

# Comparative Genomics Analysis to Identify Genetic Determinants of Influenza Virus Human Adaptation

Richard H. Scheuermann, Ph.D.  
Director of Informatics  
JCVI

# Outline

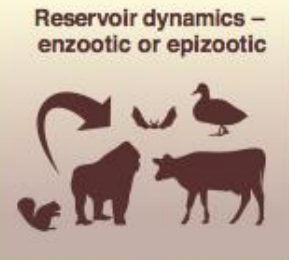
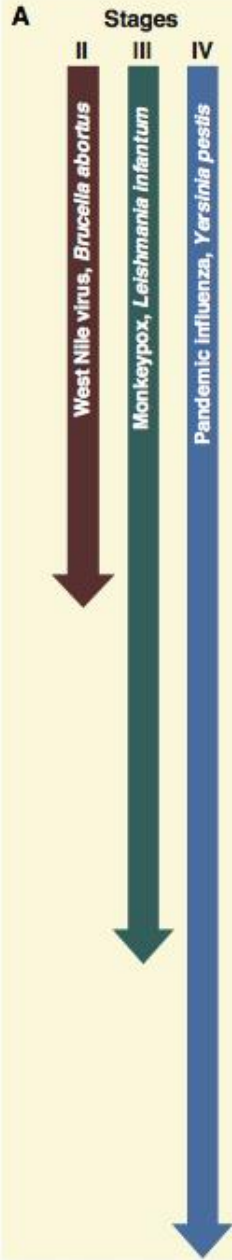
- Comparative genomics of H5N1 human adaptation
- Hands on workshop
  - NIAID Bioinformatics Resource Centers
  - Virus Pathogen Resource (ViPR)/Influenza Research Database (IRD) overview
  - Human adaptation of 2013 avian H7N9
  - Sequence annotation/submission/recombination
  - PlasmODB example
- T-shirt contest

# Zoonosis Summary

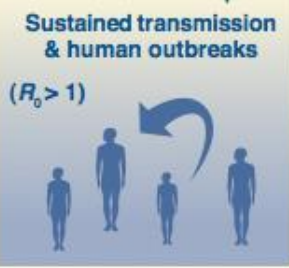
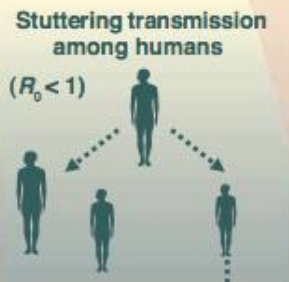
- A zoonosis is an infectious disease that is transmitted between species (sometimes by a vector) from animals other than humans to humans or from humans to other animals.
- Of the 1415 recognized species of human pathogens, 61% are of zoonotic origin [Taylor 2001].
- These include Hendra, Nipah, Machupo, Ebola, Influenza A, SARS-CoV, *Yersinia Pestis*, *Borrelia burgdorferi*, *Plasmodium knowlesi*.
- Use of comparative genomics to understand zoonotic spillover – what are the genetic determinants that allow an animal virus to adapt to human

# Flu pandemics of the 20<sup>th</sup> and 21<sup>st</sup> centuries initiated by species jump events

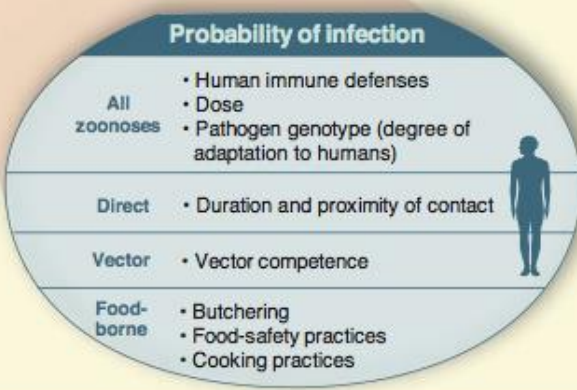
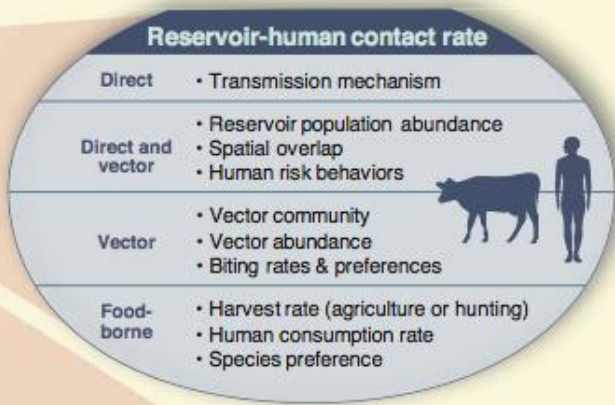
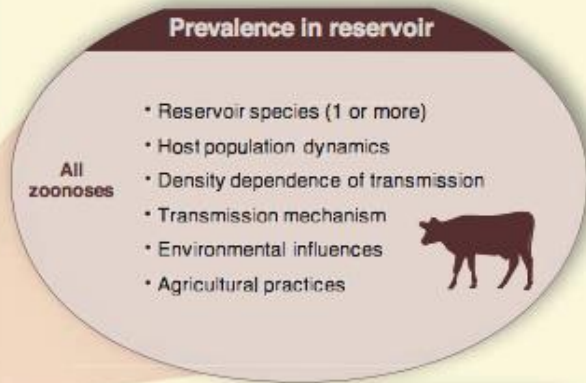
- 1918 flu pandemic (Spanish flu)
  - subtype H1N1 (avian origin)
  - estimated to have claimed between 2.5% to 5.0% of the world's population (20 - 100 million deaths)
- Asian flu (1957 – 1958)
  - subtype H2N2 (avian origin)
  - 1 - 1.5 million deaths
- Hong Kong flu (1968 – 1969)
  - subtype H3N2 (avian origin)
  - between 750,000 and 1 million deaths
- 2009 H1N1
  - subtype H1N1 (swine origin)
  - ~ 16,000 deaths as of March 2010



