# A web server: ATIVS (Analytical Tools for Influenza Virus Surviellance)

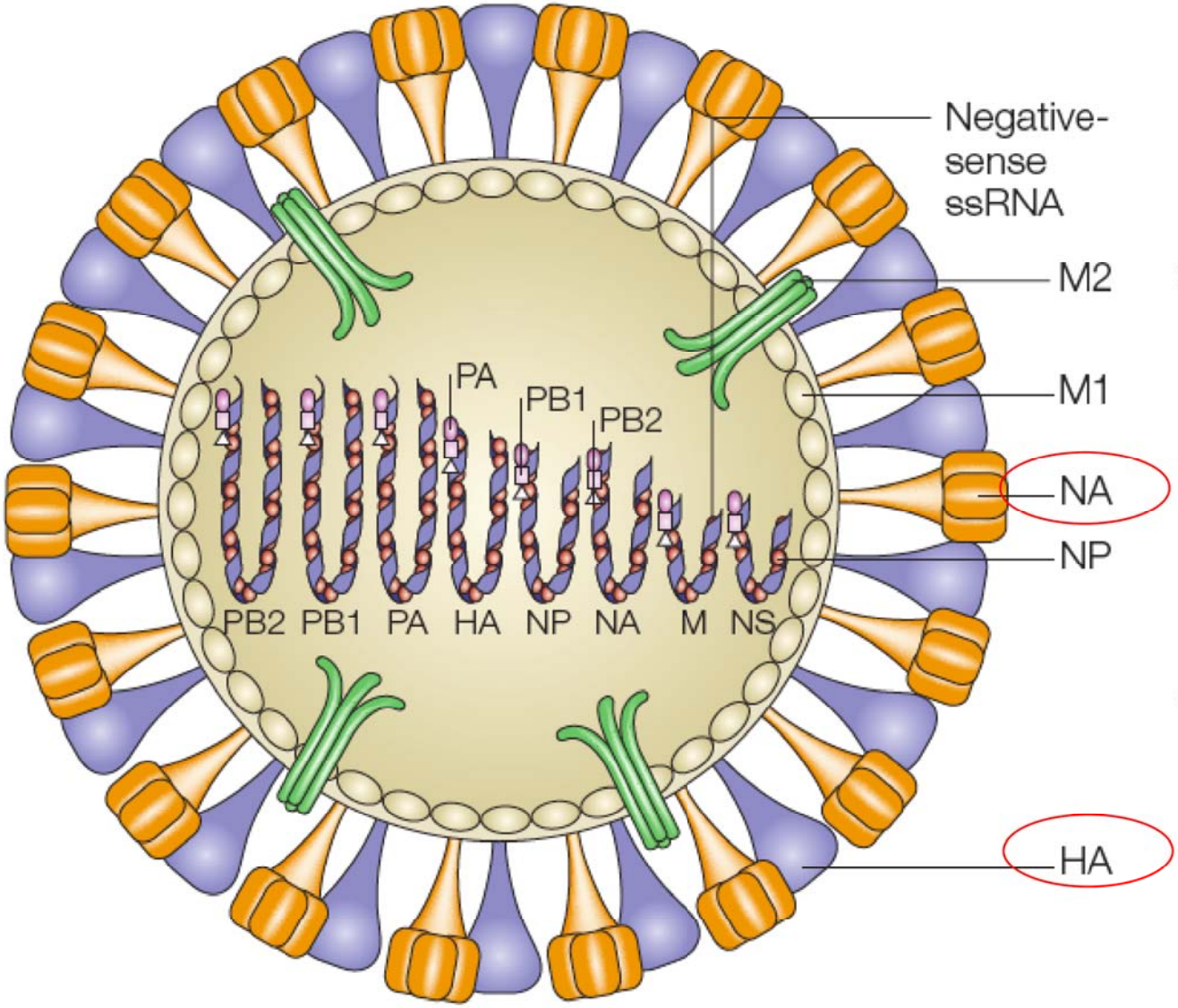
Yu-Chieh Liao, Ph.D.
Assistant Investigator
National Health Research Institutes

http://influenza.nhri.org.tw/ATVIS

(more than 2500 visits)

# Influenza viruses

# Major respiratory pathogen
# Three types




- Influenza virion contains eight pieces of segmented negative-sense RNA (13.5 kb total).
- In influenza A viruse, these eight RNAs encode 11 proteins (HA, NA, NP, M1, M2, NS1, NS2, PA, PB1, PB1-F2, PB2), while encode NB instead of M2 in influenza B virus.
- Only seven RNAs in influenza C virus.

Horimoto, Nature reviews, **3**, 591-600 (2005)

- Virulence of influenza viruses
  - Type A: most virulent and most severe
  - Type B: exclusively infect human and less common
  - Type C: less common and mild disease
- Different serotypes of influenza A
  - Based on the antigenicity of hemagglutinin (HA) and neuraminidase (NA)
  - 16 HA and 9 NA subtypes
  - A/H1N1 and A/H3N2 in human populations
- A/H1N1, A/H3N2 and B viruses ➔ vaccine are trivalent

# Epidemic and pandemic flu

- To escape ➔ New influenza viruses are constantly being produced by mutation or by reassortment.
- Accumulation of mutations
  - Antigenic drift and reduce the effectiveness of vaccines
  - Outbreak occurred in the late fall and winter
  - Result in epidemic (seasonal epidemics)
- Reassortment between different hosts
  - Antigenic shift
  - Entirely novel antigens ➔ everybody will be susceptible
  - Pandemic
- Focus on antigenic drift

# Surveillance

- Seasonal epidemics of influenza➔the WHO Global Influenza Surveillance Network was established in 1952
- The collaborative centres in the network perform antigenic and genetic analyses of viral isolates regularly.
- Antigenic analyses
  - Hemagglutinin-inhibition (HI) tests using ferret antisera.
  - It is, however, labor-intensive and time-consuming.
- Genetic analyses (RT-PCR, easily accessible)
  - Phylogenetic analysis is widely used to elucidate genetic relatedness.
  - However, phylogenetic tree still cannot confidently predict antigenicity.

