觸圖表的編碼，它可考慮每一個位置的環境資訊。在接觸圖表的編碼，每一個位置被認為是一個球形的中心。區域在這被定義為一個球體，它的中心在每一個胺基酸位置上。由於在紅血球凝集素上有三百二十七個位置。有三百二十七個區域在紅血球凝集素的三維結構上。如果在紅血球凝集素之三百二十七個區域上的任何位置被改變了，將這個區域視為改變。

四、抗原性距離

我們想要找到哪個位置改變會影響紅血球凝集素力價數值，所以需要定義哪種程度的紅血球凝集素力價視為改變。在這個研究中，我們將紅血球凝集素力價數值的差異程度分為兩類：抗原變異和抗原相同的例子。

從實驗而來的紅血球凝集素抑制測試數值不方便用來分析，所以紅血球凝集素抑制測試數值通常轉變為抗原性距離用來大量分析。我們應用相關研究上的方程式來定義抗原變異株[26]。這個方程式計算兩個病毒株之間的抗原性距離，方程式如下所示：

$$\sqrt{\frac{(\text{homologous I\_I})(\text{homologous J\_J})}{(\text{heterologous J\_I})(\text{heterologous I\_J})}}$$

這個方程式需要紅血球凝集素抑制測試數值的四個要件是指兩個抗血清需要被

交叉測試。一個抗原變異株被定義為當抗原性距離大於四[40]。這個意指兩個同源和異源紅血球凝集抑制測試需要有紅血球凝集抑制值等於或大於四倍。

五、亂度值

亂度值被用於測量一個空間內的紊亂程度。我們在這裡使用熵評估每個位置的紊亂程度作為在基因層次上的索引值。計算熵的方程式如下：

$$H(X) = -\sum_{r=1}^{20} P_r \log(P_r)$$

H(X)為在位置 X 上的熵，P_r 是胺基酸類型 r 在這個位置上出現的機率。在位置 X 上亂度為總和全部二十種胺基酸。特定胺基酸位置具有較大的熵值代表在此胺基酸位置較常發生胺基酸變異。

六、資訊獲得量

資訊獲得量是從具有統計意義的資訊理論而來的一個索引值。資訊獲得量測量兩個變數之間的關聯性。較大的資訊獲得量是指兩個變數之間有更大的關連性。在這個例子中，一個位置具有非常高的資訊獲得量意指如果這個位置發生改變，則有可能是抗原變異株。因此我們能使用資訊獲得量來建立基因與抗原演化之間的關係。

這裡我們使用資訊獲得量去測量每一個位置的改變影響抗原改變的程度。當我們了解 X 的數值時，給一個屬性 X 的資訊獲得量涉及一個屬性 Y 是減少關於 Y 值的不確定性。方程式如下所示：

$$I(Y,X) = H(Y) - H(Y \mid X)$$

關於 Y 值的不確定性是透過它的亂度測量，H(Y)。當我們知道 X 的數值在給定 X 之環境亂度 Y 下被給定 Y 值的不確定性，H(Y |X)。方程式三能被轉換成下列型式：

$$I(Y,X) = H(Y) - \sum_{V \in Value(X)} \frac{|Y_V|}{|Y|} H(Y_V)$$

當 Y 和 X 為離散變數帶有的數值是在{y1...yk}和{x1...xl}時，為方程式四。

七、透過資訊獲得量選擇重要的位置

對於選擇重要的關鍵概念如下所示：

假設有許多流感病毒可能的紅血球凝集素突變之模式造成病毒逃離免疫反應。所以我們會分類這些不同紅血球凝集素的突變成為好幾群。每一群突變能解釋部分過去歷史資料的抗原改變。

我們運用貪婪法選擇重要的位置。在第一層我們有完整的訓練集，而後我們選擇位置 P_1 具有最高的抗原關聯性(最高的資訊獲得量)。在第一層中這些例子有在 P_1 上突變認為可透過位置 P_1 解釋那些被解釋的例子會從原本的資料集內移除。然後在第二層，尚未被解釋的例子全部沒有在 P_1 上發生突變，所以我們找到在第二層中剩下的例子位置 P_2 具有最高的資訊獲得量。

透過遞迴選擇最高資訊獲得量的位置然後移除已解釋的例子，我們最終能找到一些位置解釋全部的例子。

決策樹是針對探索模式的複雜性數據挖掘工具，可使用它們來做預測。決策樹的方法論核心是資訊獲得量。這裡我們應用決策樹 C4.5[41]幫助我們在每一層選擇具有最高資訊獲得量的位置。

參、結果與討論

一、結果總覽

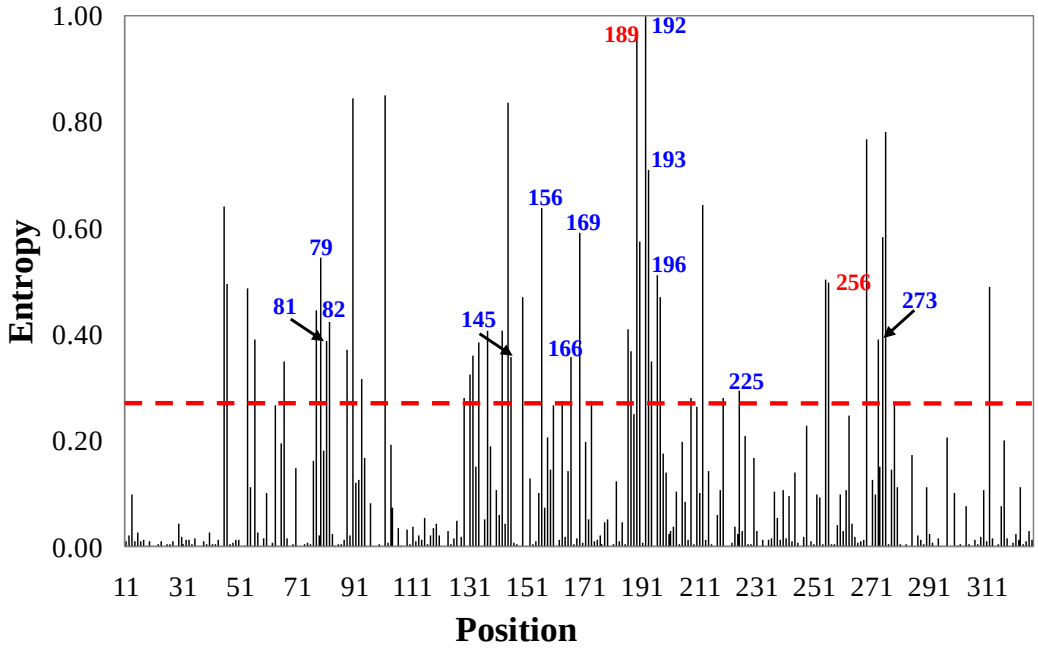
研究結果主要探討為觀察 H1N1 流感病毒在紅血球凝集素上基因性及抗原性的演化行為，而結合基因性及抗原性資訊可以幫助我們了解對於抗原性漂移相關的重要胺基酸位置，並觀察這些胺基酸位置對應於結構上 H1N1 抗原決定位關係。最後根據基因及抗原的資訊，應用決策樹透過重要的胺基酸位置預測抗原變異株並且找出規則。

二、H1N1 紅血球凝集素抗原之重要胺基酸位置

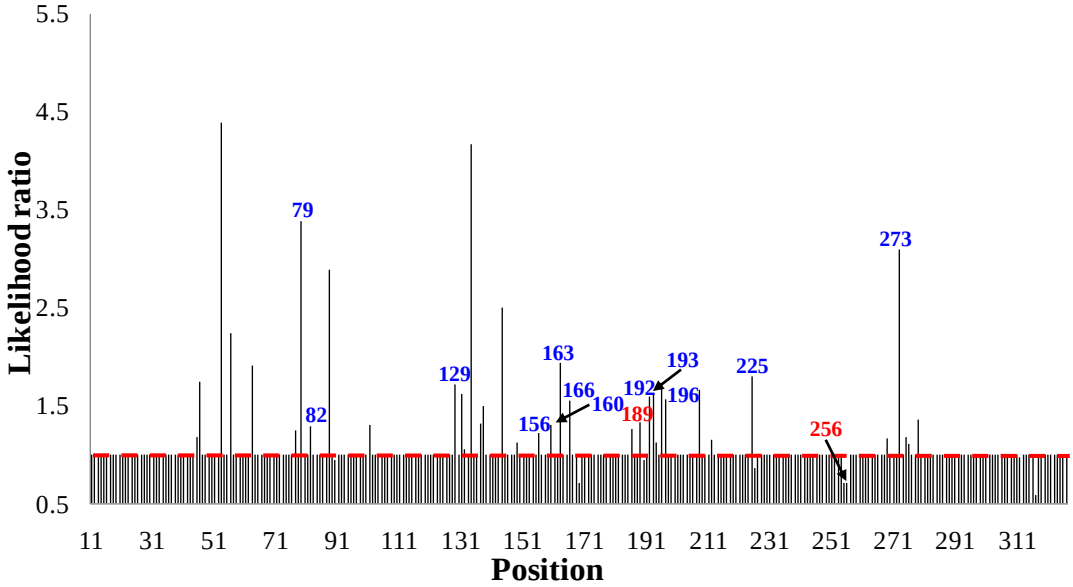
為了探討 H1N1 之紅血球凝集素抗原性飄移(Antigenic drift)的演化機制，在此研究中，分別透過三個分析構面進行研究：(1)基因層次演化分析(Genetic evolution)；(2)抗原層次分析(Antigenic evolution)；(3)結構層次分析(Structure)，主要針對抗原決定位(Epitope)及結構表面(surface)分析。透過上述構面針對紅血球凝集素上胺基酸分析，透過基因層次演化分析可挑選出常發生突變之胺基酸位置(圖八)，然而常發生突變之位置不一定影響疫苗更換(圖十四)，故透過抗原層次演化分析進一步觀察，當一個特定胺基酸位置發生改變同時需要更換疫苗的關聯程度(圖九)，透過基因及抗原層次觀察對抗原性飄移具高度影響力之胺基酸位置(圖十)，將這些位置視為對抗原性飄移之重要位置。

由於 H1N1 流感病毒之紅血球凝集素之胺基酸位置經常發生改變造成病毒株產生抗原性漂移，透過基因及抗原分析構面結合，藉由常發生突變之紅血球凝集素位置和發生改變造成病毒株產生抗原性漂移關聯之重要的位置，即同時具有熵值較高及概似比(Likelihood ratio, LR)較大的紅血球凝集素胺基酸位置，或同時具有熵值較高及資訊獲得量(Information Gain)較大的紅血球凝集素胺基酸位置，視為重要的位置。在此章節主要利用透過概似比之方法，從這兩點構面挑選出具有影響抗原性漂移的重要胺基酸位置。所挑選到的重要位置如下所示，以胺基酸