**Spillover transmission across species**



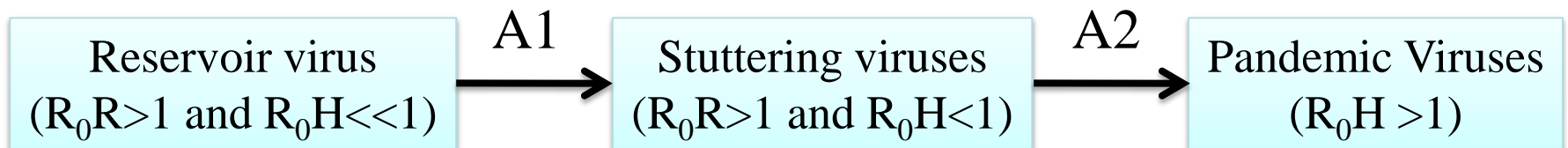
**B**



← Adaptive drivers

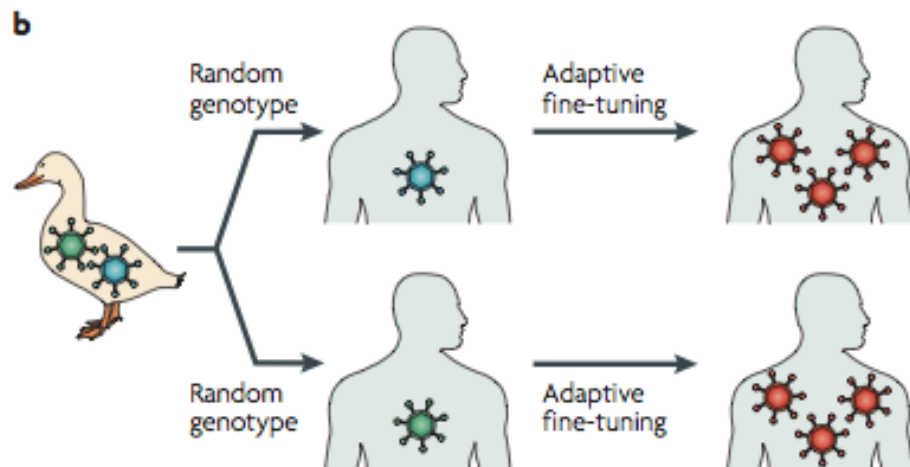
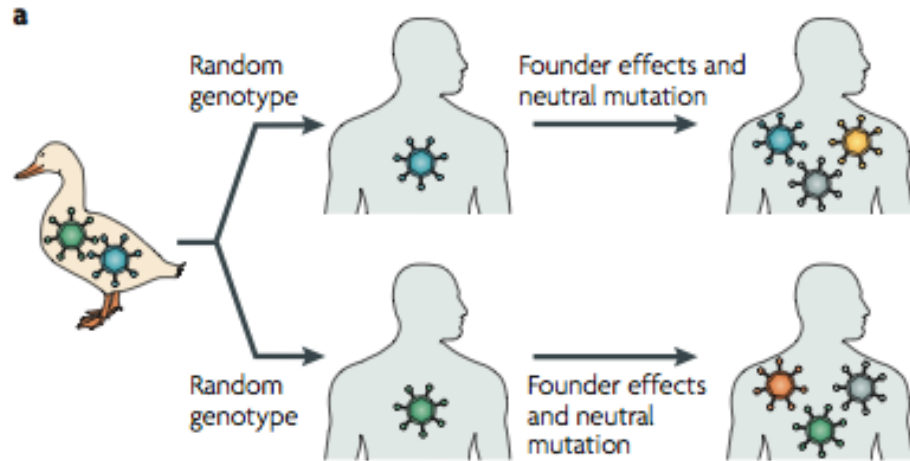
# Basic reproductive number ( $R_0$ )

- Total number of secondary cases per case
- Reasonable surrogate of fitness
- Characteristics of pandemic viruses:
  - $R_0H > 1$ , and
  - In genetic neighborhood of viruses with  $R_0R > 1$  and  $R_0H < 1$
- Adaptive drivers

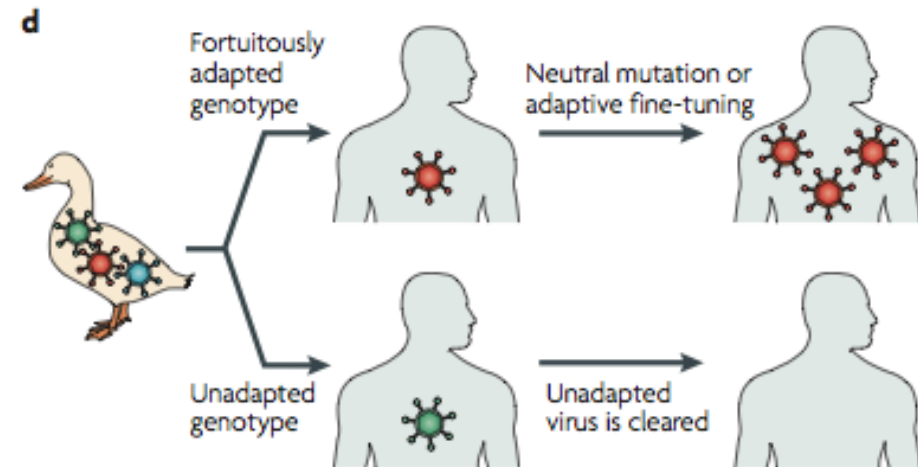
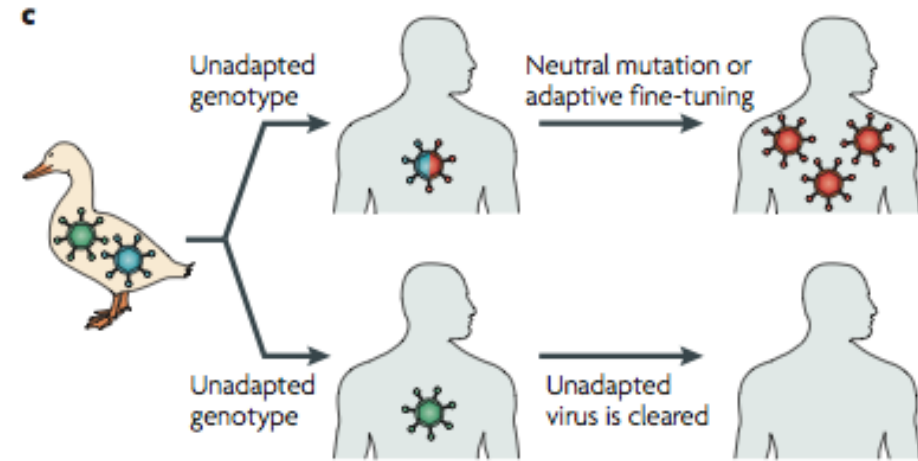