# HI table

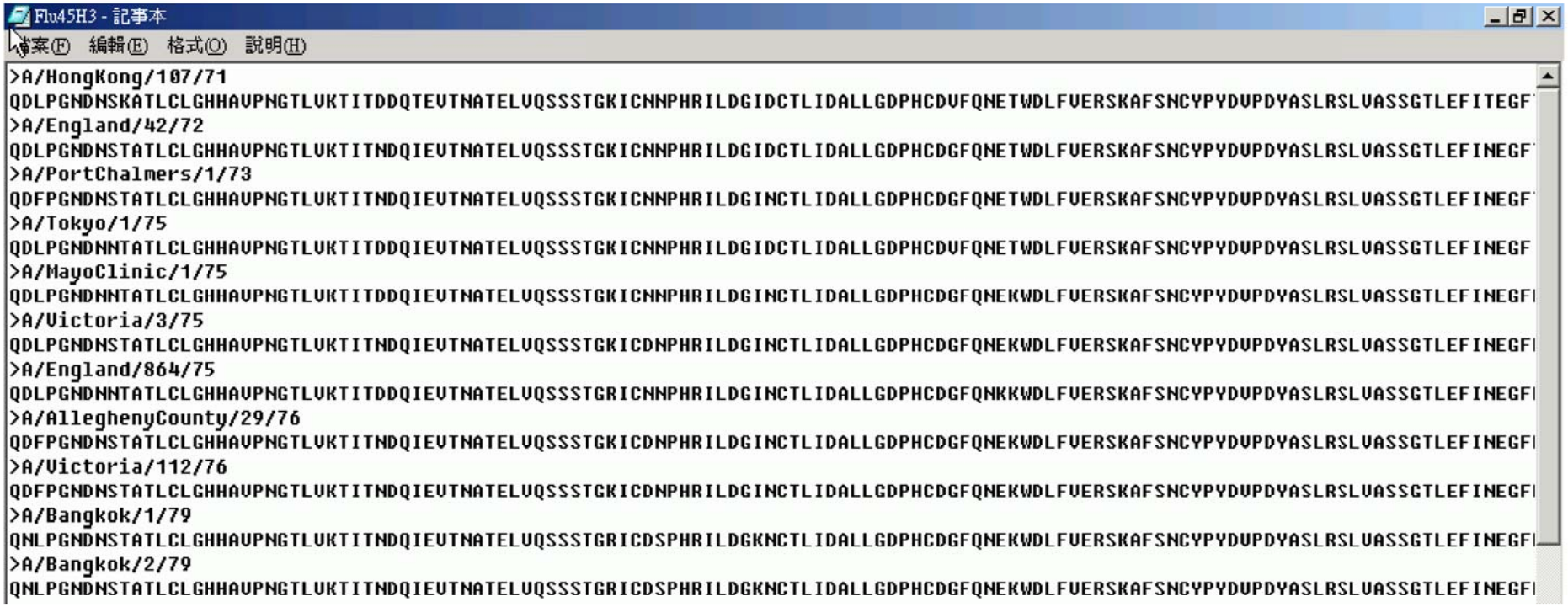
| | STRAIN DESIGNATION | REFERENCE FERRET ANTISERA | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | REFERENCE ANTIGENS | WY/3 | WEL/01 | ND/01 | CA/7 | SN/37 | Date collected | Passage |
| 1 | A/WYOMING/03/2003 X-147 * | 640 | 80 | 160 | 160 | 40 | 02/13/03 | X/E1 |
| 2 | A/WELLINGTON/01/2004 * | 160 | 320 | 320 | 320 | 320 | 01/26/04 | E3/E2 |
| 3 | A/NORTH DAKOTA/01/2004 | 80 | 160 | 640 | 320 | 160 | 10/20/04 | SpfCK1E4 |
| 4 | A/CALIFORNIA/07/2004 * | 160 | 80 | 320 | 320 | 160 | 09/16/04 | SpfCK1E3 |
| 5 | A/SINGAPORE/37/2004 | 160 | 160 | 640 | 640 | 640 | 6/7/2004 | E4/E1 |
| | TEST ANTIGENS | | | | | | | |
| 6 | A/KENTUCKY/6e/2004 | 320 | 320 | 640 | 1280 | 1280 | 11/27/04 | SpfC1KE3 |
| 7 | A/NEW YORK/57/2004 | 320 | 160 | 640 | 640 | 160 | 12/28/04 | X2/C1 |
| 8 | A/VIRGINIA/02/2005 | 320 | 320 | 1280 | 640 | 640 | 01/06/05 | M1/C1 |
| 9 | A/WISCONSIN/21e/2004 | 160 | 160 | 320 | 320 | 160 | 11/29/04 | SpfCK3E3 |
| 10 | A/NEW YORK/39e/2004 | 160 | 160 | 640 | 640 | 320 | 11/09/04 | SpfCK2E4 |

Homologous and heterologous antibody titers

More than 4 fold ➔ lower protection

# HA1 sequences

**45 viruses**




**329 amino acid residues of HA1**

# Antigenic distances

The antigenic distance between two viruses is the reciprocal of the geometric mean of two ratios between the heterologous and homologous antibody titers

