- Project presentations and papers due:
  Thursday, MARCH 15, 2007,
  1030-1220

In Genome Sciences Foege S-110

Evolution of influenza

Today:

1. Global health impact of flu
   - why should we care?
2. What is influenza?
   - what are the components of the virus and how do they change?
3. Where does influenza come from?
   - are there animal reservoirs?
4. Predicting the future of influenza
   - can we predict strains to make vaccines?

Global impact of flu

- **Flu** is a highly contagious respiratory illness which infects millions of people every year and kills hundreds of thousands
- Caused by influenza viruses (A, B, C)
- Estimated to infect 100 million people each year in the northern hemisphere alone
- Huge impacts on morbidity and mortality, but also economic impacts

Global impact of flu

- Typical patterns is a regional **epidemic** every year beginning in late winter and lasting a couple of months
- These yearly epidemics are caused by established influenza A lineages (H1N1, H3N2) that continued to circulate after being introduced to the human population
- Every 10-50 years sees a **pandemic** variant that sweeps across the globe, infecting a large proportion of humans

Global impact of flu

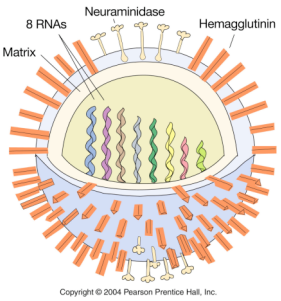
- Pandemics occurred in 1890, 1918, 1957, and 1968. The 1918 **Spanish flu** epidemic probably infected about 50% of the human population and represents the most intense culling of humans.
- It is very likely that pandemic influenza will return
- Evolutionary tools can help fight currently circulating influenza, and possibly dampen the effects of future pandemic strains
- **Antigenic drift** versus **antigenic shift**

What is influenza?

- There are 3 main types of influenza virus: A, B, and C
- We'll concentrate on influenza A, the most important from the human standpoint
- Negative-stranded RNA viruses with segmented genome
- 8 RNA segments encoding around 10 proteins

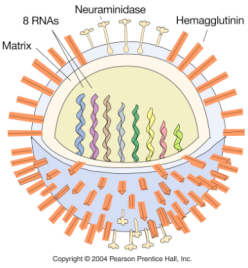
What is influenza?

- 2 glycosylated proteins on the surface, HA (hemagglutinin) and NA (neuraminidase)
- HA and NA are involved in **virus attachment and release from hosts cells**
- They are the primary targets of the immune system in humans (and swine)
- Different strains of influenza are typically named for their HA and NA genes, eg. "H1N1"



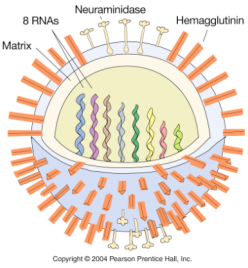
What is influenza?

- The virus is capable of generating a lot of genetic variability
- First, like other RNA viruses, the lack of proofreading and high error rate of the viral polymerase leads to high **mutation rate.**
- This high mutation rate, in turn, leads to a high **substitution rate**.
- (**Substitutions** are mutations that have become fixed through **genetic drift** or **natural selection**)
- When these substitutions occur in antigenic epitopes, they can lead to escape mutants (**antigenic drift**)



What is influenza?

- The segmented nature of the influenza genome leads to another, more dramatic source of variability
- **Reassortment** can occur when one host is co-infected with two different strains, and the progeny viruses get some gene segments from one "parent" and some from another
- For example, if you were infected simultaneously with both H1N1 and H3N2, you might generate an H1N2 virus that could infect someone else and start a "new" epidemic



INFLUENZA VIRUS

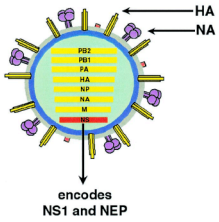
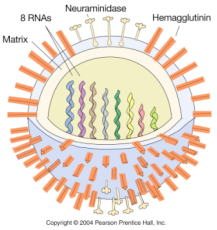


Fig. 1. Diagram of an influenza-virus particle. The surface of each influenza virion consists of a lipid envelope in which two major viral surface antigens, the hemagglutinin (HA) and the neuraminidase (NA), are found. Within the particle are the eight negative-sense viral RNA segments encoding the viral proteins. The smallest viral segment, the NS segment, encodes two proteins: the NS1, an antagonist of the cellular type I interferon system, and the nuclear export protein (NEP), which functions in viral assembly.