# Adaptive drivers

**Ecological driver  
(no selection required)**



**Adaptive driver  
(selection required)**



# Stuttering transmission and adaptive drivers

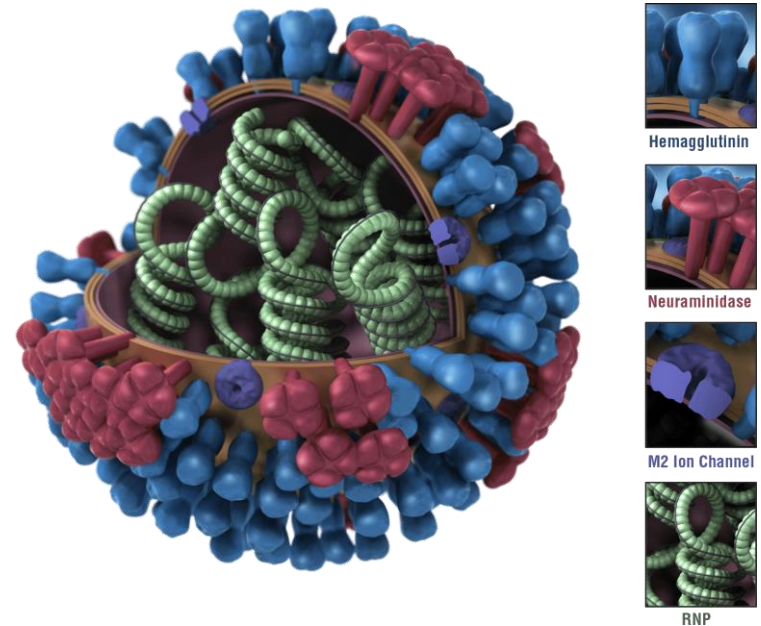
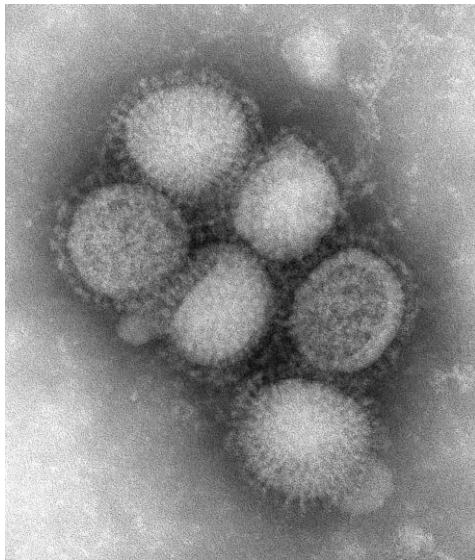
- Spillover and stuttering transmission can reveal adaptive drivers when combined with evidence of convergent evolution
  - Odds of finding the same neutral mutation by chance in multiple species jumps is low
  - Therefore, finding same mutation in multiple phylogenetically independent species jump events (convergent evolution) is strong evidence for adaptive driver
- Requires a statistical test of abundance difference combined with a test of phylogenetic independence that controls for sampling bias and founder effects



# Influenza Virus

Orthomyxoviridae family  
Negative-strand RNA  
Enveloped  
8 RNA segments encode 12 proteins  
Classified based on HA/NA serology

Zoonotic virus – broad host range  
Primary reservoir in aquatic birds  
Transmission between taxonomies is rare  
Different strains show different virulence



# H5N1 Overview

- H5N1 influenza has historically been limited to avian species
- In 1997, a lethal case of H5N1 infection was reported in a 3-year old boy in Hong Kong, cause by a highly pathogenic H5N1 avian influenza virus
- In spite of attempt to eradicate the virus in poultry markets, the lineage spread through wild aquatic birds and domestic poultry; however, no new human cases were reported after 1997
- In 2003, a new wave of human cases resulting from avian H5N1 zoonosis was reported, with increased disease severity and death
- Since 2003, sporadic H5N1 human cases continue to be reported (stuttering transmission), suggesting that the virus is capable of evading the genetic barrier to human transmission

# Underlying postulates

- Prior to their repeated zoonosis starting in 1997, H5N1 viruses circulating in the avian reservoir were maladapted for infection of humans
- H5N1 viruses isolated from humans reflect a natural selection for viruses that have acquired a subset of adaptive changes that allow them to infect humans with reasonable efficiency
- Thus, a comparison of genome sequences between human and avian H5N1 viruses would reveal candidate adaptive driver sequence variations
- The identification of the same sequence variations in independent virus lineages would be evidence of convergent evolution

# Data mining workflow

- Extract sequences for every IAV genome segment from avian H5N1 viruses isolated from Southeast Asia (avian H5N1) from the Influenza Research Database (IRD; [www.fludb.org](http://www.fludb.org)).
- Extract sequences for every IAV genome segment from human H5N1 viruses isolated from Southeast Asia (human H5N1) from the IRD.
- Align all sequences for each IAV protein and identify those amino acid positions that have significantly different residues between avian H5N1 and human H5N1 proteins and are more prevalent in human isolates as initial candidates.
- Select only those variants that are relatively rare in avian reservoir (<25%).
- Determine which of these initial candidates are found in multiple phylogenetic lineages in nucleotide segment sequence trees.
- Select those significant amino acid residues that are found in independent clades as candidates for convergent evolution, or variants generated by different codon substitutions within the same clade
- Perform analysis on all proteins/segments

# Strain Search – PB2 avian H5N1 Southeast Asia up to 2003

## Protein Sequence Search <sup>?</sup>

Search for influenza sequences, proteins, and strains using two types of searches. Use the advanced search to allow you to refine your search with the more fine grained search, and you can pick your viewing options.

### DATA TO RETURN

- Segment / Nucleotide  
 Protein  
 Strain

### VIRUS TYPE

- A  
 B  
 C

### SUB TYPE

H5N1

\* Use comma to separate multiple entries. Ex: H1N1, H7, H3N2.

### STRAIN NAME

\* Use comma to separate multiple entries.

Ex: A/chicken/Israel/1055/2008,  
A/chicken/Laos/16/2008.

### SELECT PROTEINS

- All  
1 PB2  
2 PB1  
2 PB1-F2  
3 PA  
4 HA  
5 NP  
6 NA  
7 M1  
7 M2  
8 NS1  
8 NS2

- Complete Genome Only  
 Complete CDS Only  
 Only include data from 2009 swine origin viruses (SOP)  
 Exclude data from 2009 swine origin viruses

### DATE RANGE

From: YYYY To: 2003

### HOST

- All  
Avian  
Blow Fly  
Camel  
Cheetah  
Civet

### AVIAN

- All  
African Starling  
American Black Duck  
American White Pelican  
American Wigeon  
Aquatic Bird  
Arctic Tern

### GEOGRAPHIC GROUPING

- All  
Africa  
Asia  
Europe  
North America  
Oceania

### COUNTRY

- Honduras  
Hong Kong  
Hungary  
Iceland  
India  
Indonesia  
Iran

▶ [ADVANCED OPTIONS](#) Show All

Tip: To select multiple or deselect, Ctrl-click (Windows) or Cmd-click (MacOS)

Clear

Search

# 260 PB2 records

## Protein Sequence Search Results

Your search returned 260 proteins.

[Search Criteria](#)

[Display Settings](#)

Displaying 50 per page

Your Selected Items: 0 items selected

[Add to Working Set](#)

[Save Search](#)

[Run Analysis](#) ▼

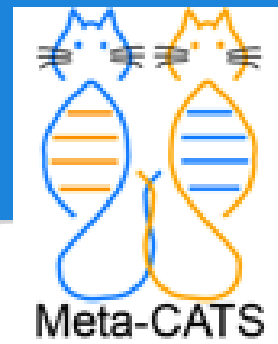
[Download](#)