**181 pairwise comparisons among 45 viruses**

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | ID | strainA | strainB | AB | lnAb |
| 2 | 1 | A/Alaska/10/95 | A/Johannesburg/33/94 | 4.00 | 1.39 |
| 3 | 2 | A/AlleghenyCounty/29/76 | A/England/42/72 | 4.00 | 1.39 |
| 4 | 3 | A/AlleghenyCounty/29/76 | A/England/864/75 | 32.00 | 3.47 |
| 5 | 4 | A/AlleghenyCounty/29/76 | A/HongKong/107/71 | 39.19 | 3.67 |
| 6 | 5 | A/AlleghenyCounty/29/76 | A/MayoClinic/1/75 | 32.00 | 3.47 |
| 7 | 6 | A/AlleghenyCounty/29/76 | A/PortChalmers/1/73 | 22.63 | 3.12 |
| 8 | 7 | A/AlleghenyCounty/29/76 | A/Tokyo/1/75 | 45.25 | 3.81 |
| 9 | 8 | A/AlleghenyCounty/29/76 | A/Victoria/3/75 | 1.89 | 0.63 |
| 10 | 9 | A/Auckland/5/96 | A/Alaska/10/95 | 2.00 | 0.69 |
| 11 | 10 | A/Auckland/5/96 | A/Fujian/47/96 | 1.00 | 0.00 |
| 12 | 11 | A/Auckland/5/96 | A/Johannesburg/33/94 | 32.00 | 3.47 |
| 13 | 12 | A/Auckland/5/96 | A/Nanchang/933/95 | 1.00 | 0.00 |
| 14 | 13 | A/Auckland/5/96 | A/NewYork/37/96 | 1.00 | 0.00 |
| 15 | 14 | A/Auckland/5/96 | A/SouthAfrica/1147/96 | 1.00 | 0.00 |
| 16 | 15 | A/Auckland/5/96 | A/Wuhan/359/95 | 2.00 | 0.69 |
| 17 | 16 | A/Bangkok/1/79 | A/AlleghenyCounty/29/76 | 78.38 | 4.36 |
| 18 | 17 | A/Bangkok/1/79 | A/England/42/72 | 78.38 | 4.36 |
| 19 | 18 | A/Bangkok/1/79 | A/England/864/75 | 6.93 | 1.94 |
| 20 | 19 | A/Bangkok/1/79 | A/HongKong/107/71 | 67.88 | 4.22 |
| 21 | 20 | A/BANGKOK/1/79 | A/LENINGRAD/360/86 | 11.31 | 2.43 |
| 22 | 21 | A/Bangkok/1/79 | A/MayoClinic/1/75 | 55.43 | 4.02 |
| 23 | 22 | A/BANGKOK/1/79 | A/MISSISSIPPI/1/85 | 5.66 | 1.73 |
| 24 | 23 | A/BANGKOK/1/79 | A/PHILLIPINES/2/82 | 11.31 | 2.43 |
| 25 | 24 | A/Bangkok/1/79 | A/PortChalmers/1/73 | 55.43 | 4.02 |
| 26 | 25 | A/BANGKOK/1/79 | A/SHANGHAI/11/87 | 64.00 | 4.16 |
| 27 | 26 | A/BANGKOK/1/79 | A/SICHUAN/2/87 | 45.25 | 3.81 |
| 28 | 27 | A/BANGKOK/1/79 | A/SYDNEY/1/87 | 32.00 | 3.47 |
| 29 | 28 | A/Bangkok/1/79 | A/Tokyo/1/75 | 39.19 | 3.67 |
| 30 | 29 | A/Bangkok/1/79 | A/Victoria/112/76 | 27.71 | 3.32 |

**AD: antigenic distance**
**>=4 antigenic variant**
**<4 similar antigens**

# Scores of pair-wise amino acid comparisons

- Quantitatively evaluate the amino acid difference of pair-wise comparisons.
- Simplest scoring method:

  0: two amino acid residues are identical
  1: otherwise

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Viruse 1 | Q | K | L | P | G | N | D | N | S | T | A | T | L | C | L | G | H | H | A | V |
| Viruse 2 | Q | K | L | V | G | N | E | N | S | T | I | T | A | C | M | G | H | H | A | V |
| Score | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |

- Each pair-wise amino acid sequences → xxx.....xx , lnAb

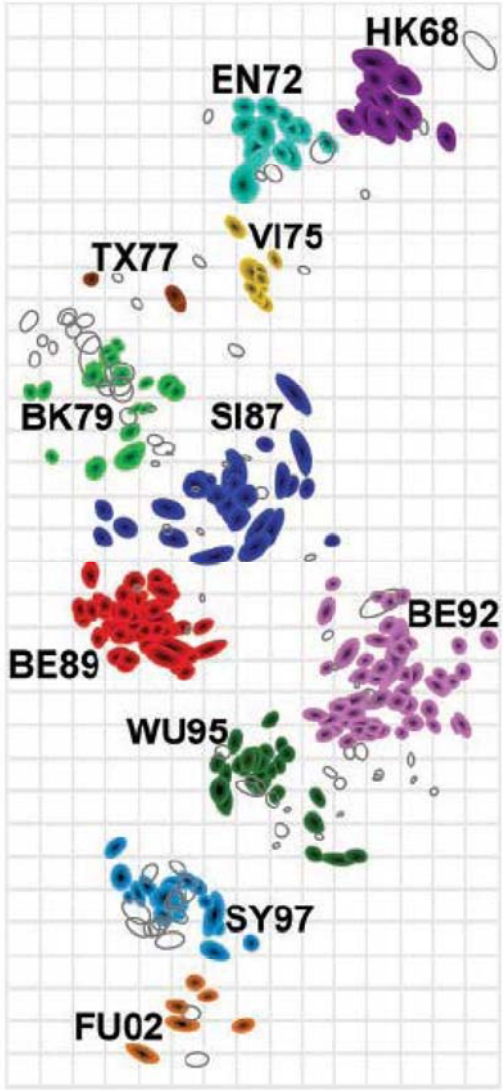
329

# Multiple regression

| Position | Antigenic domain | Residue frequency among the 45 viruses in the training dataset | GM4 | GM5 |
|---|---|---|---|---|
| AA82 | E | 17 (E), 28 (K) | 0.998 | 1.037 |
| AA92 | E | 1 (E), 44 (K) | 0.941 | 0.920 |
| AA121 | D | 28 (I), 10 (N), 7 (T) | 0.495 | 0.546 |
| AA124 | A | 13 (D), 20 (G), 2 (N), 10 (S) | 0.298 | – |
| AA129 | B | 1 (E), 44 (G) | 1.748 | – |
| AA135 | A | 1 (E), 23 (G), 6 (K), 15 (T) | 0.954 | 1.021 |
| AA144 | A | 13 (D), 3 (I), 5 (N), 24 (V) | 0.716 | 0.683 |
| AA145 | A | 1 (I), 18 (K), 22 (N), 1 (R), 3 (S) | 1.209 | 1.282 |
| AA155 | B | 30 (H), 2 (T), 13 (Y) | 1.202 | 1.582 |
| AA156 | B | 8 (E), 1 (H), 27 (K), 9 (Q) | 0.400 | 0.294 |
| AA157 | B | 26 (L), 19 (S) | 0.423 | 0.448 |
| AA158 | B | 29 (E), 7 (G), 9 (K) | 0.761 | 0.715 |
| AA160 | B | 1 (A), 35 (K), 1 (R), 1 (S), 7 (T) | 1.072 | 1.073 |
| AA173 | D | 34 (K), 11 (N) | 1.285 | 1.301 |
| AA174 | D | 40 (F), 4 (S), 1 (V) | 0.613 | 0.633 |
| AA188 | B | 42 (D), 1 (E), 1 (N), 1 (Y) | 1.087 | 1.234 |
| AA189 | B | 8 (K), 5 (Q), 8 (R), 24 (S) | 0.721 | 0.684 |
| AA240 | D | 44 (G), 1 (R) | 0.690 | 0.708 |
| AA273 | C | 44 (P), 1 (S) | 0.779 | 0.738 |
| AA276 | C | 9 (K), 14 (N), 22 (T) | 1.830 | 2.287 |
| Agreement rate in the training dataset ($N = 181$) | | | 93.37% | 92.82% |
| Agreement rate in the validation dataset ($N = 96$) | | | 91.67% | 91.67% |

Yu-Chieh Liao, Min-Shi Lee, Chin-Yu Ko and Chao A. Hsiung, Bioinformatics models for predicting antigenic variants of influenza A/H3N2 virus, Bioinformatics, **24** (2008), 505-512.

| | STRAIN DESIGNATION | REFERENCE FERRET ANTISERA | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | REFERENCE ANTIGENS | WY/3 | WEL/01 | ND/01 | CA/7 | SN/37 | Date collected | Passage |
| 1 | A/WYOMING/03/2003 X-147 * | 640 | 80 | 160 | 160 | 40 | 02/13/03 | X/E1 |
| 2 | A/WELLINGTON/01/2004 * | 160 | 320 | 320 | 320 | 320 | 01/26/04 | E3/E2 |
| 3 | A/NORTH DAKOTA/01/2004 | 80 | 160 | 640 | 320 | 160 | 10/20/04 | SpfCK1E4 |
| 4 | A/CALIFORNIA/07/2004 * | 160 | 80 | 320 | 320 | 160 | 09/16/04 | SpfCK1E3 |
| 5 | A/SINGAPORE/37/2004 | 160 | 160 | 640 | 640 | 640 | 6/7/2004 | E4/E1 |
| | TEST ANTIGENS | | | | | | | |
| 6 | A/KENTUCKY/6e/2004 | 320 | 320 | 640 | 1280 | 1280 | 11/27/04 | SpfC1KE3 |
| 7 | A/NEW YORK/57/2004 | 320 | 160 | 640 | 640 | 160 | 12/28/04 | X2/C1 |
| 8 | A/VIRGINIA/02/2005 | 320 | 320 | 1280 | 640 | 640 | 01/06/05 | M1/C1 |
| 9 | A/WISCONSIN/21e/2004 | 160 | 160 | 320 | 320 | 160 | 11/29/04 | SpfCK3E3 |
| 10 | A/NEW YORK/39e/2004 | 160 | 160 | 640 | 640 | 320 | 11/09/04 | SpfCK2E4 |




**Antigenic map**
Smith et al., Science, 2004

***CLASSICAL MDS***

The identifying aspect of *classical MDS* (CMDS) is that there is only one similarity matrix. Table I is a matrix of similarity data suitable for CMDS. It contains the flying mileages between 10 American cities. The cities are the "objects." and the mileages are the "similarities." An MDS of these data gives the picture in Fig. 1, a map of the relative locations of these 10 cities in the United States. This map has 10 points, one for each of the 10 cities. Cities that are similar (have short flying mileages) are represented by points that are close together, and cities that are dissimilar (have large mileages) by points far apart.

**Table 1 Flying Mileages Between 10 American Cities**

| Atlanta | Chicago | Denver | Houston | Los Angeles | Miami | New York | San Francisco | Seattle | Washington, DC | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 587 | 1212 | 701 | 1936 | 604 | 748 | 2139 | 2182 | 543 | Atlanta |
| 587 | 0 | 920 | 940 | 1745 | 1188 | 713 | 1858 | 1737 | 597 | Chicago |
| 1212 | 920 | 0 | 879 | 831 | 1726 | 1631 | 949 | 1021 | 1494 | Denver |
| 701 | 940 | 879 | 0 | 1374 | 968 | 1420 | 1645 | 1891 | 1220 | Houston |
| 1936 | 1745 | 831 | 1374 | 0 | 2339 | 2451 | 347 | 959 | 2300 | Los Angeles |
| 604 | 1188 | 1726 | 968 | 2339 | 0 | 1092 | 2594 | 2734 | 923 | Miami |
| 748 | 713 | 1631 | 1420 | 2451 | 1092 | 0 | 2571 | 2408 | 205 | New York |
| 2139 | 1858 | 949 | 1645 | 347 | 2594 | 2571 | 0 | 678 | 2442 | San Francisco |
| 2182 | 1737 | 1021 | 1891 | 959 | 2734 | 2408 | 678 | 0 | 2329 | Seattle |
| 543 | 597 | 1494 | 1220 | 2300 | 923 | 205 | 2442 | 2329 | 0 | Washington, DC |