Reassortment possible since HA and NA are encoded on different RNA segments.

Where does flu come from?

- Reassortment gets particularly bad when HA and/or NA genes that are new to the human population are introduced
- There are 15 HA subtypes in the gene pool of influenza that infects wild birds (H1-H15)
- Birds are a **reservoir** of human influenza, the source from which new viruses may periodically emerge
- Importation of a variant to which few or no humans have prior immunity (**antigenic shift**) is the cause of the periodic pandemics



Where does flu come from?

- Since pigs can be infected with both avian and human influenza, and various reassortants have been recovered from pigs, it has been suggested that pigs might play the role of intermediary in the generation of reassortant pandemic strains
- In 1979, for example, an avian influenza A began infecting swine in Northern Europe. This lineage has since clearly mixed with locally circulating human lineages, and has picked up human H and N2 HA and NA segment via reassortment

Where does flu come from?

- 1997, it became clear that avian influenza could also jump directly from birds into humans
- The Hong Kong 1997 variant was an avian H5N1 virus that infected 18 people and killed 6
- Luckily, the virus was poorly transmissible in humans (if at all)
- What would happen if someone got infected with avian H5N1 from their chicken, and also human H1N1 from their co-worker?

1918 Flu

- 1918 Spanish flu probably infected about 1 billion of the world's 1.8 billion people, and led to the death of perhaps 50 million (>30% US population)
- Two waves, first not too deadly. Most deaths occurred in a second 8-week window, October-November 1918
- Most deaths due to complications like pneumonia, dehydration
- Unusual pattern of mortality, with healthy young adults, 20-25, most affected.

How would you determine origin of 1918 flu strain?

-Recent adaptation human strain?
- Avain origin?
-Swine origin?

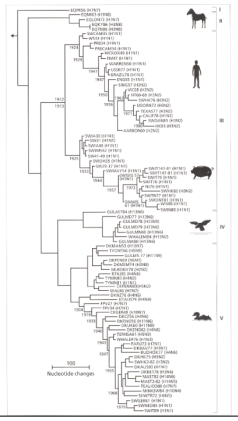
Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 1651–1656, February 1999
Microbiology

Origin and evolution of the 1918 "Spanish" influenza virus hemagglutinin gene

ANN H. REID*, THOMAS G. FANNING, JOHAN V. HULTIN, AND JEFFERY K. TAUBENBERGER

Phylogenetic analysis to determine origin of 1918 flu - was it avian?

- Use fixed tissue from 1918 to PCR amplify flu strain
- Sequence genes and place in known phylogeny of influenza from different taxa
  - Problem with not have complete data for all taxa

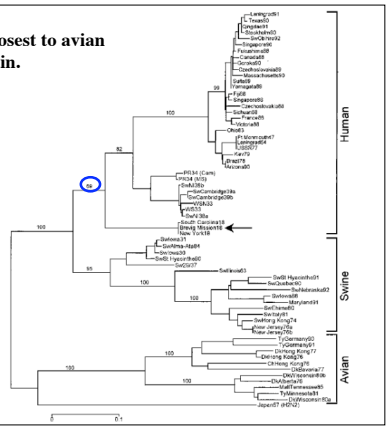


Note diversity in avain.

Now place in 1918 influenza

- Use distance NJ method
  - What about parsimony and ML?
- Bootstrap tree (100 replicates)
  - Need more replicates

1918 flu group is closest to avian than any other strain.