Select all 260 proteins

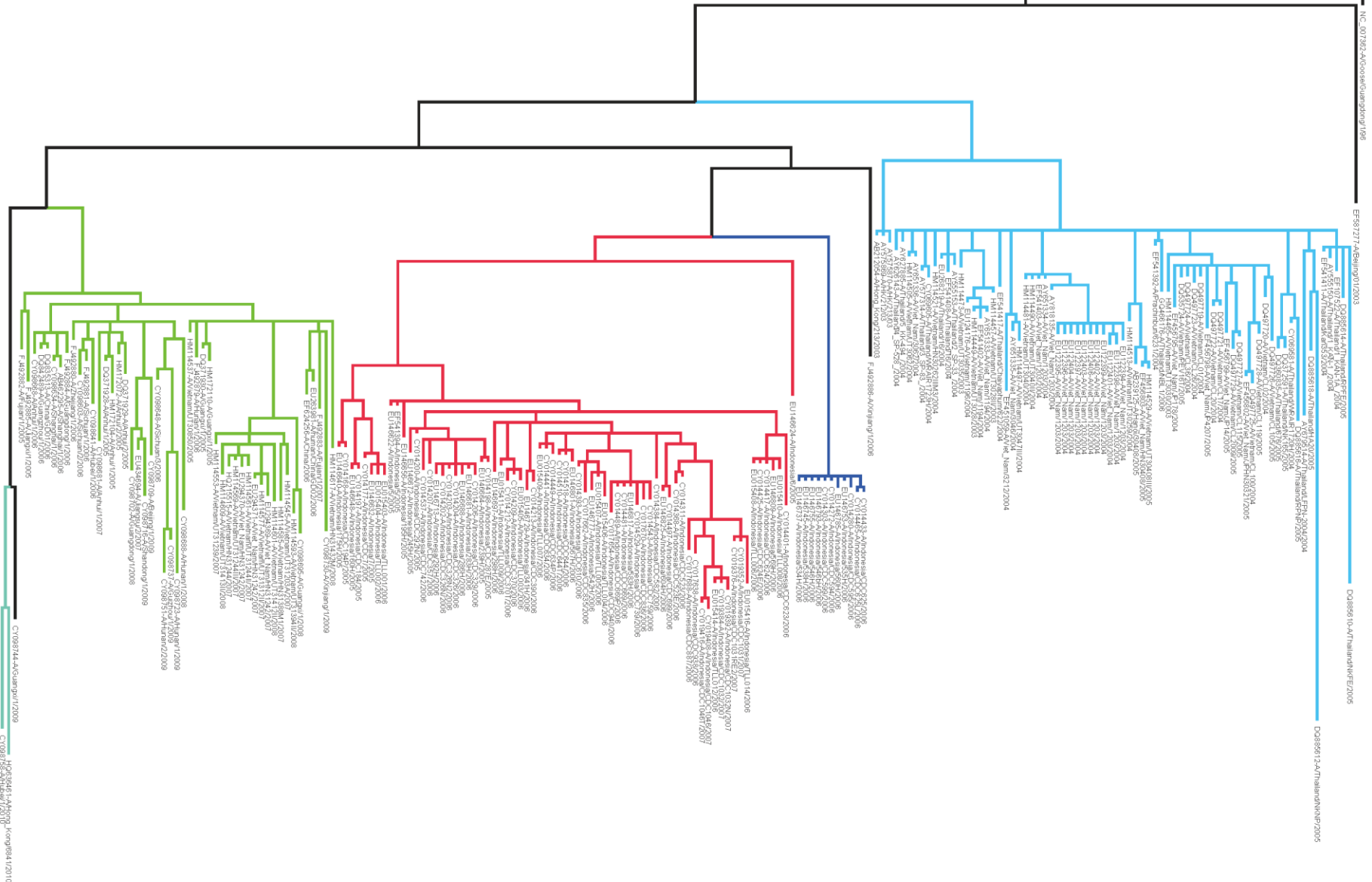
1 2 3 4 5 6 Next > Page:  of 6

<input type="checkbox"/>	<a href="#">i</a>	Name	Sequence Accession	Complete Genome	Segment	Segment Length	Subtype	Date	Host Species	Country	Flu Season	Strain Name
<input type="checkbox"/>	<a href="#">i</a>	PB2	AY651743	No	1	1292	H5N1	2003	Black-Headed Gull/Avian	Hong Kong	U	*A/Black Headed Gull/HK/12.1/2003(H5N1)
<input type="checkbox"/>	<a href="#">i</a>	PB2	GU052126	No	1	2303	H5N1	1997	Chicken/Avian	China	NT	A/chicken/China/27402/1997
<input type="checkbox"/>	<a href="#">i</a>	PB2	GQ122430	Yes	1	2280	H5N1	11/15/2003	Chicken/Avian	Indonesia	U	*A/chicken/East Java/UT1006/2003(H5N1)
<input type="checkbox"/>	<a href="#">i</a>	PB2	DQ351872	Yes	1	2280	H5N1	2002	Chicken/Avian	China	NT	*A/chicken/Hebei/108/02(H5N1)
<input type="checkbox"/>	<a href="#">i</a>	PB2	DQ351870	Yes	1	2280	H5N1	2001	Chicken/Avian	China	NT	*A/chicken/Hebei/718/2001(H5N1)
<input type="checkbox"/>	<a href="#">i</a>	PB2	AY651739	No	1	1449	H5N1	2003	Chicken/Avian	Hong Kong	U	A/chicken/HK/2133.1/2003
<input type="checkbox"/>	<a href="#">i</a>	PB2	AY576386	No	1	2274	H5N1	2002	Chicken/Avian	Hong Kong	U	A/chicken/HK/31.4/02
<input type="checkbox"/>	<a href="#">i</a>	PB2	AY651736	No	1	1340	H5N1	2002	Chicken/Avian	Hong Kong	U	A/chicken/HK/3169.1/2002
<input type="checkbox"/>	<a href="#">i</a>	PB2	AY651735	No	1	1226	H5N1	2002	Chicken/Avian	Hong Kong	U	A/chicken/HK/3176.3/2002
<input type="checkbox"/>	<a href="#">i</a>	PB2	AY651732	No	1	2271	H5N1	2002	Chicken/Avian	Hong Kong	U	A/chicken/HK/37.4/2002
<input type="checkbox"/>	<a href="#">i</a>	PB2	AY576390	No	1	1293	H5N1	2002	Chicken/Avian	Hong Kong	U	A/chicken/HK/409.1/02
<input type="checkbox"/>	<a href="#">i</a>	PB2	AY576387	No	1	1287	H5N1	2002	Chicken/Avian	Hong Kong	U	A/chicken/HK/61.9/02

# Meta-CATS



- Statistical algorithm written in the R programming language
- Algorithm steps
  - 1) identify and segregate sequence records of interest based on associated metadata
  - 2) performing a multiple sequence alignment using the Multiple Sequence Comparison by Log-Expectation (MUSCLE) and UCLUST algorithms
  - 3) performing the chi-square test of independence and Pearson's chi-square test in tandem to calculate a p-value
  - 4) compare with other information about sequence feature structure/function
- Publication
  - Brett E. Pickett, et al. (2013) "Metadata-driven Comparative Analysis Tool for Sequences (meta-CATS): an Automated Process for Identifying Significant Sequence Variations Dependent on Differences in Viral Metadata." *Virology* (in press).
- Comparative analysis of avian versus human H5N1 for all 10 proteins
  - 126 candidates
  - Next step – multiple phylogenetic lineages (convergent evolution)