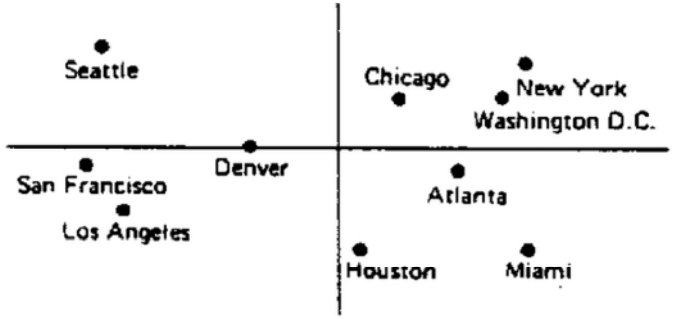



**Figure 1** CMDS of flying mileages between 10 American cities.

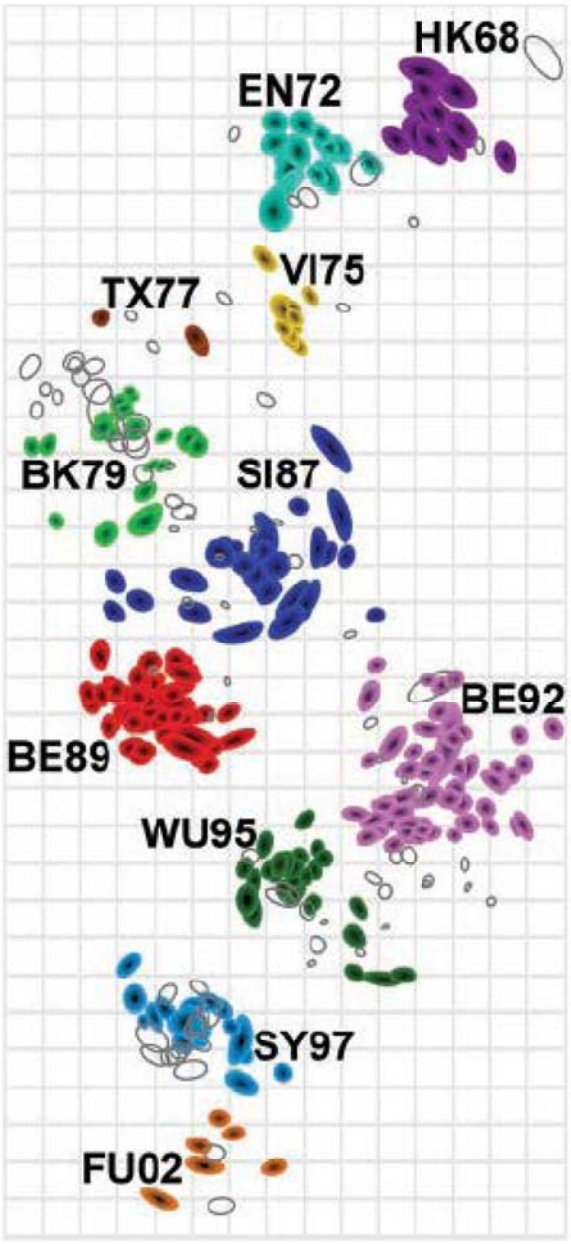
http://forrest.psych.unc.edu/teaching/p208a/mds/mds.html

# Antigenic Map of Human Influenza A (H3N2) virus

**(Smith et al., Science, 2004)**

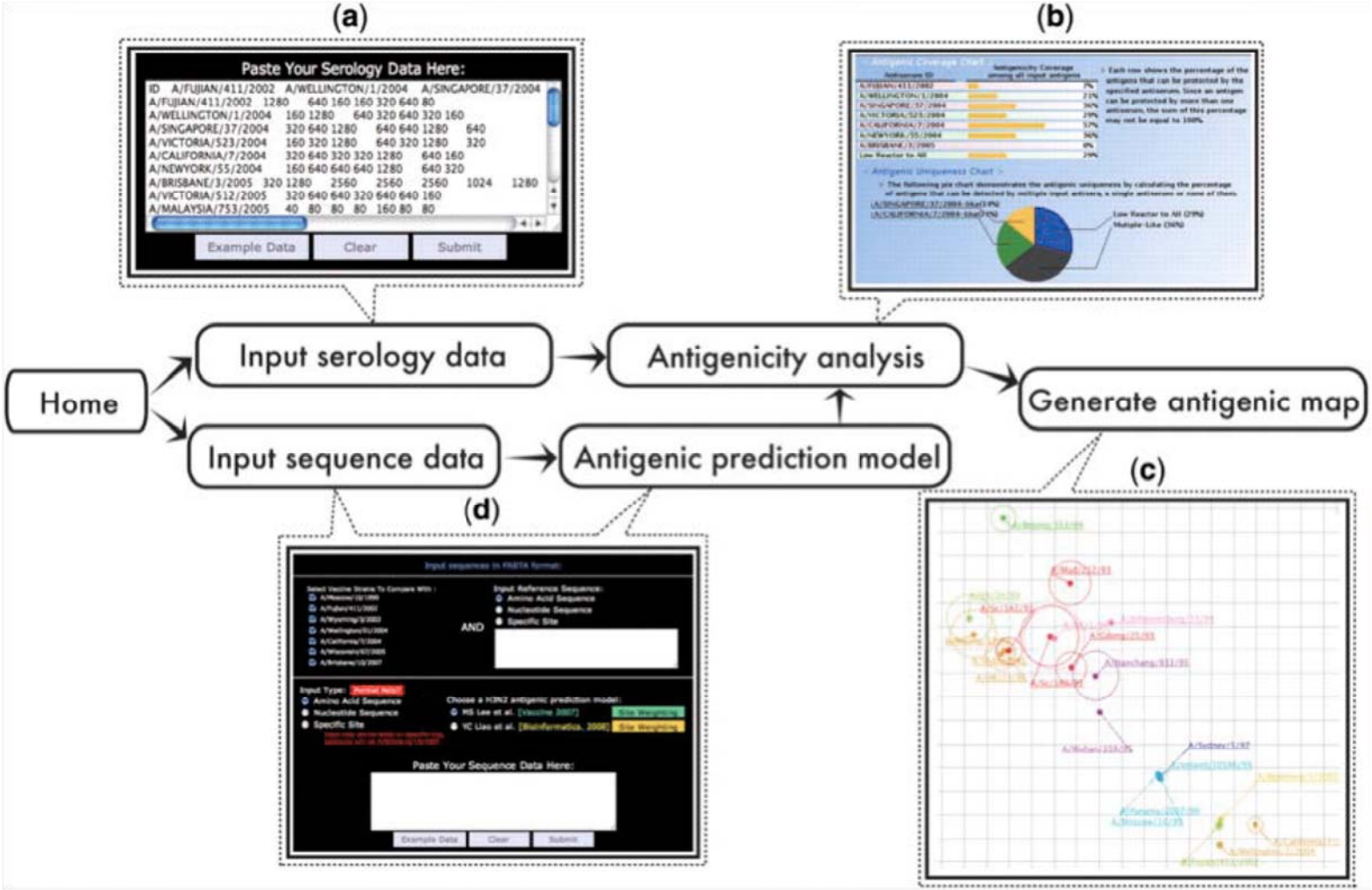
- Each antigen and antiserum is assigned a point in an "antigenic map" such that the distance between an antigen / antiserum in the map corresponds to the HI measurement.
- Use multidimensional scaling to position the antigens and antisera in the map.
- Provide the antigenic distances, define antigenic clusters