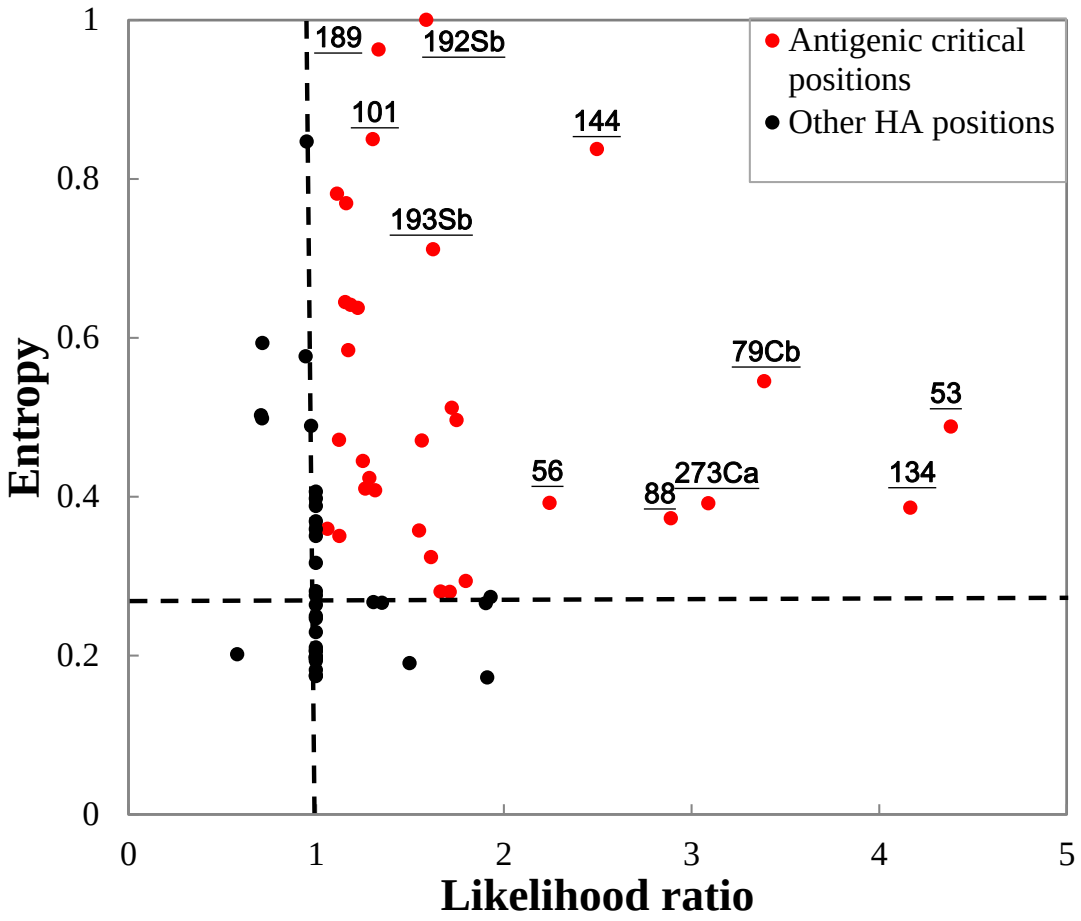
53 為例，透過亂度量度此位置之熵值為 0.49(表一)，藉由概似比評估此位置產生突變與抗原性漂移相關之重要程度為 4.38(表一)。最終總共挑選了紅血球凝集素上三十個胺基酸位置(表一)。例如：胺基酸 53、56、79、144、189 等。



圖八、H1N1 紅血球凝集素上每個胺基酸位置所對應之亂度(Entropy)。



九、H1N1 紅血球凝集素上胺基酸位置所對應之概似比(Likelihood ratio)。



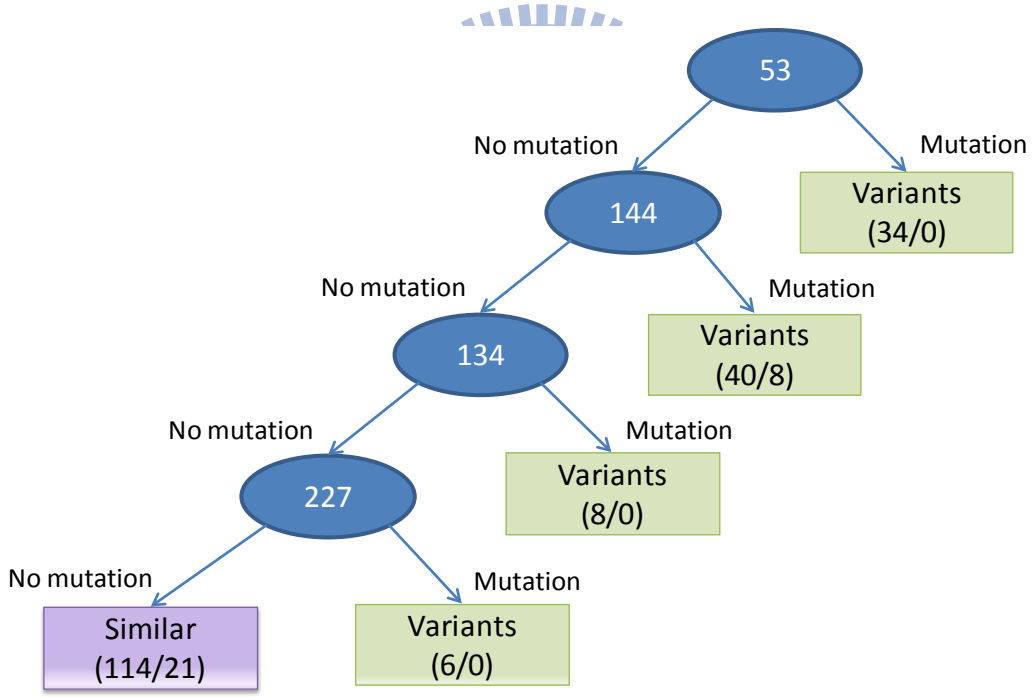
圖十、H1N1 紅血球凝集素上胺基酸位置之亂度與概似比。

表一、H1N1 之紅血球凝集素上重要胺基酸位置

Position	H1N1 Epitope	H3N2 Epitope	Surface	Entropy	Likelihood ratio
45		C	+	0.6412	1.1848
46		C	+	0.4965	1.7484
53		C	+	0.4880	4.3831
56				0.3921	2.2448
77			+	0.4450	1.2492
79	Cb			0.5452	3.3881
82	Cb	E	+	0.4232	1.2847
88		E		0.3727	2.8906
101			+	0.8500	1.3026
131		A	+	0.3239	1.6130
132		A	+	0.3594	1.0605
134			+	0.3861	4.1675
137		A	+	0.4077	1.3150
144		A	+	0.8375	2.4959
149			+	0.4713	1.1223
156	Sb	B	+	0.6375	1.2216
166	Sa			0.3574	1.5486
186		B		0.4100	1.2625
189		B	+	0.9627	1.3325
192	Sb	B	+	1.0000	1.5874
193	Sb	B	+	0.7113	1.6237
194		B	+	0.3506	1.1246
196	Sb	B	+	0.5116	1.7240
197		B	+	0.4704	1.5640
212		D		0.6447	1.1560
225	Ca2		+	0.2937	1.7984
269			+	0.7692	1.1610
273	Ca1	C	+	0.3915	3.0896
275		C	+	0.5844	1.1721
276		C	+	0.7814	1.1109

三、H1N1 紅血球凝集素抗原之決策樹

藉由決策樹程式(C4.5)建構之 H1N1 的決策樹模型中，有四個紅血球凝集素上的胺基酸位置(如：53,144,134 和 227)被挑選，挑選的原則為越上層的胺基酸位置當發生改變時對於更換疫苗相關程度越高，在這棵樹中的第一個規則是當胺基酸位置 53 發生突變時，在總共兩百零二對紅血球凝集素抑制的測試對中，總共有三十四對在此胺基酸位置發生改變且此三十四對皆需要換疫苗(圖十一)，透過決策樹應用在這個規則上可正確預測，並無預測錯誤之配對，基於此模型的分析結果，對於 H1N1 流感病毒發生改變疫苗之預測的準確率大約為 85.6% (173/202)。下圖中方框內數值斜線左邊數值代表此層中符合條件的對數，右邊數值為預測錯誤之對數。