Conclusions from tree

- 1918 strain clusters within mammalian lineage
- Closest mammalian strain to avian
  - Low bootstrap support
- Could have jumped avian, then resided in mammalian before becoming highly virulent.

Other evidence of avian origin:

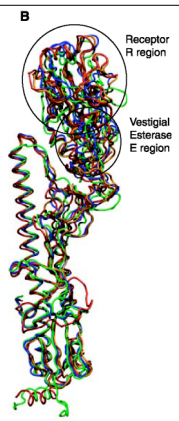
Structure of the Uncleaved Human H1 Hemagglutinin from the Extinct 1918 Influenza Virus

James Stevens,[1] Adam L. Corper,[1] Christopher F. Basler,[3] Jeffery K. Taubenberger,[4] Peter Palese,[3] Ian A. Wilson[1,2]*

19 MARCH 2004 VOL 303 SCIENCE

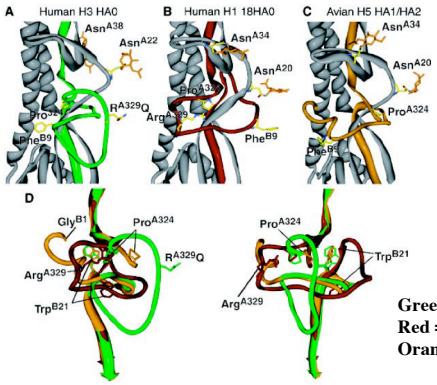
Structural studies suggest avian origin

- Purify/express HA from 1918 virus
- Crystalize structures and solve
  - Problem with structures crystalized in different solutions. Structures not static images as we view them.
- Compare structures between human H1 (1918 strain), H5 and avian.
- Avian strain has unique fold.
- 3D structures more conserved than primary structure?



Comparison of HA from 1918 (red), Human H5 (green) and avian (blue)

Detail of region characteristic of avian HA



Green = human H5
Red = 1918
Orange = Avian

Summary from structural studies

- 1918 contains structures unique to avian influenza
- Similar to human influenza strains, worked out details of how it evolved to bind human.

Where did 1918 influenza come from

- Phylogenetics points to mammalian
- Structural biology suggests avian
- May have jumped from avian into mammalian host, adapted then caused epidemic
  - Young were hardest hit in 1918, could older adults had antibodies from prior less pathogenic strain?

Newest results:
Tumpey TM, et al. A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission. Science. 2007 Feb 2;315(5812):655-9.

- Took 1918 flu strain used site directed mutagenesis to mutate 2 residues back to avian form (**involved in sialic acid binding**)
- Infected ferrets
  - Ferrets became infected and sickened, but flu did not transmit
  - Suggests strain adapted from avian to human
- Could the sites have been predicted?
- Concerns regarding reconstructing pandemic flu virus?

How to design a vaccine against influenza?

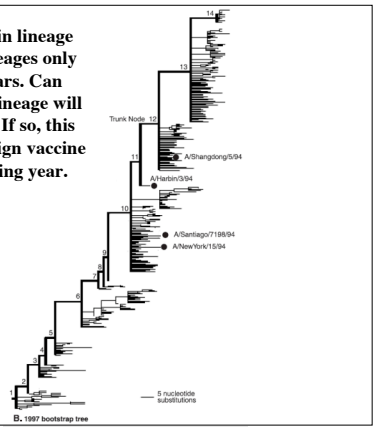
How to design a vaccine against influenza?