- Antigenic data not available
- Software not available



# ATIVS (Analytical Tools for Influenza Virus Surviellance)

- Analyzing serological data of all influenza viruses and hemagglutinin sequence data of human influenza A/H3N2 viruses
  ➔ generate antigenic maps



# Serological data analysis

- Two supporting to summarize the antigenic relationship

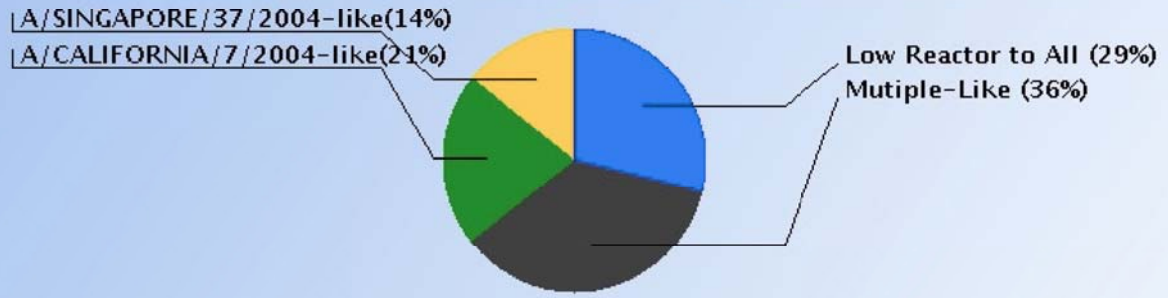
< Antigenic Coverage Chart >

| Antiserum ID | Antigenicity Coverage among all input antigens |
|---|---|
| A/FUJIAN/411/2002 | 7% |
| A/WELLINGTON/1/2004 | 21% |
| A/SINGAPORE/37/2004 | 36% |
| A/VICTORIA/523/2004 | 29% |
| A/CALIFORNIA/7/2004 | 57% |
| A/NEWYORK/55/2004 | 36% |
| A/BRISBANE/3/2005 | 0% |
| Low Reactor to All | 29% |

> The bar chart is to determine which reference strains provide optimal antigenic coverage rate against the circulating strains. Since an antigen can be protected by more than one antiserum, the sum of this percentage

< Antigenic Uniqueness Chart >

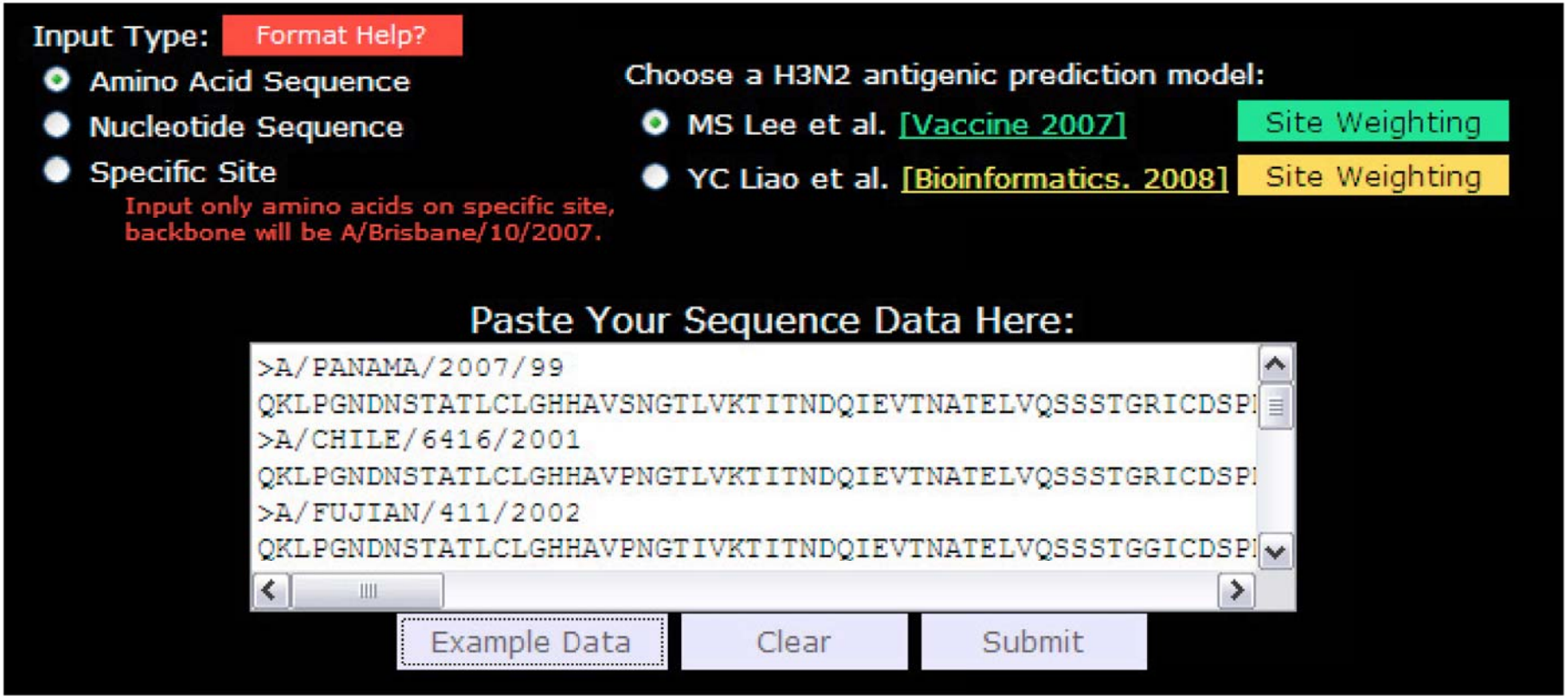
> The pie chart demonstrates the antigenicity distribution of the circulating strains that can be covered by the reference strains (single, multiple or low-reactors to all).