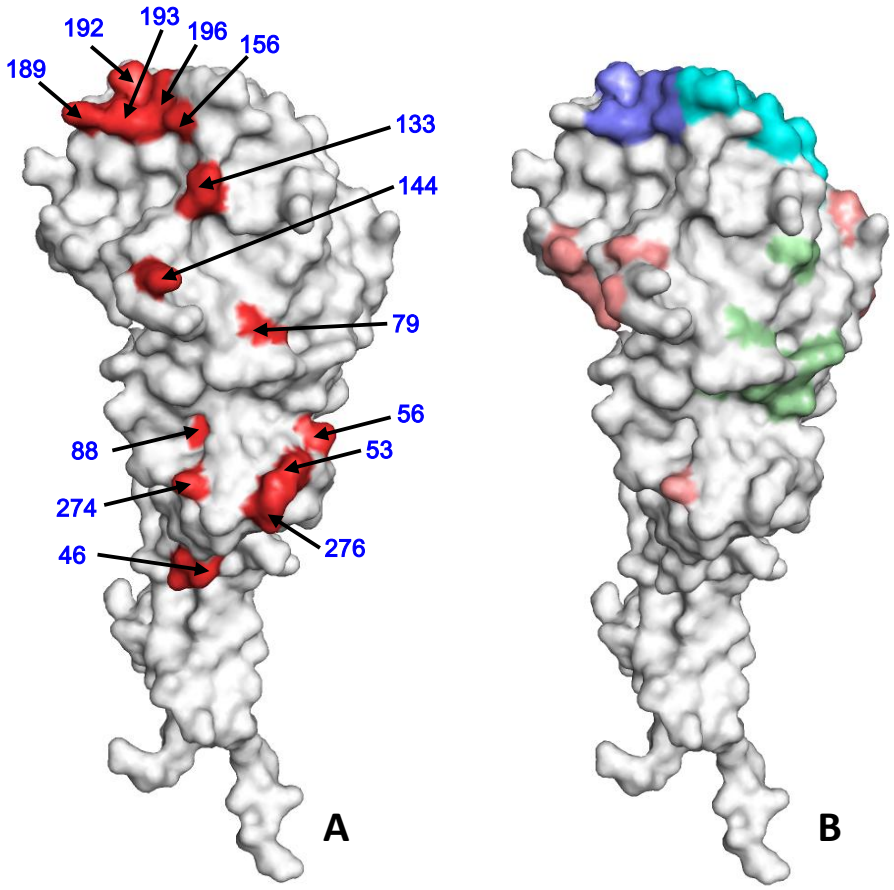
圖十一、H1N1 紅血球凝集素決策樹和預測抗原變異株規則。

四、H1N1 紅血球凝集素重要位置對應結構與抗原決定位

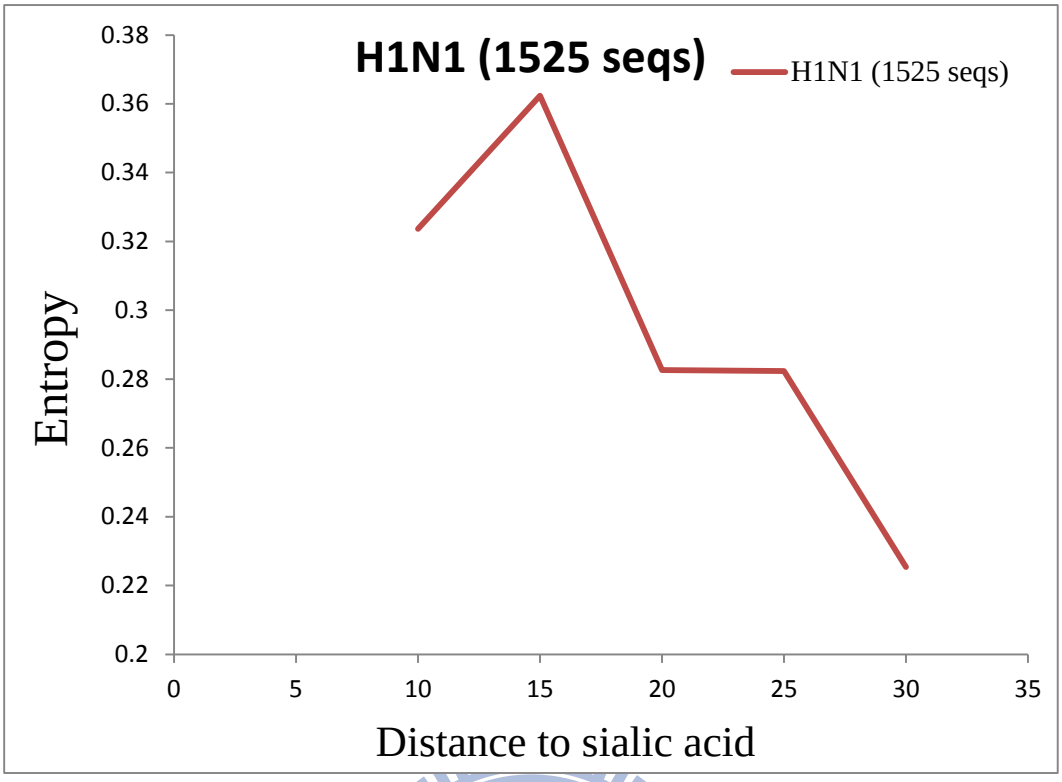
根據上述從同時具有熵值較高及概似比較大(圖十)的紅血球凝集素胺基酸位置或由 H1N1 之決策樹模型(圖十一)所挑選 H1N1 流感病毒之紅血球凝集素的關鍵胺基酸位置，透過對應於 H1N1 蛋白質結構上觀察，在這些挑選到之重要位

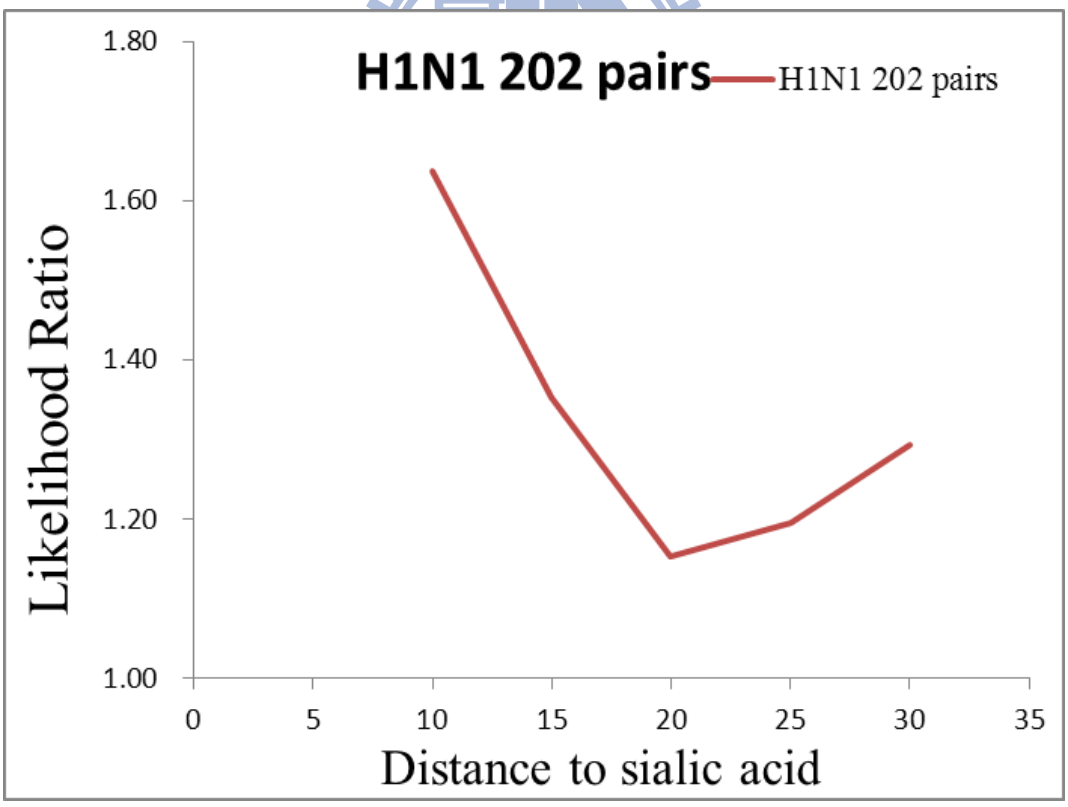
置發現對應到蛋白質結構上主要位於紅血球凝集素之表面(圖十二)，並且觀察到這些位置有部分存在於 H1N1 流感病毒總共四個抗原決定位上面(圖十二)。

另外統計在 H1N1 之紅血球凝集素上 327 個胺基酸位置對應唾液酸(sialic acid)之距離分析亂度-唾液酸之距離關係，可以發現在靠近唾液酸周圍也就是距離唾液酸 15Å 以內的區域具有較高的胺基酸熵值(圖十三)，統計在 H1N1 之紅血球凝集素上 327 個胺基酸位對應唾液酸之距離分析抗原關聯-唾液酸之距離關係，亦可發現在靠近唾液酸周圍也就是距離唾液酸 15Å 以內的區域具有較高的抗原關聯程度(圖十三)。此結果以生物角度上來說可能因為靠近唾液酸接合位的區域也是抗體辨識流感病毒之抗原辨識位，若在此處發生胺基酸突變則有機會使病毒逃離免疫辨識，進而產生抗原性漂移的現象。



圖十二、H1N1 紅血球凝集素重要胺基酸位置與對應抗原決定位。(A)在此圖中紅色標定為上述紅血球凝集素上重要之胺基酸位置。(B)抗原決定位對應圖。藍色為 Sb、青色為 Sa、紅色為 Ca、綠色為 Cb。

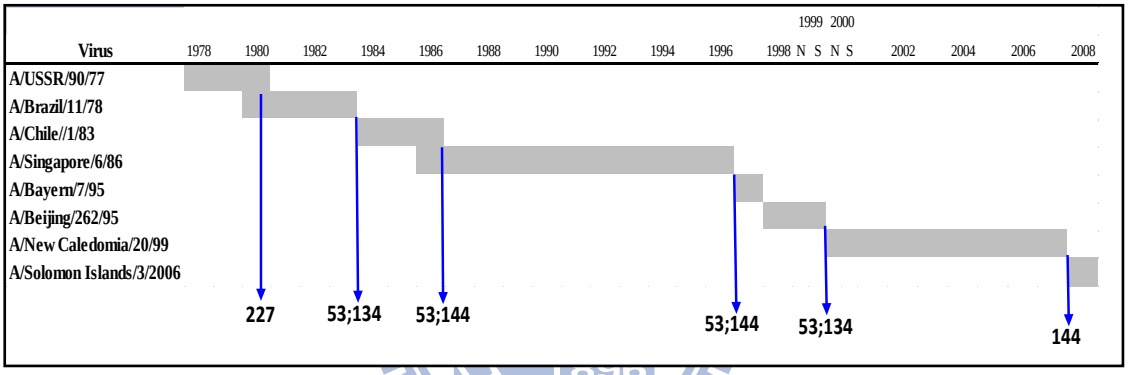




圖十三、H1N1 紅血球凝集素亂度-唾液酸之距離圖與抗原關聯-唾液酸之距離圖。

五、H1N1 紅血球凝集素歷史資料重要改變位置

觀察 H1N1 病毒疫苗株隨時間而改變的紅血球凝集素胺基酸位置，由決策樹所歸納出重要的四個胺基酸位置(如：53、134、144、227)中，透過世界衛生組織所提供由 1978 年至 2008 年流感病毒的疫苗株資訊(圖十四)，可發現至 2004 年，H1N1 流感病毒總共更換過七次疫苗株，位置 53 發生改變為四次，位置 134 發生改變兩次，位置 144 發生改變三次，位置 227 發生改變一次。由此結果可見，透過歷史資料觀察疫苗株改變的胺基酸的確符合決策樹挑選重要的胺基酸位置。