Predicting the Evolution of Human Influenza A

Robin M. Bush,[1]* Catherine A. Bender,[2] Kanta Subbarao,[2] Nancy J. Cox,[2] Walter M. Fitch[1]

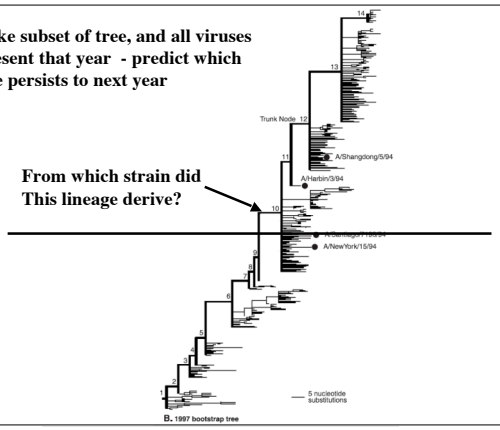
Eighteen codons in the HA1 domain of the hemagglutinin genes of human influenza A subtype H3 appear to be under positive selection to change the amino acid they encode. Retrospective tests show that viral lineages undergoing the greatest number of mutations in the positively selected codons were the progenitors of future H3 lineages in 9 of 11 recent influenza seasons. Codons under positive selection were associated with antibody combining site A or B or the sialic acid receptor binding site. However, not all codons in these sites had predictive value. Monitoring new H3 isolates for additional changes in positively selected codons might help identify the most fit extant viral strains that arise during antigenic drift.

Influenza tree has main lineage termed trunk. Tip lineages only last on average 1.5 years. Can we predict which tip lineage will continue along trunk. If so, this is a good strain to design vaccine against for the following year.

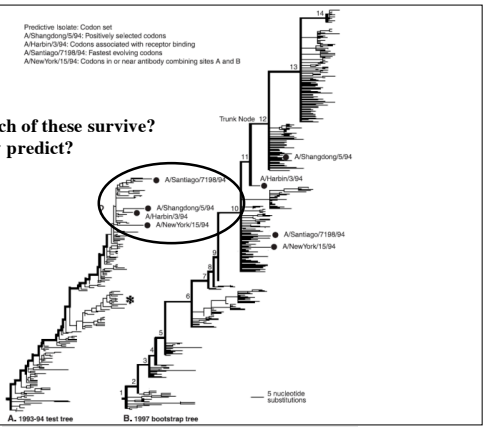


Take subset of tree, and all viruses present that year - predict which one persists to next year

From which strain did This lineage derive?



Which of these survive? How predict?



How to design a vaccine against influenza

- Predict sites under positive selection
- Look at lineages with most changes in these sites

Determining sites under positive selection

- Calculate probability of observing 8 nonsynonymous changes in 10 substitutions
  - Need to know proportion of nonsynonymous sites (p) and synonymous sites (q)

What is probability of getting exactly eight non-silent mutations in examining ten mutations?

$$P = a!p^n q^s / n!s!$$

This is same probability theory as with coin toss

$$P = 10!\,(0.45)^8\,(0.55)^2/(8!\,2!) = 0.023$$

The eighteen positions of H3 hemagglutinins from human influenza viruses under positive selection on the off-tip branches of the tree

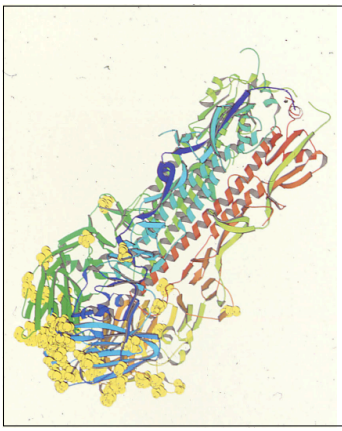
Position	NS/S	Probability		Position	NS/S	Probability	
121	5/0	0.0135		186	9/1	0.0025	t
124	5/0	0.0135		190	4/0	0.0320	t
133	8/0	0.0010		193	4/0	0.0320	t
135	5/0	0.0135	t	194	4/0	0.0320	t
138	6/0	0.0057	t	197	4/0	0.0320	
142	4/0	0.0320		201	4/0	0.0320	
145	8/0	0.0010	t	226	20/1	< 0.0001	t
156	9/1	0.0025	t	262	4/0	0.0320	
158	5/0	0.0135		275	6/0	0.0057	

Position is the amino acid position in the protein. NS/S is the number of non-silent over the number of silent substitutions. Probability is the binomial probability that NS/S would be observed by chance. The letter t indicates that the position also had a significant excess of non-silent changes in the tip branches.