IC\_07302-AcebesaQuampur/1208

EF83771-ABHijp/1203

CD88614-4-AThailandKRF/E/2005

EF10752-2-AThailand KRF/12/2005

EF41431-1-AThailandKRF/2004

CD88612-2-AThailandKRF/2005

CD88610-1-AThailandKRF/2005

CD88591-10-AThailandKRF/2005

CD88591-13-AThailandKRF/2005

CD88576-1-AThailandKRF/2005

CD88575-1-AThailandKRF/2005

CD88574-1-AThailandKRF/2005

CD88573-1-AThailandKRF/2005

CD88572-1-AThailandKRF/2005

CD88571-1-AThailandKRF/2005

CD88570-1-AThailandKRF/2005

CD88569-1-AThailandKRF/2005

CD88568-1-AThailandKRF/2005

CD88567-1-AThailandKRF/2005

CD88566-1-AThailandKRF/2005

CD88565-1-AThailandKRF/2005

CD88564-1-AThailandKRF/2005

CD88563-1-AThailandKRF/2005

CD88562-1-AThailandKRF/2005

CD88561-1-AThailandKRF/2005

CD88560-1-AThailandKRF/2005

CD88559-1-AThailandKRF/2005

CD88558-1-AThailandKRF/2005

CD88557-1-AThailandKRF/2005

CD88556-1-AThailandKRF/2005

CD88555-1-AThailandKRF/2005

CD88554-1-AThailandKRF/2005

CD88553-1-AThailandKRF/2005

CD88552-1-AThailandKRF/2005

CD88551-1-AThailandKRF/2005

CD88550-1-AThailandKRF/2005

CD88549-1-AThailandKRF/2005

CD88548-1-AThailandKRF/2005

CD88547-1-AThailandKRF/2005

CD88546-1-AThailandKRF/2005

CD88545-1-AThailandKRF/2005

CD88544-1-AThailandKRF/2005

CD88543-1-AThailandKRF/2005

CD88542-1-AThailandKRF/2005

CD88541-1-AThailandKRF/2005

CD88540-1-AThailandKRF/2005

CD88539-1-AThailandKRF/2005

CD88538-1-AThailandKRF/2005

CD88537-1-AThailandKRF/2005

CD88536-1-AThailandKRF/2005

CD88535-1-AThailandKRF/2005

CD88534-1-AThailandKRF/2005

CD88533-1-AThailandKRF/2005

CD88532-1-AThailandKRF/2005

CD88531-1-AThailandKRF/2005