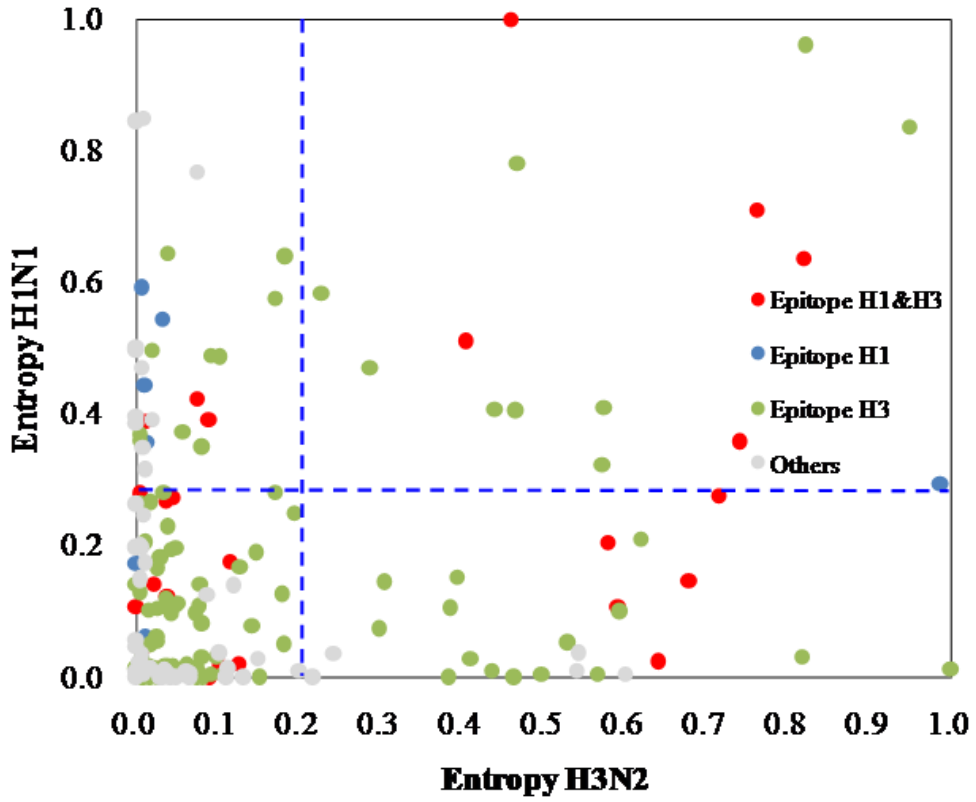


圖十四、H1N1 疫苗株時間表對應紅血球凝集素之重要胺基酸改變位置。

六、 H1N1 及 H3N2 紅血球凝集素基因性多樣性之比較

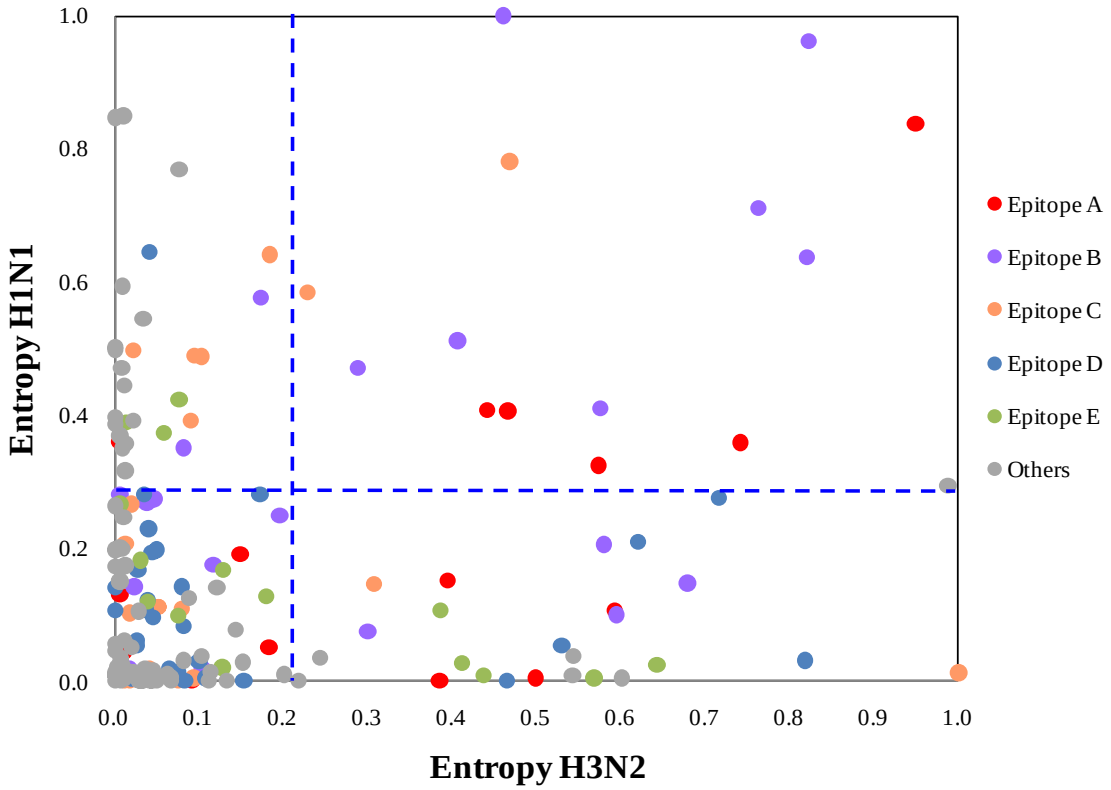
為了觀察 H3N2 和 H1N1 流感病毒之間紅血球凝集素上每個胺基酸位置改變程度，我們透過熵評估兩病毒的紅血球凝集素各個胺基酸位置改變程度，熵最大值為 1 最小值為 0。依據 H3N2 和 H1N1 熵的平均數值將紅血球凝集素上總共三百九十二個胺基酸位置劃分大致可以分為四群探討(圖十五)，首先我們觀察 H3N2 和 H1N1 兩者熵較高的胺基酸位置群聚，在此群中以胺基酸位置 189 為代表例子觀察，在這個位置上所對應的熵值在 H3N2 和 H1N1 分別為 0.82 及 0.96，此外流感病毒紅血球凝集素上胺基酸位置 189 常發生改變且在抗原變異株上亦曾發現此種位置的突變。另外過去的研究主要是集中在其中一種亞型的流感病毒紅血球凝集素觀察，在 H3N2 及 H1N1 兩者熵較高的紅血球凝集素十五個胺基酸位置上，如果僅觀察 H3N2 則有十四個胺基酸分布在 H3N2 的抗原決定位，若僅

觀察 H1N1 則有六個胺基酸分布在 H1N1 的抗原決定位。而在本研究中與過去最大差別的特色在於，同時考慮 H3N2 及 H1N1 之紅血球凝集素蛋白質之抗原決定位，發現在 H3N2 及 H1N1 兩者熵較高的紅血球凝集素十五個胺基酸位置上皆為 H3N2 或 H1N1 抗原決定位(圖十五)，因此同時考慮 H3N2 及 H1N1 抗原性功能位，可能對於紅血球凝集素上較常改變之胺基酸位置判斷可能有幫助。



圖十五、H3N2 及 H1N1 紅血球凝集素蛋白質熵分布圖。紅色為同時是 H1 及 H3 之抗原決定位上的胺基酸位置，藍色為 H1 之抗原決定位上的胺基酸位置，綠色為 H3 之抗原決定位上的胺基酸位置，灰色為非抗原決定位上的胺基酸位置。

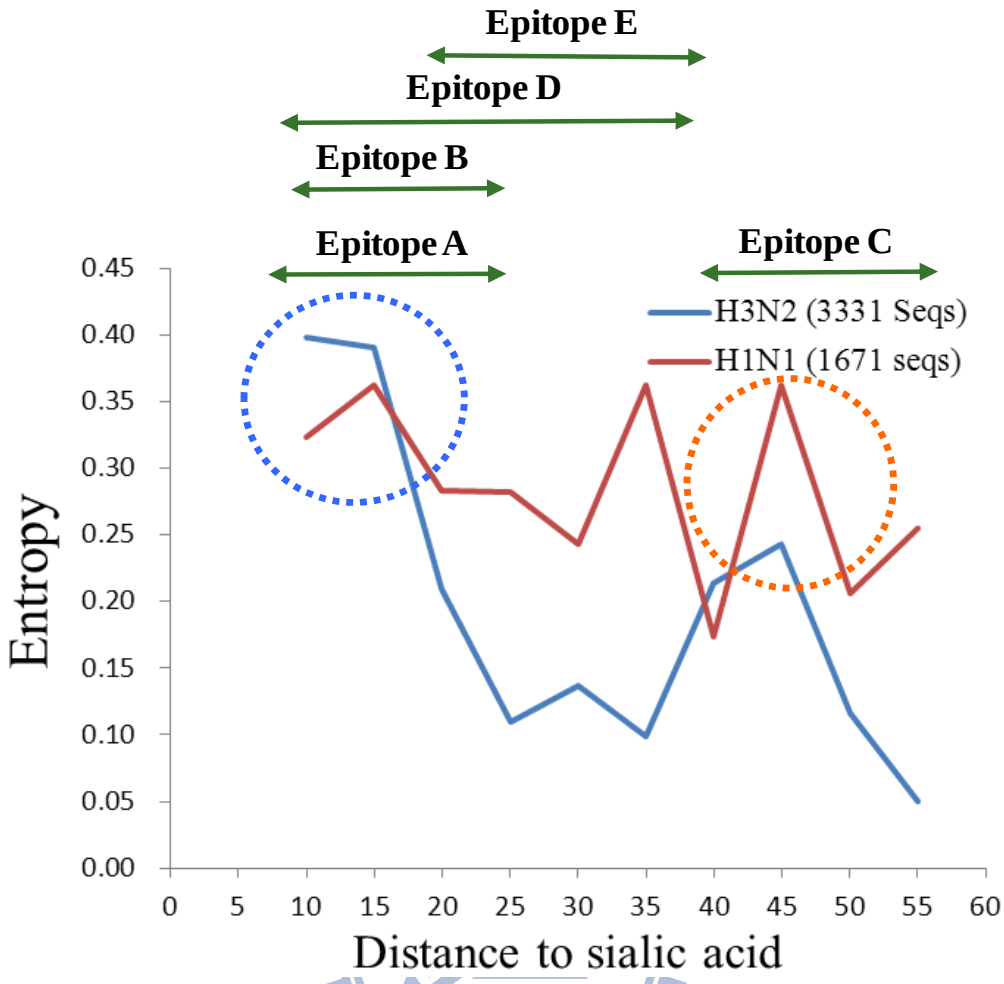
其次是觀察 H3N2 及 H1N1 之紅血球凝集素胺基酸位置在各個抗原決定位(A, B, C, D 及 E)的分布關係，在兩種亞型熵皆較高的象限中，紅血球凝集素的胺基酸位置主要集中在抗原決定位 B 上，次者為抗原決定位 A(圖十六)。由生物意義來看，抗原決定位 A 和 B 是主要接近宿主細胞上面的受體接合之處，若發生突變則有機會使流感病毒逃離宿主的免疫系統辨識造成抗原性漂移。另外僅觀察 H3N2 之紅血球凝集素熵較高且 H1N1 之紅血球凝集素熵較低的胺基酸位置，發現主要都分布在抗原決定位 D 和抗原決定位 E(圖十六)；而僅觀察 H1N1 之紅血球凝集素熵較高且 H3N2 之紅血球凝集素熵較低的胺基酸位置，分布在抗原決定位 C，顯示兩者各自紅血球凝集素常改變胺基酸位置的在抗原決定位上的差異，對於往後從基因演化角度分析此兩種不同亞型之流感病毒上，某種程度來說可能提供另一種新的分析線索。