Table 1. The 18 positively selected codons in the HA1 hemagglutinin gene and their membership in alternative codon sets. R, codons associated with the sialic acid receptor binding site; A or B, codons in or near antibody combining site A or B, respectively; F, codons with rapid rates of amino acid replacement.

Codon	Codon set	
121	—	— F
124	—	A —
133	—	A F
135	R	A F
138	R	A F
142	—	A —
145	—	A F
156	—	B F
158	—	B —
186	—	B F
190	R	B F
193	—	B F
194	R	B F
197	—	B —
201	—	— —
226	R	— F
262	—	— —
275	—	— F

Positively selected codons fall in several functional regions



Positively selected sites fall in receptor region

Subset of codons used for prediction

- Positively selected
- Antigen Binding site A and B (AB)
- Antigen Binding site C,D, E (CDE)
- Sialic acid receptor binding site (RBS)
- Fast - undergoing largest replacements
- Random

Basics of test

- Make tree with all extant lineages
- Group codons based upon predetermined codon sets.
- Count number of NS changes for each codon set.
- Assign the most likely to persist as that strain with most NS changes.
  - Those with most changes likely to escape immune response.

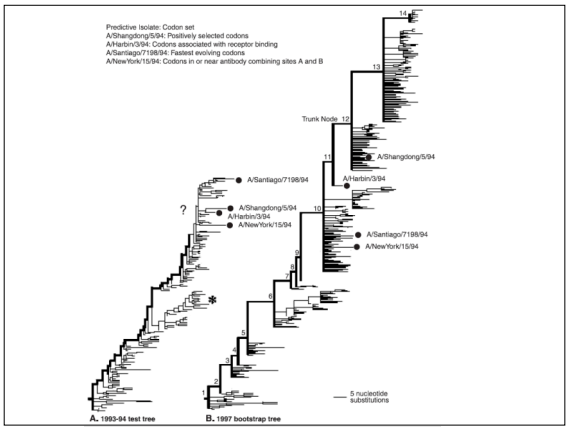


Table 2. Sample sizes for retrospective test data sets, and comparison of prediction results with random expectations.

	Influenza season										
	86–87	87–88	88–89	89–90	90–91	91–92	92–93	93–94	94–95	95–96	96–97	Mean
Number of isolates per test tree*	39	50	61	65	76	97	141	173	222	288	357	
Probability of success using randomly chosen isolates†	15.4	30.0	42.6	1.5	1.3	8.3	31.9	0.6	1.8	16.0	3.9	13.9
Probability of success using randomly chosen codons‡	4.1	20.6	41.8	0.8	0.4	14.4	31.0	0.2	0	1.4	1.8	10.6

*Cumulative number of isolates used to construct retrospective test trees representing the evolution of the hemagglutinin gene from 1983 through 11 consecutive influenza seasons. †Percent of randomly chosen predictive isolates that descend from trunk nodes as far or farther up the trunk of the 1997 bootstrap tree (Fig. 1B) as the predictive isolate found by counting replacements at the 18 positively selected condons. ‡Percent of 1000 randomly chosen sets of 18 codons that produced predictive isolates that descend from trunk nodes as far or farther up the trunk of the 1997 bootstrap tree (Fig. 1B) as the predictive isolate found by counting changes at the 18 positively selected codons.