CD88530-1-AThailandKRF/2005

CD88529-1-AThailandKRF/2005

CD88528-1-AThailandKRF/2005

CD88527-1-AThailandKRF/2005

CD88526-1-AThailandKRF/2005

CD88525-1-AThailandKRF/2005

CD88524-1-AThailandKRF/2005

CD88523-1-AThailandKRF/2005

CD88522-1-AThailandKRF/2005

CD88521-1-AThailandKRF/2005

CD88520-1-AThailandKRF/2005

CD88519-1-AThailandKRF/2005

CD88518-1-AThailandKRF/2005

CD88517-1-AThailandKRF/2005

CD88516-1-AThailandKRF/2005

CD88515-1-AThailandKRF/2005

CD88514-1-AThailandKRF/2005

CD88513-1-AThailandKRF/2005

CD88512-1-AThailandKRF/2005

CD88511-1-AThailandKRF/2005

CD88510-1-AThailandKRF/2005

CD88509-1-AThailandKRF/2005

CD88508-1-AThailandKRF/2005

CD88507-1-AThailandKRF/2005

CD88506-1-AThailandKRF/2005

CD88505-1-AThailandKRF/2005

CD88504-1-AThailandKRF/2005

CD88503-1-AThailandKRF/2005

CD88502-1-AThailandKRF/2005

CD88501-1-AThailandKRF/2005

CD88500-1-AThailandKRF/2005

CD88499-1-AThailandKRF/2005

CD88498-1-AThailandKRF/2005

CD88497-1-AThailandKRF/2005

CD88496-1-AThailandKRF/2005

CD88495-1-AThailandKRF/2005

CD88494-1-AThailandKRF/2005

CD88493-1-AThailandKRF/2005

CD88492-1-AThailandKRF/2005

CD88491-1-AThailandKRF/2005

CD88490-1-AThailandKRF/2005

CD88489-1-AThailandKRF/2005

CD88488-1-AThailandKRF/2005

CD88487-1-AThailandKRF/2005

CD88486-1-AThailandKRF/2005

CD88485-1-AThailandKRF/2005

CD88484-1-AThailandKRF/2005

CD88483-1-AThailandKRF/2005

1

2.2

2.1.2

2.1.3

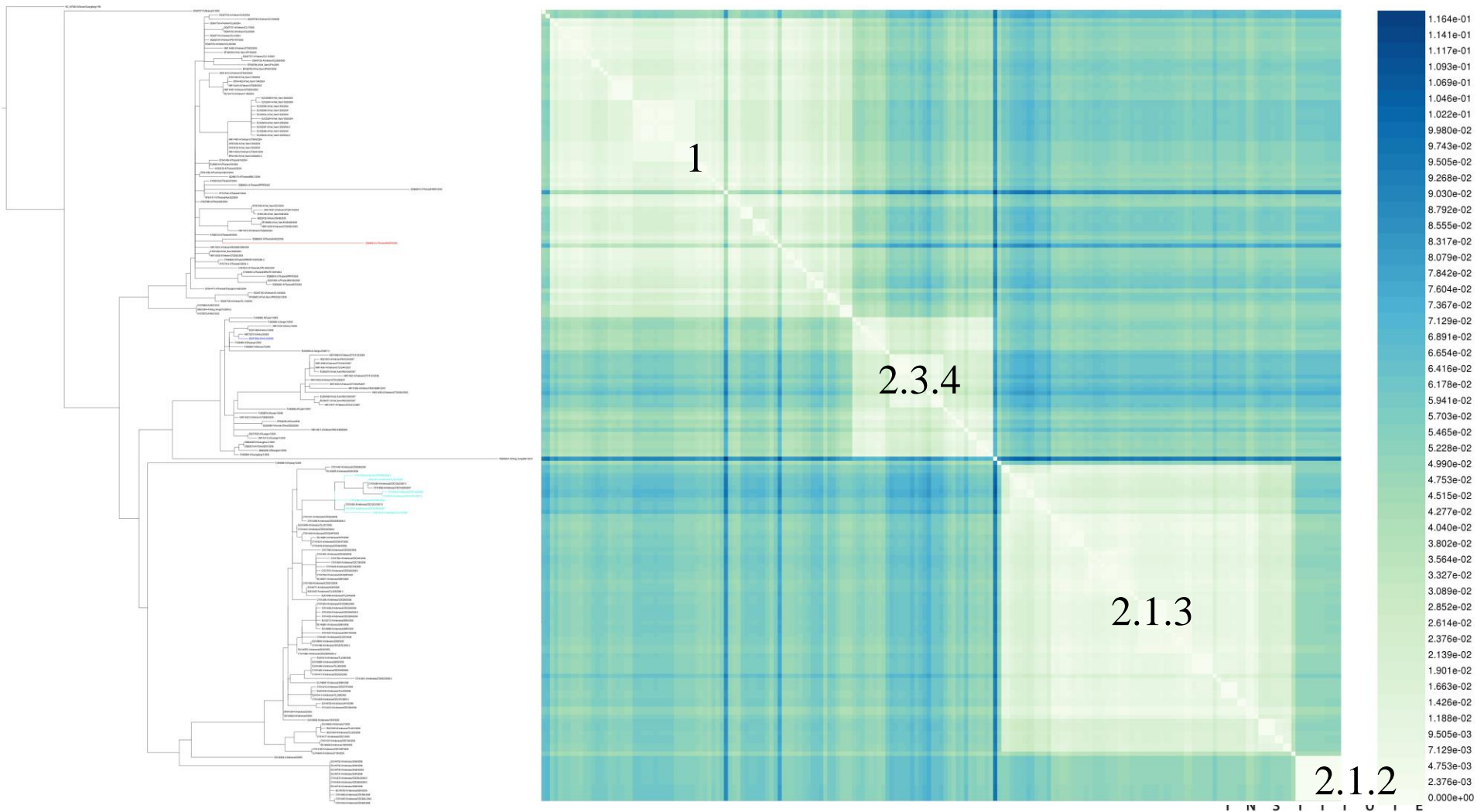
2.3.4

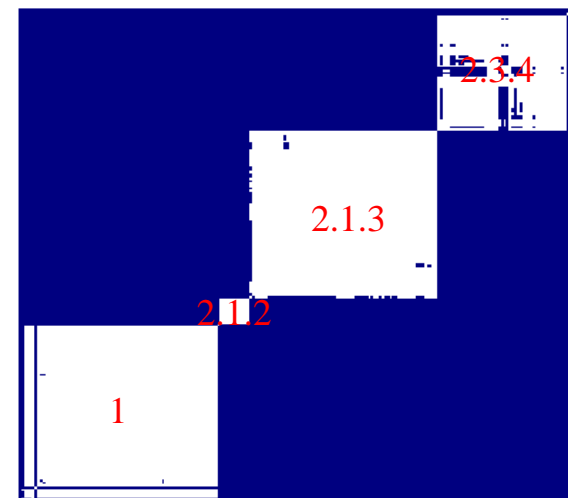
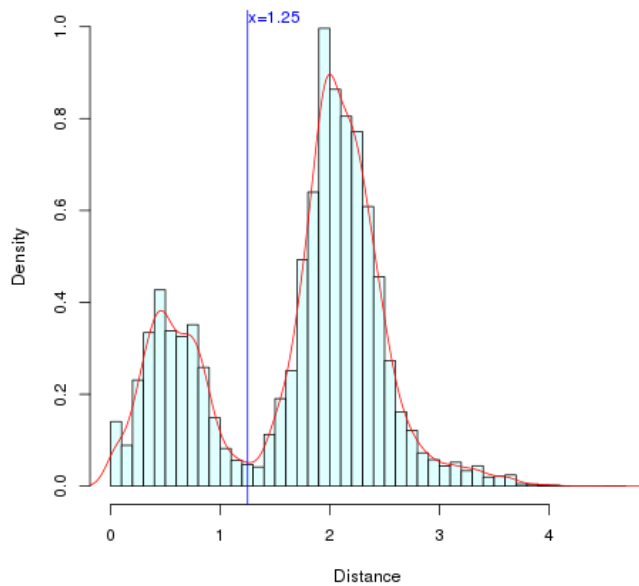
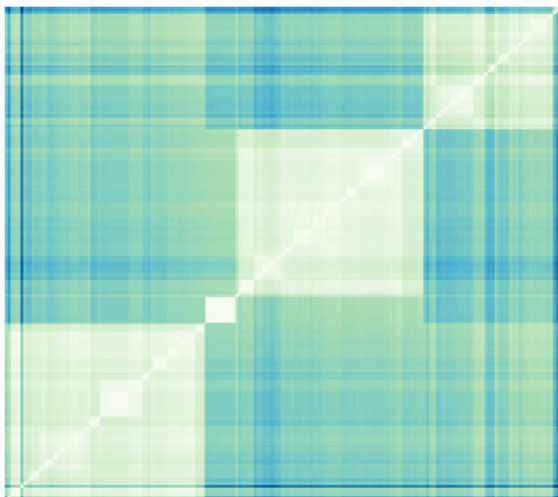
2.3.2.1