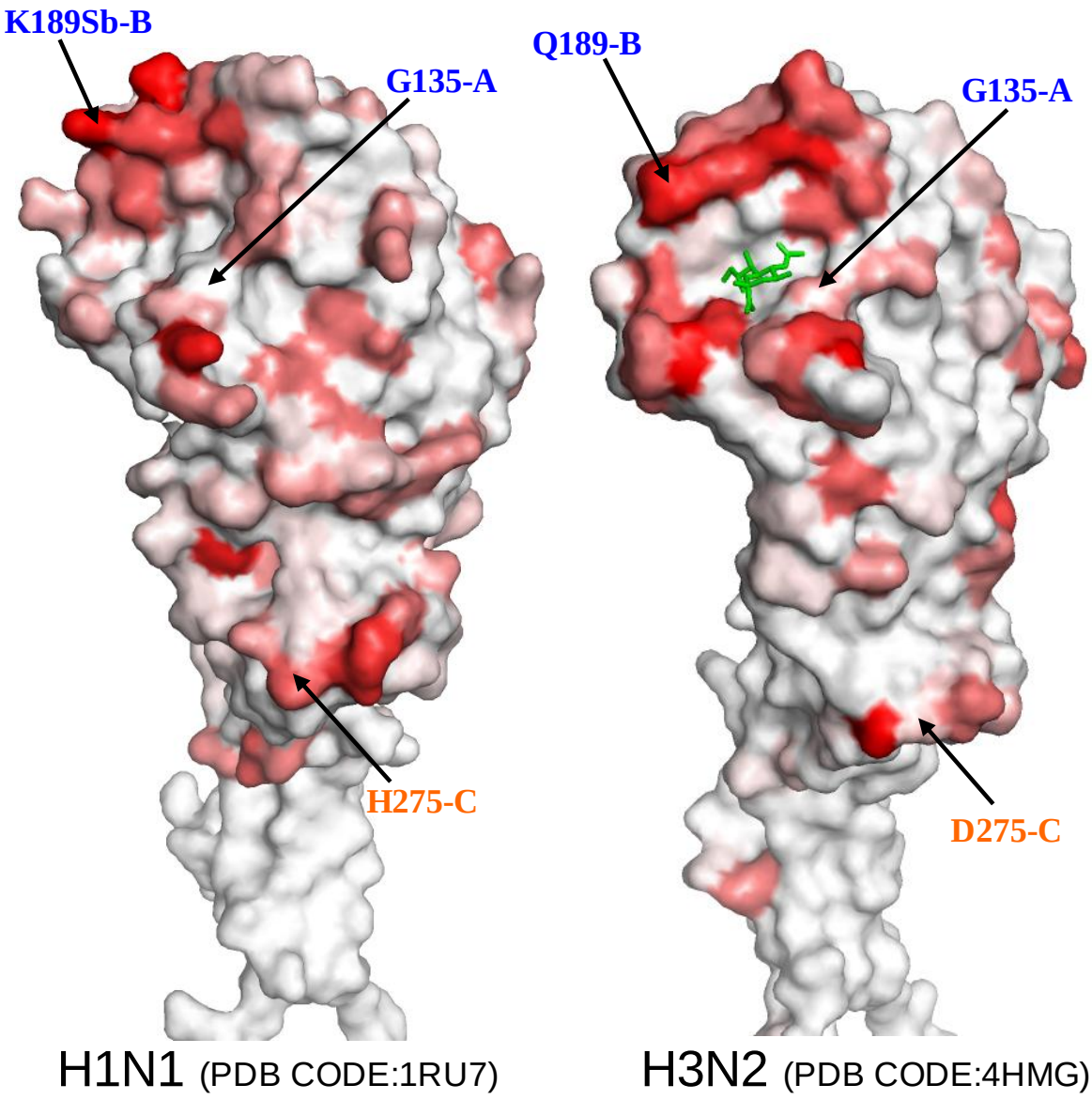
圖十六、H3N2 及 H1N1 紅血球凝集素蛋白質熵對應抗原決定位分布圖。顏色紅，紫，橘，藍，綠，灰依序代表 A 至 E 的抗原決定位上胺基酸位置。

唾液酸(sialic acid)在感染人類流感病毒中扮演著重要的角色，在流感病毒感染宿主細胞的過程中，因為流感病毒表面上具有紅血球凝集素的醣蛋白可接合在人類紅血球表面的唾液酸，透過此一機制流感病毒可以順利接合在宿主細胞上後，進入細胞體本體，達到感染細胞的目的。

為了觀察紅血球凝集素胺基酸位置的改變程度與距離唾液酸的遠近關係，我們作了以下關係圖(圖十七)，以唾液酸為中心，計算流感病毒紅血球凝集素上三百二十九個胺基酸與之距離，根據此圖可以觀察到不論是 H3N2 與 H1N1 都有共同的趨勢，在對應熵之值較大分布的兩群紅血球凝集素的胺基酸位置中，其中一群為距離唾液酸 15Å 以內的位置，在這個範圍位置包含抗原決定位有 A 與 B 等，另一群距離唾液酸在 45Å 附近的距離，其範圍位置包含了抗原決定位 C，從生物角度來看，抗原決定位 A 和 B 是靠近紅血球凝集素與宿主細胞接合之處，也是宿主的免疫抗體所辨識區域[42]，在過去研究中指出抗體為 Y 字型的蛋白質，在接合紅血球凝集素時，會透過 Y 字型的抗原接合位接觸抗原上的兩個區塊，此 Y 字型抗原接合位距離大約為 30Å，上述對應熵之值較大分布的兩群位置亦差距 30 Å，顯示在紅血球凝集素較常發生突變的位置上，對應到紅血球凝集素的蛋白質結構上觀察，看出突變較常發生的胺基酸位置同時也是在抗體辨識的區域上，此外由 H3N2 分析距離唾液酸 15Å 的胺基酸熵值最大，其次才為距離唾液酸在 45Å 處的胺基酸熵值(圖十八)，然而在 H1N1 距離唾液酸 15Å 及 45Å 的胺基酸熵值皆為最大(圖十八)，從基因構面探討以 H3N2 來說，常改變的重要位置主要分布於距離唾液酸 15Å 的 A、B、D 抗原決定位上，反觀在 H1N1 上除了上述這些抗原決定位為重要位置之外，更包含了抗原決定位 C，這也顯示出兩者之間重要改變位置的差異。



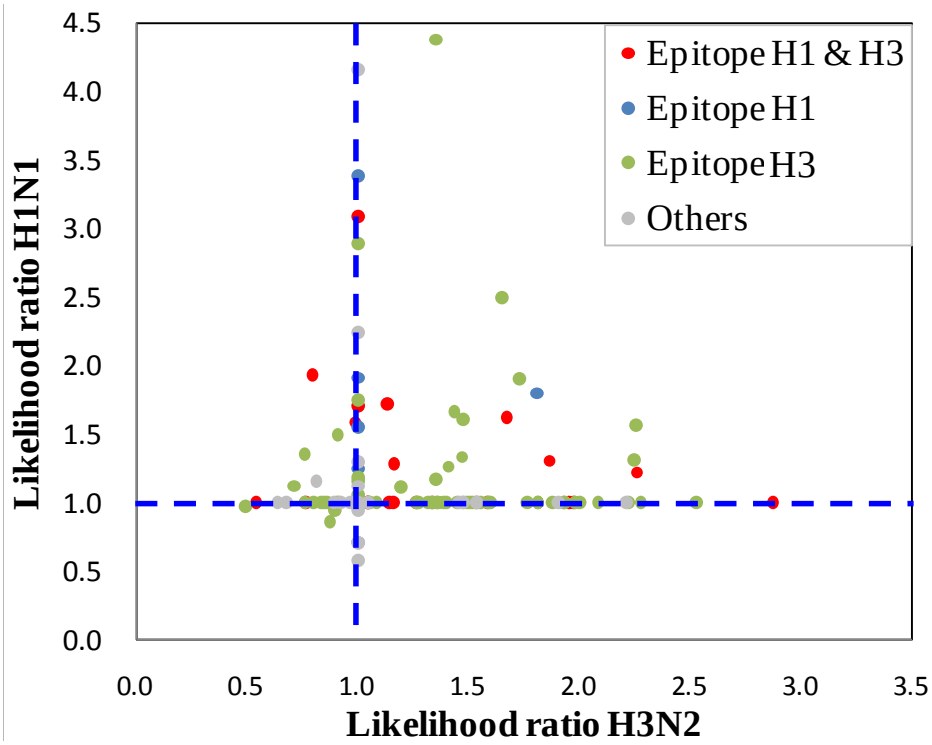
圖十七、H3N2 及 H1N1 紅血球凝集素熵與唾液酸距離對應圖。上方綠色線條分別代表各個抗原決定位分布距離。



圖十八、H3N2 及 H1N1 紅血球凝集素熵對應結構圖。標示綠色部分為唾液酸。結構上紅色代表胺基酸發生改變的位置，紅色越深代表胺基酸越常發生改變。

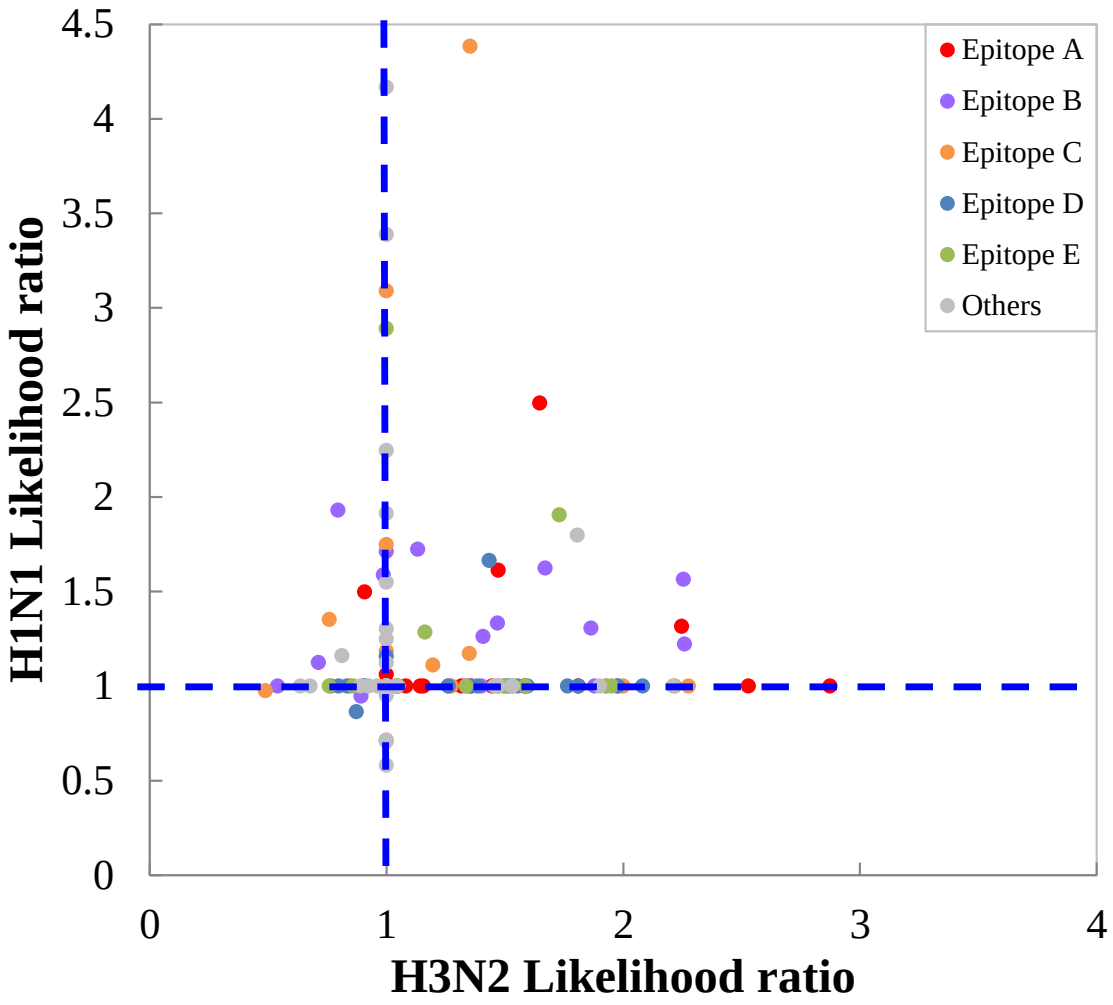
七、 H1N1 及 H3N2 紅血球凝集素抗原性演化之比較

觀察 H3N2 和 H1N1 流感病毒之間紅血球凝集素各個胺基酸位置發生改變造成病毒株產生抗原性漂移的相關程度，透過概似比(likelihood ratio, LR)及資訊獲得量(information gain, IG)評估，此章節主要利用概似比作分析方式，首先透過概似比觀察依照兩者的概似比數值為 1 區分成四個區域，從 H3N2 和 H1N1 兩者所包含的概似比(LR)皆大於 1 的區塊分析(圖十九)，與熵評估 H3N2 和 H1N1 兩病毒的紅血球凝集素各個胺基酸位置改變程度時有相似的趨勢，若僅考慮 H3N2 或 H1N1 抗原決定位時，在 H3N2 及 H1N1 兩者概似比較高的紅血球凝集素十八個胺基酸位置上，如果僅觀察 H3N2 則有十七個胺基酸包含在 H3N2 的抗原決定位，若僅觀察 H1N1 則有六個胺基酸包含在 H1N1 的抗原決定位，而同時考慮 H3N2 和 H1N1 兩者抗原決定位時，則可以看到十八個胺基酸位置皆在抗原決定位上，這說明了同時觀察 H3N2 及 H1N1 抗原決定位在評估重要之抗原性漂移的相關胺基酸時可協助挑選。