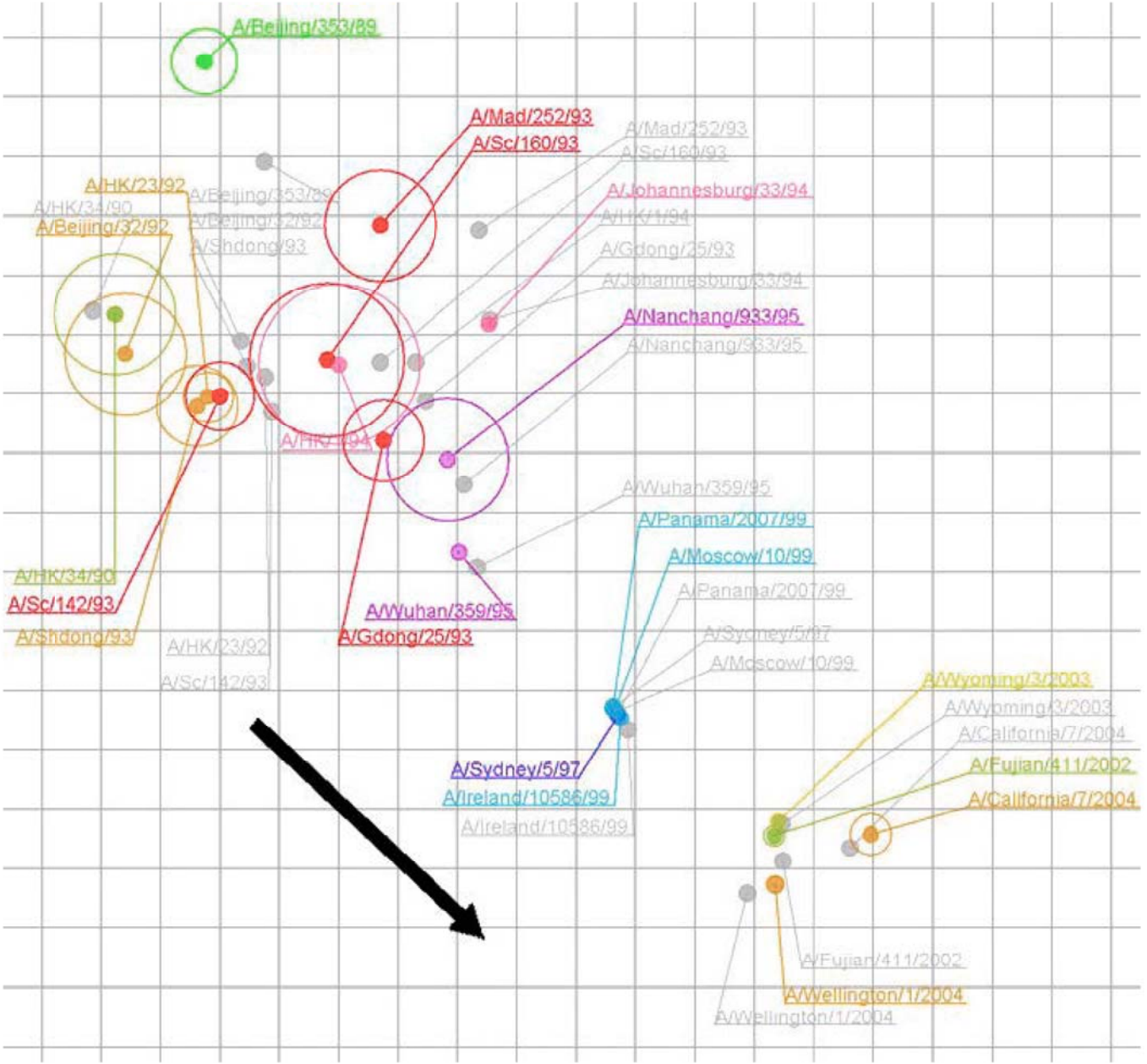

- The bar chart: determine which reference strains provide optimal antigenic coverage rate against the circulating strains.
- The pie chart: demonstrates the antigenicity distribution of the circulating strains that can be covered by the reference strains