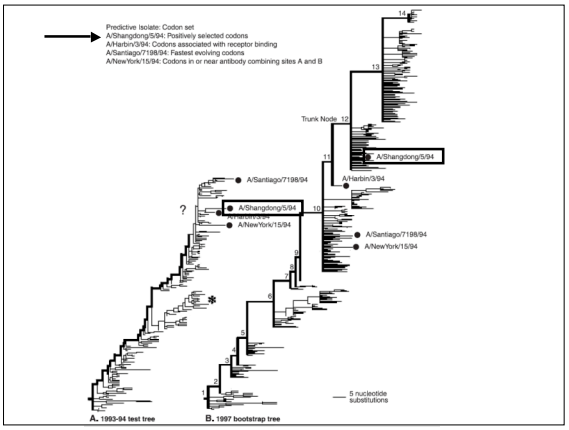
Table 3. Results of retrospective prediction tests for 11 recent influenza seasons using codons under positive selection and seven alternative codon sets. Cells indicate the trunk nodes on the 1997 bootstrap tree (Fig. 1B) from which the predictive isolates resulting from each test descended. Successful tests (bold underlined numbers) are those in which the predictive isolate descended from the uppermost possible node on the 1997 tree (bottom row). Right-hand column shows the total number of seasons in which each codon set produced a successful test. Cell entries with a decimal place are mean trunk nodes (across the 10 replicate test trees) when these tests produced predictive isolates descending from different trunk nodes. Positively selected, the set of 18 codons under positive selection; AB, codons in or near antibody combining sites A and B; CDE, codons in or near antibody combining sites C, D, and E; RBS, codons associated with the sialic acid receptor binding site; Fast, codons undergoing the largest number of amino acid replacements.

	Codons (*n*)	Influenza season										Success (*n*)
		86–87	87–88	88–89	89–90	90–91	91–92	92–93	93–94	94–95	95–96	96–97	
Positively selected	18	**5**	5	5	**8**	**8**	**10**	**10**	**12**	**13**	**13**	**14**	9
AB	41	**5**	5	**6**	**8**	7.0	**10**	**10**	10	12.0*	**13**	**14**	7
AB but not under positive selection	28	**5**	5	5.5	5.5	5.5	8.7	9.6	10	12	12	12	1
CDE	90	1	1	**6**	6	6	8	**10**	10	10	**13**	13	3
RBS	16	3.7	5	**6**	6	6	**10**	**10**	11	12.5	**13**	13.1*	4
RBS but not under positive selection	11	4.1	5	5	5	5	5	5	5	5	5	5	0
Fast	20	3	3	**6**	6.7*	7.0	**10**	**10**	10	**13**	**13**	**14**	6
Fast but not under positive selection	18	3	3	**6**	6	6	6	**10**	10	10	12	13	2
Top possible trunk node		5	6	6	8	8	10	10	12	13	13	14	

*Prediction tests that failed to identify the same predictive isolate in 7 or more of the 10 replicate tests. For these tests the cell contains the mean trunk node taken across results from all 10 replicate tests.

Success with + selected sites = 82%

Functional subsets not as predictive.



Summary

- Influenza phylogenetic tree has single main lineage
- Lineages with most changes at positively selected sites are most likely to predict evolution of influenza.
- No casual explanation for adaptive evolution
  - Sites with known function not predictive.

Could method be improved?

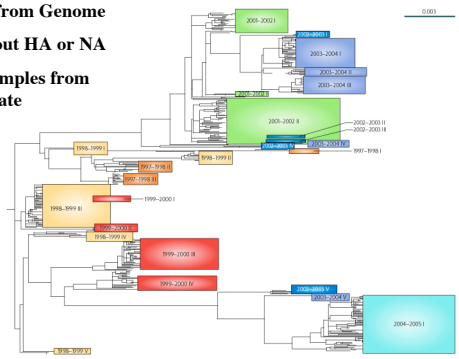
- Use maximum likelihood to predict sites under positive selection.

This study used just HA gene. What about phylogenies using whole genome?

Tree from Genome

Without HA or NA

All samples from NY state



Questions

- What are two ways the 1918 flu was shown to be of avian origin.
- What is antigenic drift and antigenic shift
- How does reassortment work in influenza virus?
- Describe one way to predict the evolution of human influenza.
- How would 3D structures help determine relatedness of 1918 strain and avian influenza?
- What is difference between epidemic and pandemic?