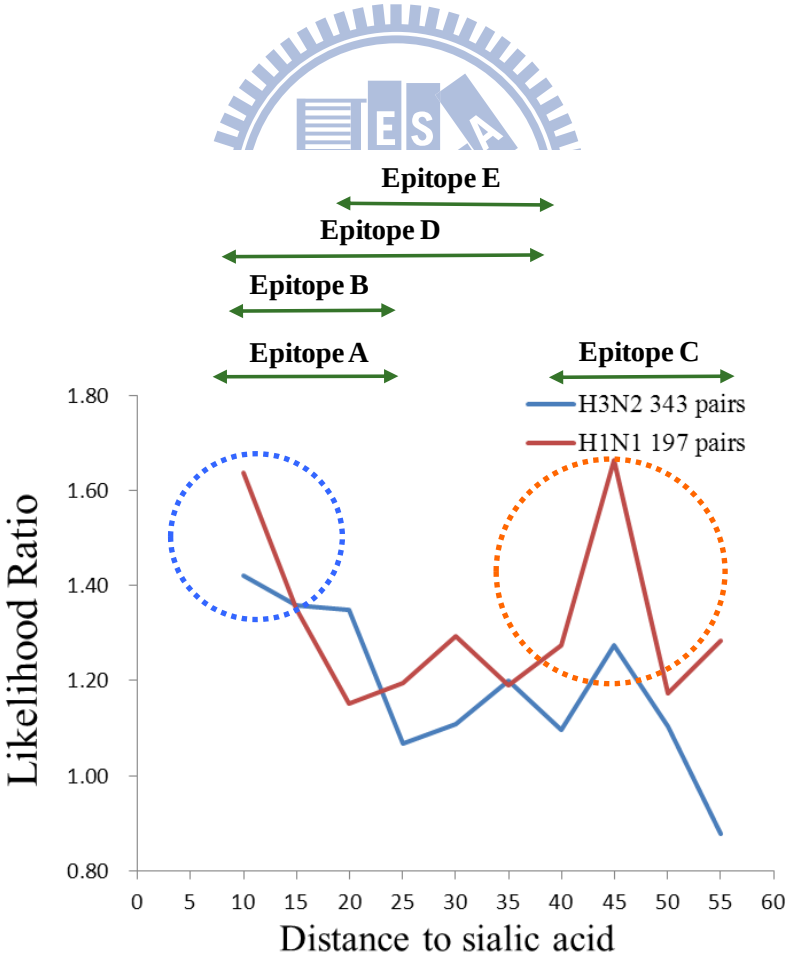
圖十九、H3N2 及 H1N1 紅血球凝集素蛋白質概似比分布圖。

觀察 H3N2 和 H1N1 流感病毒之間紅血球凝集素各個胺基酸位置發生改變造成病毒株產生抗原性漂移的相關程度，先分析兩者概似比皆大於 1 之紅血球凝集素胺基酸這些位置，在此群胺基酸總共十八個位置上，有七個胺基酸位置分布在抗原決定位 B 由前面對於熵的分布亦有相似的趨勢(圖二十)，從基因的多樣性和抗原性分析兩者都可看到，抗原決定位 A 和 B 在抗原性漂移上皆具有影響。在單看僅分布在 H3N2 流感病毒抗原決定位上的位置時，概似比數值較大的位置主要分布在抗原決定位 D 和抗原決定位 A，在單看僅分布在 H1N1 流感病毒抗原決定位上的位置時，概似比數值較大的位置卻主要分布在抗原決定位 B 和抗原決定位 C，這種差異是否可提供在流感病毒不同亞型的紅血球凝集素分析。



圖二十、H3N2 及 H1N1 紅血球凝集素蛋白質概似比對應抗原決定位分布圖。

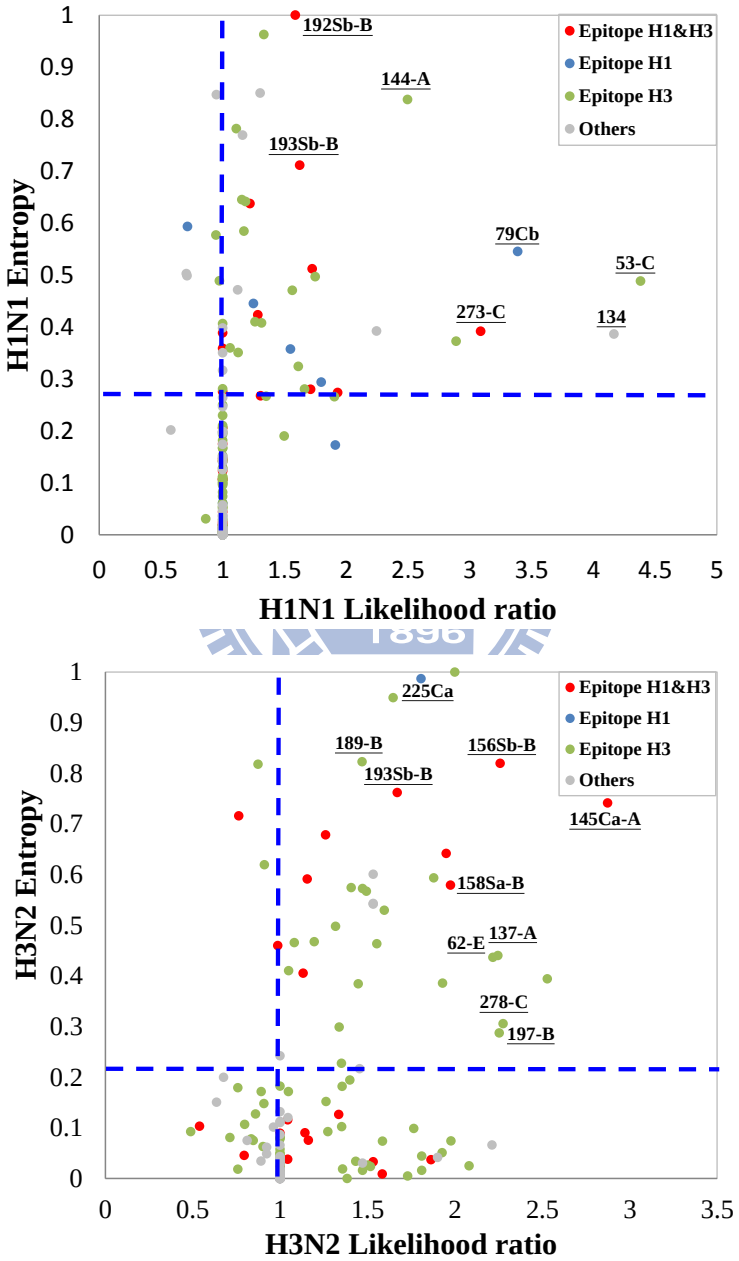
觀察概似比大小對應唾液酸接合位距離分布關係，以唾液酸為中心計算量度與紅血球凝集素上三百二十九個胺基酸上的距離，觀察下圖，X 軸為各個胺基酸對應唾液酸的距離，Y 軸為各個胺基酸之概似比數值，在 H3N2 的概似比分布趨勢發現位於距離唾液酸 10 Å 具有較高的概似比(圖二十一)，其次為距離唾液酸 45 Å 兩個區域(圖二十一)，在 10 Å 的位置附近對應到抗原決定位 A、B 及 D，而 45 Å 對應到抗原決定位 C，在這兩個區域之中分布著較高之概似比的胺基酸，顯示在這些區域中的胺基酸位置對於流感病毒抗原性漂移的影響可能扮演著重要的角色，此外觀察 H3N2 及 H1N1 亦可發現與基因多樣性角度分析有相似的趨勢，概似比較高的重要位置主要分布於距離唾液酸 15Å 的 A、B、D 抗原決定位上，而在 H1N1 上除了上述這些抗原決定位為重要位置之外，更包含了抗原決定位 C。



圖二十一、H3N2 及 H1N1 紅血球凝集素蛋白質概似比分布圖。

八、 H1N1 及 H3N2 之紅血球凝集素重要位置比較

觀察 H3N2 和 H1N1 流感病毒之間紅血球凝集素各個胺基酸位置常常發生改變造成病毒株產生抗原性漂移的重要位置，透過上述兩種分析構面結合，藉由常發生突變之紅血球凝集素位置和發生改變造成病毒株產生抗原性漂移高度相關的位置作為重要位置比較兩者(圖二十二)。