# Sequence data analysis

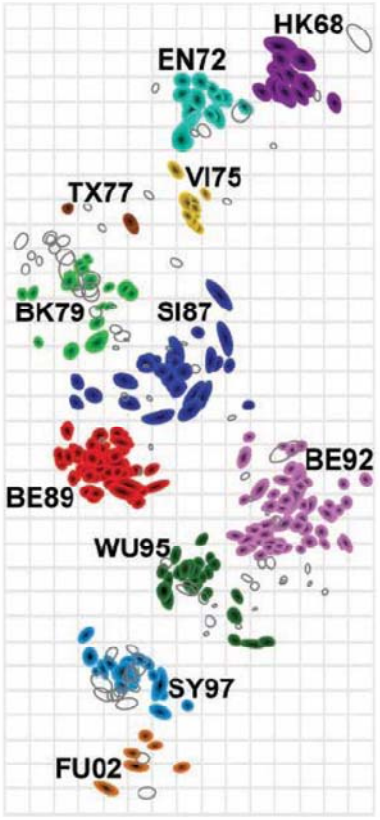
- Based on the relationship between the genetic differences and the antigenic distances for predicting antigenic variants.

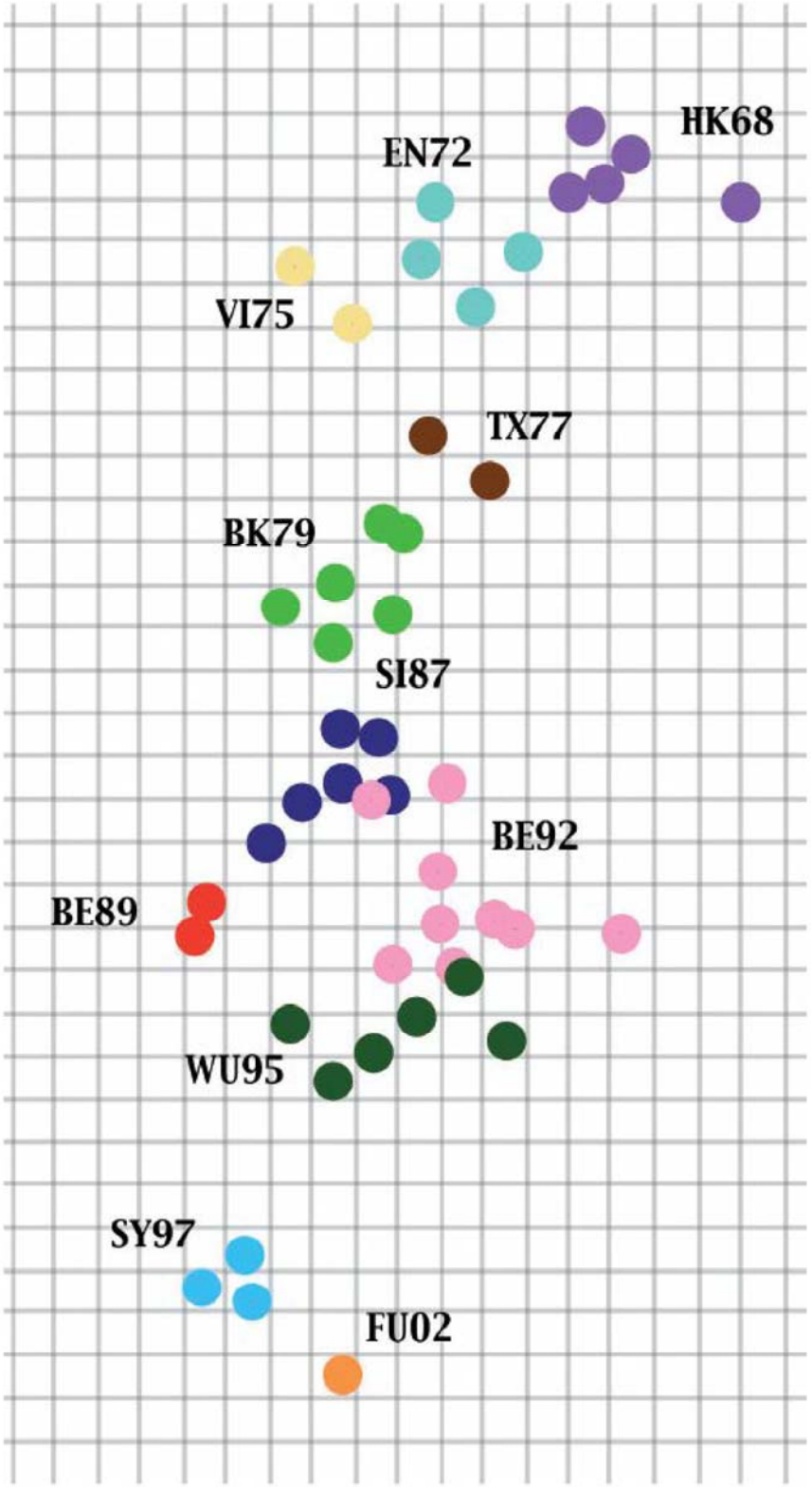


- We combine five datasets, obtained at different times, to form the HI table.



- We use sequence data to generate antigenic map. (253 sequences extracted from the Supplementary Data of Smith et al.)
- This antigenic map is highly consistent with the Smith's map, which shows the robustness of our method



# Summary

- ATIVS is a java-based web server built on Linux.
- Both serology data of all influenza viruses and HA1 sequence data of human influenza A/H3N2 viruses can be utilized to generate antigenic maps
- Useful in influenza virus surveillance and vaccine strain selection.

Yu-Chieh Liao, Chin-Yu Ko, Min-Hsin Tsai, Min-Shi Lee and Chao A. Hsiung, ATIVS: Analytical tool for influenza virus surveillance, Nucleic Acids Research, **37** (2009), W643-W646.

# Acknowledgements


- Chin-Yu Ko
- Ming-Hsin Tsai

- Dr. Min-Shi Lee
- Dr. Chao A. Hsiung

http://influenza.nhri.org.tw