0.4



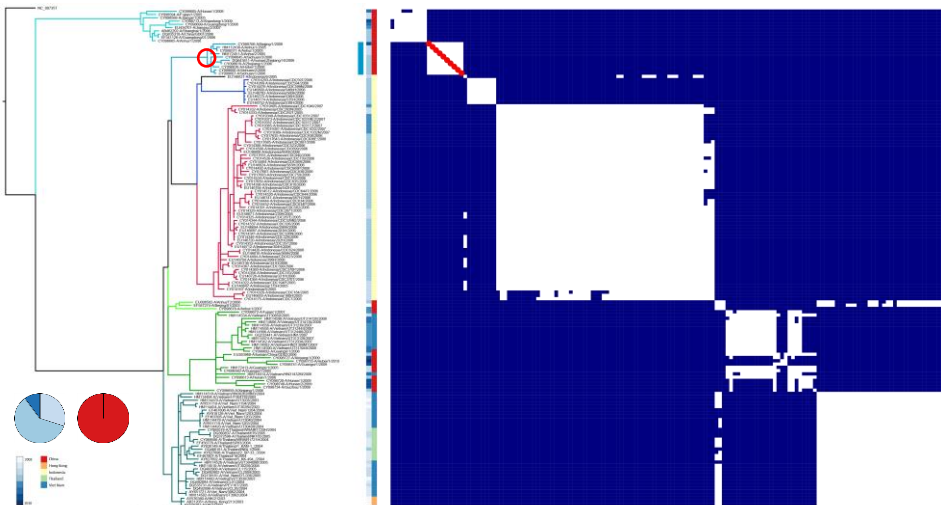
# Global analysis of human H5 HA clades



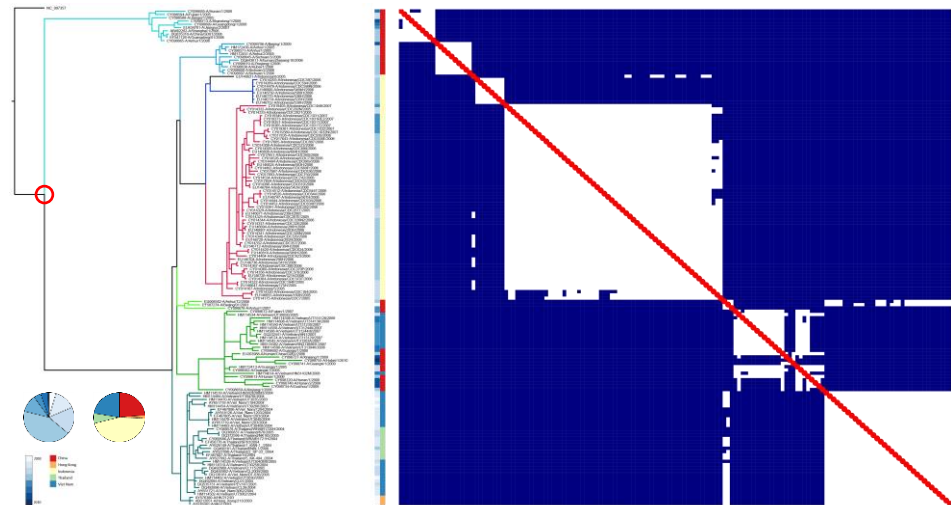


# PB2 Examples

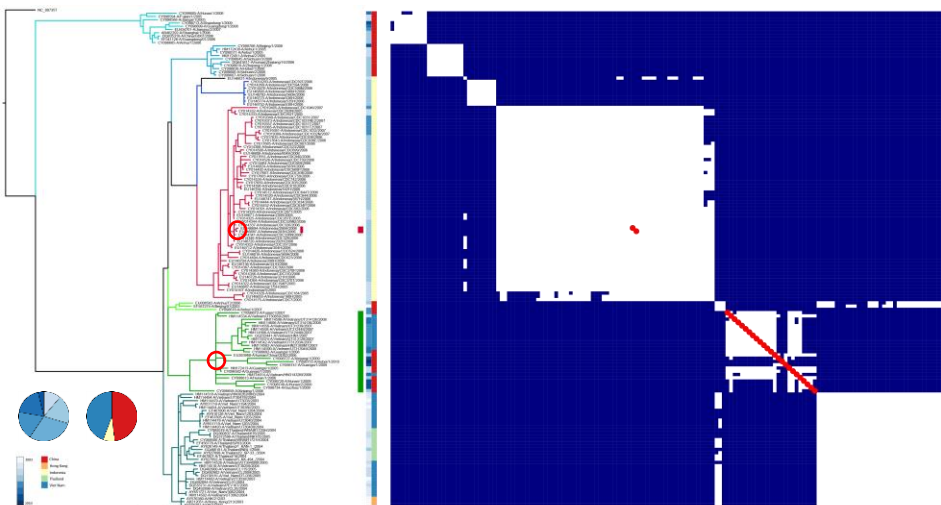
64T



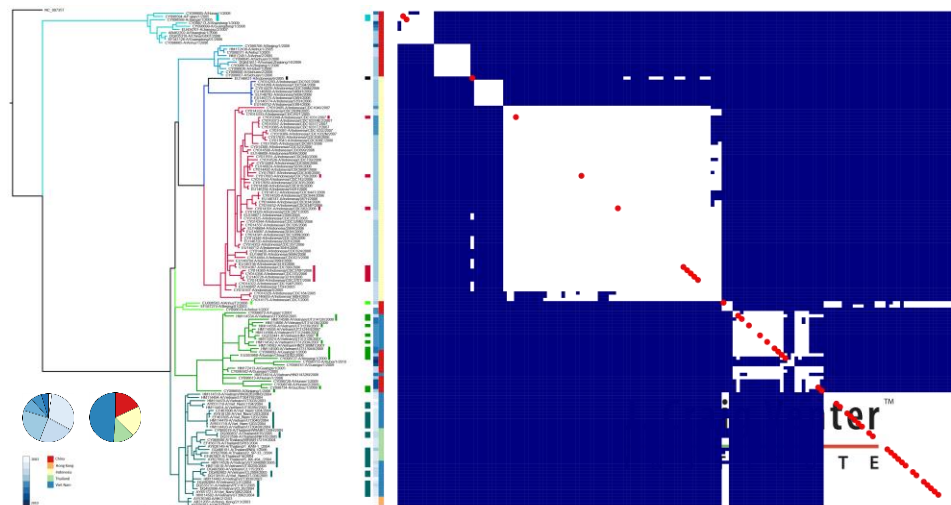
93P

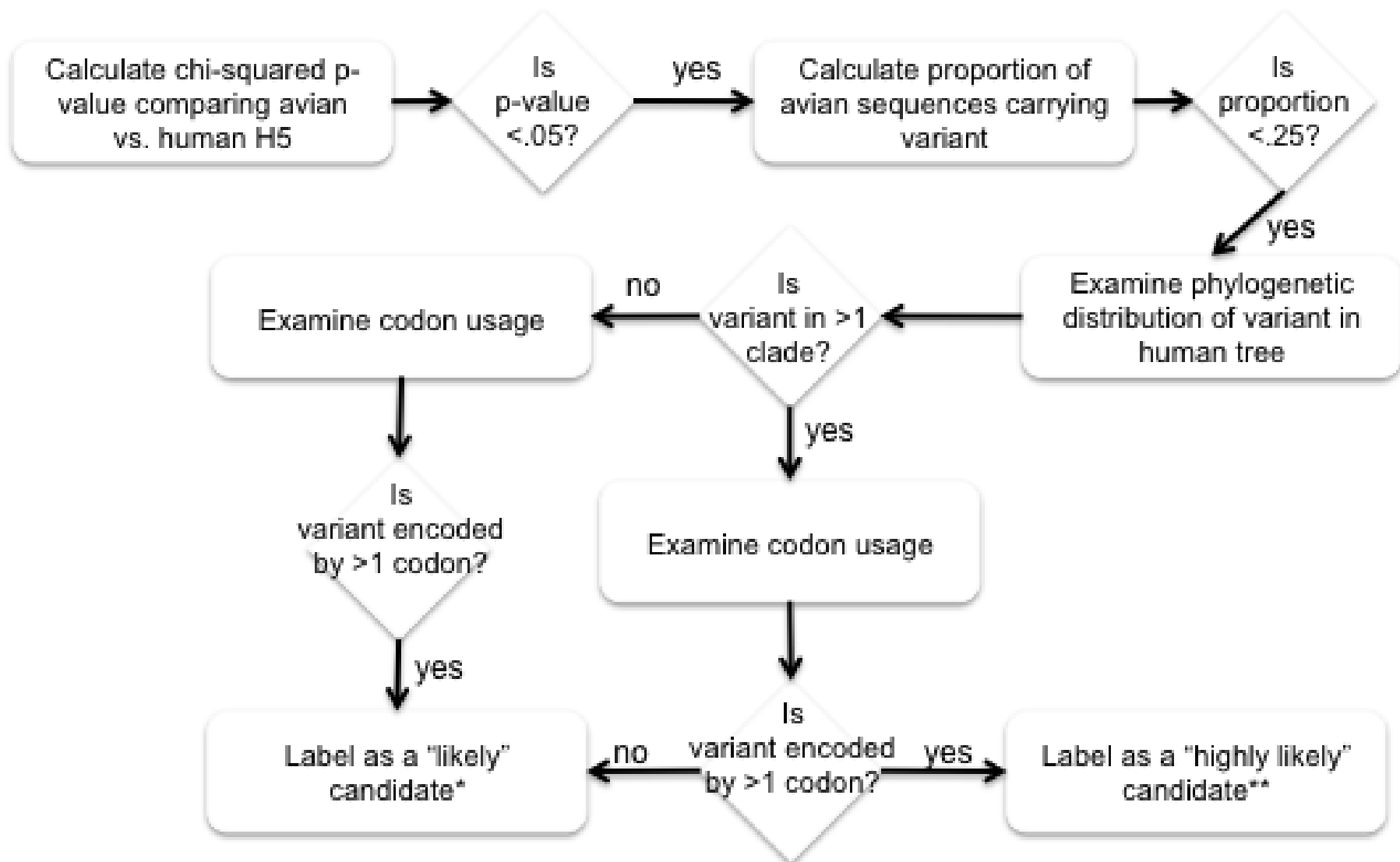


390N



627K





# Analysis of H5N1 NA protein

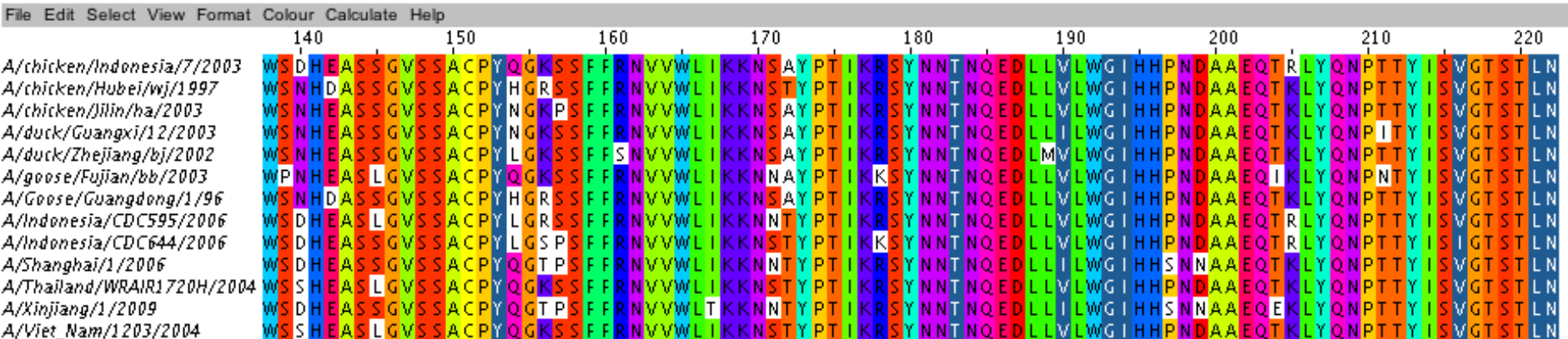
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Ref.	Avian (283)			Human (224)			P-value	Codon	Result 1	Proportion of avian sequences	Significantly different between human and avian (p-value <.05)	Relatively rare in avian (<.25 proportion)	Multiple clades	Multiple codonS	High confidence
2	Position	Position	# Sequences	Amino acid	Position	# Sequences	Amino acid									
3	29	35	0	-	39	26	L	2.40E-10	CTT (1)	F	0	Y	Y	Y	Y	**
4									TTG (20)			Y	Y	Y	Y	**
5			125	M		182	M	< 2.20E-16	ATG (143)	F	0.441696113	Y	N	Y	N	
6	95	101	13	N	106	72	N	< 2.20E-16	AAC (38)	F	0.045936396	Y	Y	Y	Y	**
7									AAT (23)			Y	Y	Y	Y	**
8			21	R		140	R	< 2.20E-16	AGA (107)	F	0.074204947	Y	Y	Y	N	*
9	105	111	3	N	116	72	N	< 2.20E-16	AAT (50)	F	0.010600707	Y	Y	N	N	
10			134	S		145	S	6.68E-05	AGC (5)	F w/R	0.473498233	Y	N	Y	Y	
11									AGT (117)			Y	N	Y	Y	
12	454	463	15	S	466	139	S	< 2.20E-16	AGC (7)	C	0.053003534	Y	Y	Y	Y	**
13									AGT (99)			Y	Y	Y	Y	**
14	46	52	0	-	57	4	D	0.03699	GAT (3)	F	0	Y	Y	N	N	
15			0	-		14	S	8.10E-06	TCT (13)	F	0	Y	Y	N	N	
16			1	V		22	V	1.09E-07	GTT (20)	F	0.003533569	Y	Y	Y	N	*
17	48	54	15	S	59	89	S	< 2.20E-16	TCA (60)	C	0.053003534	Y	Y	Y	N	*
18	70	77	0	-	81	11	I	1.04E-04	ATC (9)	F	0	Y	Y	N	N	