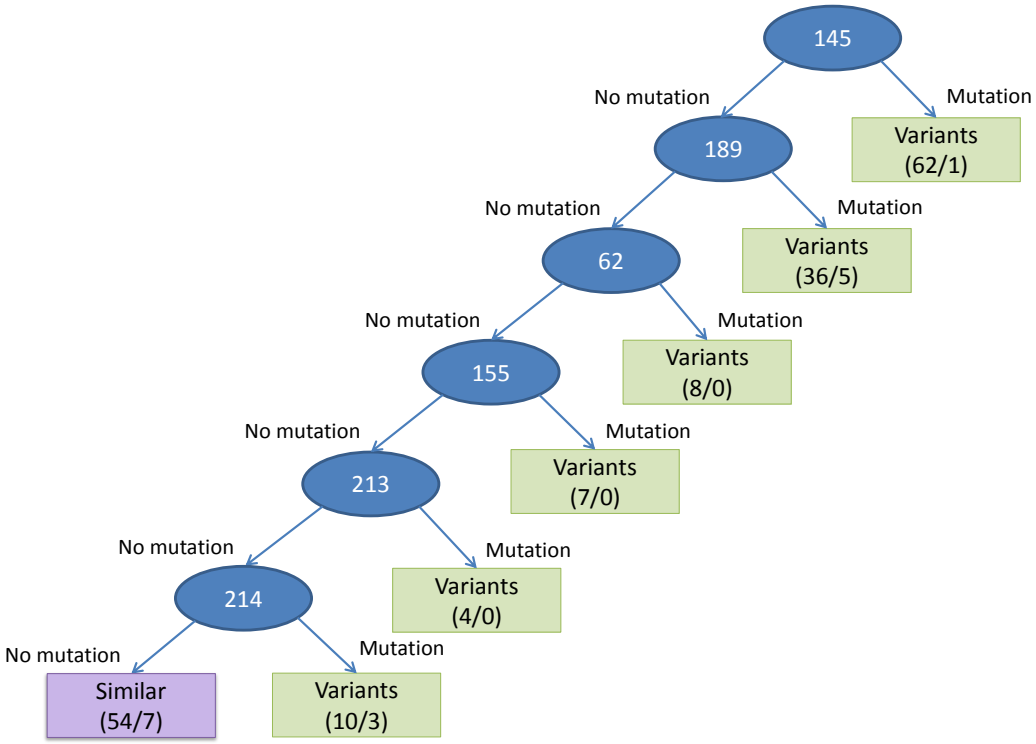


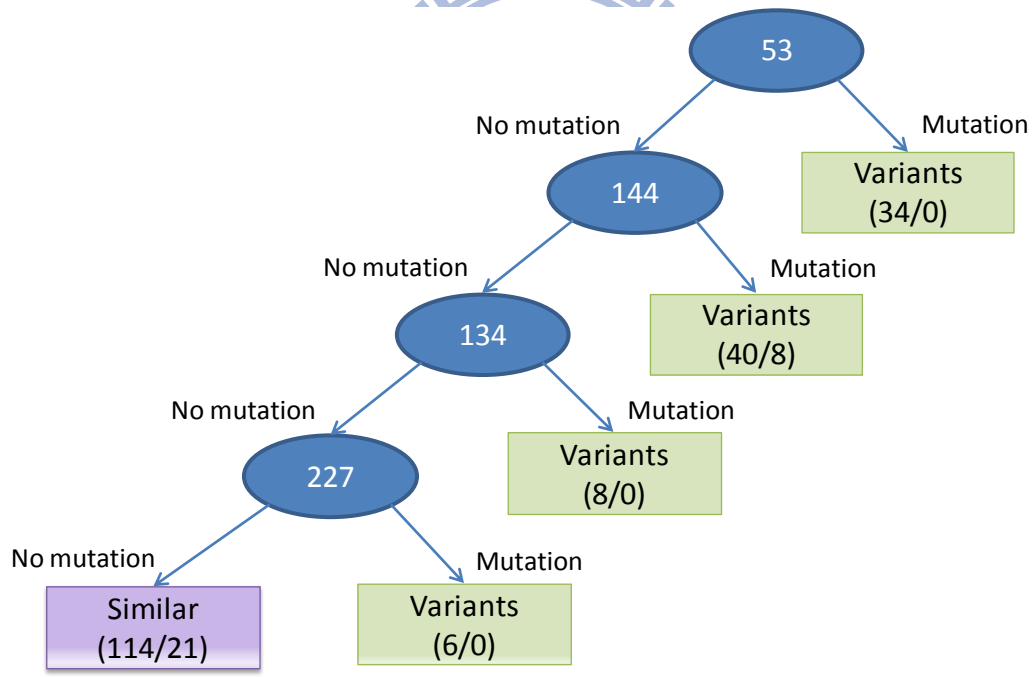
圖二十二、H1N1 及 H3N2 紅血球凝集素之熵-概似比重要位置圖。

九、 H1N1 及 H3N2 決策樹之比較

我們透過決策樹建構預測流感病毒 H3N2 及 H1N1 的抗原變異模型，在 H3N2 的決策樹中，總共有六個胺基酸位置(例如：145,189,62,155,213 和 214)被挑選到，在此決策樹中的第一個規則是當胺基酸位置 145 發生突變時，在總共一百八十一對紅血球凝集素抑制的測試對中，即測試對中序列在位置 145 的殘基類型發生改變，抗原類型為抗原變異株且可以被預測，總共有六十二對可應用在這個規則上，而有六十一對則可以被正確預測，最終決策樹的規則抗原類型為相似病毒株而且如果六個位置沒有發生突變可以被預測。基於上述決策樹的模型我們可以導出七條規則(圖二十三)，預測的準確率大約為 91.2%(165/181)結果呈現在圖中[32]。

在 H1N1 的決策樹模型中，有四個紅血球凝集素上的胺基酸位置(例如：53,144,134 和 227)被挑選，在這棵樹中的第一個規則是當胺基酸位置 53 發生突變時，在總共兩百零二對紅血球凝集素抑制的測試對中，即測試對中序列在位置 53 的胺基酸類型發生改變，抗原類型為抗原變異株且可以被預測，總共有三十四對可應用在這個規則上，三十四對則可以被正確預測，基於此決策樹模型的分析結果(圖二十三)，對於 H1N1 流感病毒發生改變疫苗之預測的準確率大約為 85.6% (173/202)。觀察兩者的發現在決策樹所預測的重要胺基酸位置皆不相同，似乎提供了兩者演化上的差異的一些線索。





圖二十三、H3N2[32]及 H1N1 紅血球凝集素決策樹和預測抗原變異株規則。

肆、結語

一、總結

為了探討 A 型流感病毒亞型 H1N1 之紅血球凝集素，透過以下幾點構面進行分析。在基因序列層次，由流感序列資料庫(Influenza Virus Sequence Database)收集了 H1N1 相關序列總共七千多條，並且透過亂度值(Shannon entropy)分析觀察紅血球凝集素上每個胺基酸位置之改變程度。在抗原層次，我們收集了近四十年的疫苗週報(Weekly Epidemiological Record)及相關文獻總共約二百筆的 HIN1 之血球凝集抑制試驗值(hemagglutination inhibition，HI)，藉由統計的方式量度每個胺基酸位置對於抗原性變化影響的程度。最終利用決策樹(C4.5)的方式來找出 H1N1 病毒抗原性漂移的規則並且進而去預測 H1N1 病毒之抗原性漂變株(antigenic variants)。另外挑選出總共 30 個 H1N1 紅血球凝集素之重要的胺基酸位置，透過兩個構面協助定義重要位置，分別為是否在紅血球凝集素之表面與抗原決定位上。最後在比較 H1N1 病毒與 H3N2 病毒在上述的分析構面下，此兩種不同亞型之間的差異。

二、主要貢獻與未來研究

本篇論文主要貢獻有下述兩者，首先，透過基因與抗原構面分析之下，本論文發現在流感病毒 H1N1 亞型重要區域主要可大致分為兩個區塊，包含鄰近受體嵌合區以及遠離受體嵌合區之抗原決定位(epitope)，相較之下，對於 H3N2 亞型之重要區域主要分布於鄰近受體嵌合區。其次，同時透過比較 H1N1 與 H3N2 亞型，發現了許多在 H3N2 亞型抗原決定位之胺基酸位置同時在 H1N1 亞型也很有可能是抗原決定位。此外，透過決策樹方法建立之模型對於 H1N1 抗原性漂移可達到 85%的預測率。

由以上的結果，顯示我們的方法具有穩健之特性並且有助於了解 H1N1 病毒之基因與抗原性演化，同時在 A 型流感病毒不同亞型 H1N1 及 H3N2 了解彼此差異，並且整理了 H1N1 之紅血球凝集素抑制測試的相關資料，對於未來預防流感病毒之疫苗設計發展及了解抗原性漂移的部份機制上提供了幫助。更進一步，未來我們將探討藉由上述構面針對數種不同流感病毒亞型進行分析，能夠大規模觀察更多不同亞型的抗原性漂移的機制與規則，並且本研究也期望對於未來流感病毒疫苗研究發展上能有所幫助。

表二、世界衛生組織(WHO)建議從 1977 至 2008 年 H1N1 之流感疫苗株

Year	Vaccine Strain	Influenza Season
1977~1980	A/USSR/90/77	1977/10/01~1980/09/30
1980~1982	A/Brazil/11/78	1980/10/01~1982/09/30
1983~1986	A/Chile/1/83	1982/10/01~1986/09/30
1985~1996	A/Singapore/6/86	1985/10/01~1996/09/30
1997	A/Bayern/7/95	1996/10/01~1997/09/30
1998~1999	A/Beijing/262/95	1997/10/01~1999/04/30
2000~2007	A/New Caledomia/20/99	1999/05/01~2007/09/30
2008	A/Solomon Islands/3/2005	2007/10/01~2008/09/30

表三、資料集內 20 張 H1N1 紅血球凝集素抑制測試圖表

Table	From To	Seq	Pairs	Variant	Equal	Period Years	Vaccine	Strain
CDC-2000	1991 2000	6	9	6	3	9	3	Beijing/262/95; Johannesburg/82/96; New_Caledoni20/99; Buenos_Aires/344/99; Peru/1621/99; Peru/1798/99
CDC-2004	1995 2003	3	2	1	1	8	2	Beijing/262/95; New Caledoni20/99; Virgini20/2003
CDC-2005	1995 2004	6	8	2	6	9	2	Beijing/262/95; Florid3/2004; Florid4/2004; Malaysi1513/2004; New Caledoni20/99; New Caledoni9/2004
CDC-2006~2007	1995 2006	7	12	1	11	11	3	Arizon1/2006; Beijing/262/95; Kentucky/1/2005; New Caledoni20/99; Pennsylvani1/2006; Singapore/107/2005; Wisconsin/73/2005
CDC-2007	1999 2007	15	55	21	34	8	5	Hong Kong/2652/2006; Hong Kong/2652/2006; Illinois/02/2006; Kentucky/2/2006; Massachusetts/03/2007; Minnesot04/2007; Missouri/12/2006; New Caledoni20/99; New Mexico/02/2006; New York/01/2007; Singapore/66/2006; Solomon Islands/3/2006; St Petersburg/8/2006; Texas/25/2006;
CDC-2008	1999 2002	12	39	24	15	9	7	Barcelon00083/2008; Brisbane/59/2007; Cambodi0371/2007; Fukushim141/2006; Hong Kong/2652/2006; Lisbon/3/2008; New Caledoni20/99; New Mexico/01/2008; New Mexico/07/2007; Solomon Islands/3/2006; South Dakot06/2007
WER-1978	1946 1977	6	6	5	1	31	4	Cameron/46; Denver/57; Fort Monmouth/1/47; Hong Kong/117/77; Malay302/54; USSR/90/77
WER-1981	1977 1980	4	6	5	1	3	3	Brazil/11/78; England/333/80; Indi6263/80; USSR/90/77
WER-1983	1977 1983	6	9	9	0	6	5	Brazil/11/78; Dunedin/27/83; England/333/80; Hong Kong/2/82; Indi6263/80; USSR/90/77
WER-1984	1978 1983	5	7	4	3	5	4	Brazil/11/78; Chile/1/83; Dunedin/27/83; England/333/80; Victori7/83
WER-1986	1978 1986	7	11	9	2	8	6	Brazil/11/78; Chile/1/83; Dunedin/27/83; England/333/80; Singapore/6/86; Taiwan/1/86; Victori7/83
WER-1992	1986 1991	4	5	3	2	5	3	Sichuan/4/88; Singapore/6/86; Taiwan/1/86; Texas/36/91
WER-1997	1986 1996	5	9	6	3	10	4	Bayern/7/95; Singapore/15/96; Taiwan/1/86; Texas/36/91; Wuhan/371/95
WER-1998	1995 1997	5	9	5	4	2	4	Bayern/7/95; Beijing/262/95; Johannesburg/159/97; Johannesburg/82/96; Wuhan/371/95
WER-1999	1995 1999	4	5	5	0	5	2	Bayern/7/95; Beijing/262/95; Johannesburg/82/96; New Caledoni20/99
WER-2007	1999 2006	3	3	2	1	7	2	Hong Kong/2652/2006; New Caledoni20/99; Solomon Islands/3/2006
WER-2008	2006 2007	5	7	2	5	1	2	Brisbane/59/2007; Norway/1735/2007; Paris/577/2007; Solomon Islands/3/2006; South Dakot06/2007

表四、流感病毒株列表

Full name	Accession no.	Vaccine	Reference
A/Alabama/22/2006	EU199334		
A/Alabama/23/2006	EU199321		
A/Arizona/1/2006	CY016369		
A/Barcelona/00083/2008	FJ654312		
A/Bayern/7/95	AJ457907	Yes	J Gen Virol. 2007 Dec;88(Pt 12):3209-13
A/Beijing/262/95	AY289928	Yes	Influenza and Other Respiratory VirusesVolume 1, Issue 3
A/Brazil/11/78	HQ008267	Yes	J Immunol. 2010 Oct 1;185(7):4284-91
A/Brisbane/59/2007	CY030230		
A/Buenos Aires/344/99	AF534031		
A/Cambodia/0371/2007	FJ375205		
A/Cameron/46	CY009596		
A/Chile/1/83	CY020437	Yes	
A/Denver/57	CY008988		
A/Dunedin/27/83	Direct entry		
A/England/333/80	Direct entry		
A/Florida/10/2007	EU516250		
A/Florida/13/2007	EU887022		
A/Florida/3/2004	CY016349		
A/Florida/4/2004	CY016350		
A/Fort Monmouth/1/47	AB043478		
A/Fukushima/141/2006	FJ654301		
A/Hawaii/02/2007	EU199322		
A/Hong Kong/117/77	CY009292		
A/Hong Kong/2/82	FJ215855		
A/Hong Kong/2652/2006	CY031342		
A/Illinois/01/2006	EU199302		
A/Illinois/02/2006	EU199340		
A/India/6263/80	CY020453		
A/Johannesburg/159/97	AJ457897		
A/Johannesburg/82/96	AJ457906		
A/Kentucky/1/2005	CY016355		
A/Kentucky/2/2006	CY016388		
A/Lisbon/3/2008	FJ654310		
A/Malaya/302/54	CY021053		
A/Malaysia/1513/2004	EF566126		
Full name	Accession no.	Vaccine	Reference

A/Massachusetts/03/2007	EU199342		
A/Memphis/01/2007	EU199314		
A/Minnesota/04/2007	EU199291		
A/Missouri/12/2006	EU199288		
A/New Caledonia/20/99	AJ344014	Yes	J Gen Virol. 2002 Apr;83(Pt 4):735-45
A/New Caledonia/9/2004	CY031330		
A/New Jersey/04/2006	EU199324		
A/New Mexico/01/2008	EU566989		
A/New Mexico/02/2006	EU100708		
A/New Mexico/07/2007	EU567015		
A/New York/01/2007	EU199329		
A/Norway/1735/2007	CY036679		
A/Paris/577/2007	EU551835		
A/Pennsylvania/1/2006	CY016370		
A/Peru/1621/99	AF268313		
A/Peru/1798/99	AF268312		
A/Puerto Rico/8/34	CY009444		
A/Sichuan/4/88	L19024		
A/Singapore/107/2005	EF566319		
A/Singapore/14/2004	EF566180		
A/Singapore/15/96	AJ457896		
A/Singapore/6/86	D00406	Yes	Nature. 2003 Mar 27;422(6930):428-33
A/Singapore/66/2006	CY030235		
A/Solomon Islands/3/2006	EU100724	Yes	Nat Biotechnol. 2009 Jun;27(6):510-3
A/South Dakota/06/2007	EU516090		
A/St Petersburg/8/2006	CY035126		
A/Taiwan/1/86	DQ508873		
A/Tennessee/02/2007	EU199305		
A/Texas/25/2006	EU199298		
A/Texas/36/91	AJ457908		
A/USSR/90/77	K01330	Yes	Science. 2001 Sep 7;293(5536):1842-5
A/Victoria/7/83	FJ743456		
A/Virginia/1/2006	CY016372		
A/Virginia/20/2003	CY016347		
A/Wisconsin/01/2007	EU100710		
A/Wisconsin/73/2005	CY016362		
A/Wuhan/371/95	AJ344022		

參考文獻

1. Stohr, K., *Influenza--WHO cares.* The Lancet infectious diseases, 2002. **2**(9): p. 517.
2. Johnson, N.P. and J. Mueller, *Updating the accounts: global mortality of the 1918-1920 "Spanish" influenza pandemic.* Bulletin of the history of medicine, 2002. **76**(1): p. 105-15.
3. Garten, R.J., et al., *Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans.* Science, 2009. **325**(5937): p. 197-201.
4. Webster, R.G. and W.J. Bean, Jr., *Genetics of influenza virus.* Annual review of genetics, 1978. **12**: p. 415-31.
5. Hayashida, H., et al., *Evolution of influenza virus genes.* Molecular biology and evolution, 1985. **2**(4): p. 289-303.
6. Fouchier, R.A., et al., *Characterization of a novel influenza A virus hemagglutinin subtype (H16) obtained from black-headed gulls.* Journal of virology, 2005. **79**(5): p. 2814-22.
7. Neumann, G., T. Noda, and Y. Kawaoka, *Emergence and pandemic potential of swine-origin H1N1 influenza virus.* Nature, 2009. **459**(7249): p. 931-939.
8. Palese, P., *Influenza: old and new threats.* Nat Med, 2004. **10**(12 Suppl): p. S82-7.
9. McHardy, A.C. and B. Adams, *The role of genomics in tracking the evolution of influenza A virus.* PLoS Pathog, 2009. **5**(10): p. e1000566.
10. Knossow, M., et al., *Mechanism of neutralization of influenza virus infectivity by antibodies.* Virology, 2002. **302**(2): p. 294-8.
11. Skehel, J.J. and D.C. Wiley, *Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin.* Annual review of biochemistry, 2000. **69**: p. 531-69.
12. Wilson, I.A. and N.J. Cox, *Structural basis of immune recognition of influenza virus hemagglutinin.* Annual Review of Immunology, 1990. **8**: p. 737-771.
13. Nelson, M.I. and E.C. Holmes, *The evolution of epidemic influenza.* Nature Reviews Genetics, 2007. **8**(3): p. 196-205.
14. Both, G.W., et al., *Antigenic Drift in Influenza Virus-H3 Hemagglutinin from 1968 to 1980 - Multiple Evolutionary Pathways and Sequential Amino-Acid Changes at Key Antigenic Sites.* Journal of Virology, 1983. **48**(1): p. 52-60.

15. Russell, C.A., et al., *Influenza vaccine strain selection and recent studies on the global migration of seasonal influenza viruses.* Vaccine, 2008. **26**: p. 0p.
16. Smith, D.J., et al., *Mapping the antigenic and genetic evolution of influenza virus.* SCIENCE, 2004. **305**(5682): p. 371-376.
17. Treanor, J., *Influenza vaccine--outmaneuvering antigenic shift and drift.* The New England journal of medicine, 2004. **350**(3): p. 218-20.
18. Gething, M.J., et al., *Cloning and DNA sequence of double-stranded copies of haemagglutinin genes from H2 and H3 strains elucidates antigenic shift and drift in human influenza virus.* Nature, 1980. **287**(5780): p. 301-6.
19. Belshe, R.B., *The origins of pandemic influenza--lessons from the 1918 virus.* N Engl J Med, 2005. **353**(21): p. 2209-11.
20. Kawaoka, Y., S. Krauss, and R.G. Webster, *Avian-to-Human Transmission of the Pb1 Gene of Influenza-a Viruses in the 1957 and 1968 Pandemics.* Journal of virology, 1989. **63**(11): p. 4603-4608.
21. Viboud, C., et al., *Multinational impact of the 1968 Hong Kong influenza pandemic: Evidence for a smoldering pandemic.* Journal of Infectious Diseases, 2005. **192**(2): p. 233-248.
22. Uyeki, T.M., *2009 H1N1 virus transmission and outbreaks.* The New England journal of medicine, 2010. **362**(23): p. 2221-3.
23. Carrat, F. and A. Flahault, *Influenza vaccine: the challenge of antigenic drift.* Vaccine, 2007. **25**(39-40): p. 6852-62.
24. WHO, *WHO guidelines on the use of vaccines and antivirals during influenza pandemics.* 2004: p. 5.
25. WHO, *WHO position paper influenza vaccines.* The Weekly Epidemiological Record, 2005. **80**: p. 277-288.
26. Lee, M.S. and J.S. Chen, *Predicting antigenic variants of influenza A/H3N2 viruses.* Emerging infectious diseases, 2004. **10**(8): p. 1385-90.
27. Finkenstadt, B.F., A. Morton, and D.A. Rand, *Modelling antigenic drift in weekly flu incidence.* Stat Med, 2005. **24**(22): p. 3447-61.
28. Cox, N.J., T.L. Brammer, and H.L. Regnery, *Influenza - Global Surveillance for Epidemic and Pandemic Variants.* European Journal of Epidemiology, 1994. **10**(4): p. 467-470.
29. Fitch, W.M., et al., *Long term trends in the evolution of H(3) HA1 human influenza type A.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(15): p. 7712-8.
30. Gerdil, C., *The annual production cycle for influenza vaccine.* Vaccine, 2003. **21**(16): p. 1776-9.
31. Gupta, V., D.J. Earl, and M.W. Deem, *Quantifying influenza vaccine efficacy*

and antigenic distance. Vaccine, 2006. **24**(18): p. 3881-3888.

32. Huang, J.W., C.C. King, and J.M. Yang, *Co-evolution positions and rules for antigenic variants of human influenza A/H3N2 viruses.* BMC Bioinformatics, 2009. **10 Suppl 1**: p. S41.
33. Bush, R.M., et al., *Predicting the evolution of human influenza A.* SCIENCE, 1999. **286**(5446): p. 1921-5.
34. Bush, R.M., et al., *Positive Selection on the H3 Hemagglutinin Gene of Human Influenza Virus A.* Molecular Biology and Evolution, 1999. **16**: p. 1457-1465.
35. Plotkin, J.B., J. Dushoff, and S.A. Levin, *Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(9): p. 6263-6268.
36. Salzberg, S., *The contents of the syringe.* Nature, 2008. **454**(7201): p. 160-1.
37. Bao, Y.M., et al., *The influenza virus resource at the national center for biotechnology information.* Journal of Virology, 2008. **82**(2): p. 596-601.
38. Macken, C., et al., *The value of a database in surveillance and vaccine selection.* Options for the Control of Influenza Iv, 2001. **1219**: p. 103-106.
39. Bernstein, F.C., et al., *Protein Data Bank - Computer-Based Archival File for Macromolecular Structures.* Journal of Molecular Biology, 1977. **112**(3): p. 535-542.
40. Schild, G.C., et al., *Antigenic variation in current human type A influenza viruses: antigenic characteristics of the variants and their geographic distribution.* Bull World Health Organ, 1973. **48**(3): p. 269-78.
41. Quinlan, J.R., *C4.5: Programs for Machine Learning*. 1993, San Mateo, CA: Morgan Kaufmann.
42. Gamblin, S.J., et al., *The structure and receptor binding properties of the 1918 influenza hemagglutinin.* Science, 2004. **303**(5665): p. 1838-42.