19			0	-		1	K	0.4402	AAA (1)	NS	0	N	Y	N	N	
20			0	-		29	R	1.55E-11	AGA (27)	F	0	Y	Y	N	N	
21	74	80	14	P	85	88	P	< 2.20E-16	CCT (59)	F	0.049469965	Y	Y	Y	N	*
22	78	84	1	N	89	30	N	1.19E-10	AAC (27)	F	0.003533569	Y	Y	N	Y	*
23									AAT (1)			Y	Y	N	Y	*
24	84	90	0	-	95	3	I	0.08467	ATA (3)	NS	0	N	Y	N	N	
25			13	K		72	K	< 2.20E-16	AAA (61)	F	0.045936396	Y	Y	Y	N	*
26	155	161	202	H	166	221	H	< 2.20E-16	CAC (174)	F	0.713780919	Y	N	Y	Y	
27									CAT (2)			Y	N	Y	Y	
28	201	207	0	-	212	70	E	< 2.20E-16	GAG (48)	F	0	Y	Y	N	N	
29	253	261	13	H	264	74	H	< 2.20E-16	CAT (61)	F	0.045936396	Y	Y	Y	N	*
30	258	266	178	M	269	223	M	< 2.20E-16	ATG (167)	F	0.628975265	Y	N	Y	N	
31			0	-		1	W	0.4407	TGG (1)	NS	0	N	Y	N	N	
32	340	349	2	H	352	13	H	8.64E-04	CAT (11)	F	0.007067138	Y	Y	N	N	
33			50	S		89	S	2.64E-08	TCT (70)	C	0.176678445	Y	Y	Y	N	*
34			0	-		1	T	0.4396	ACT (1)	NS	0	N	Y	N	N	
35	382	391	36	E	394	104	E	2.43E-14	GAA (91)	C	0.127208481	Y	Y	Y	N	*

A	B				C	D	E	F	G	
Protein	Reference strain				Position	Human Residue	Avian Residue	p-value	Confidence level	
HA	A/Viet Nam/1203/2004(H5N1)				102	T	A	< 2.20E-16	*	
						V		2.04E-14	*	
					140	D	N	< 2.20E-16	**	
						S		4.70E-06	*	
					156	T	K/R	3.75E-13	**	←
					157	P	S	< 2.20E-16	**	←
					172	T	A	< 2.20E-16	*	
					178	K	R	< 2.20E-16	*	
					216	I	V	< 2.20E-16	*	←
					228	R	K	< 2.20E-16	*	
					498	I	V	< 2.20E-16	*	
					504	N	D	< 2.20E-16	*	
549	I	V	5.41E-16	*						
	M		9.38E-07	*						

NP	A/WSN/1933(H1N1)	400	K	R	< 2.20E-16	**
NS1	A/Udorn/8/1972(H3N2)	7	L	S	< 2.20E-16	*
		79	K	M	< 2.20E-16	*
		214	F	L	< 2.20E-16	*
NS2	A/Udorn/8/1972(H3N2)	7	L	S	< 2.20E-16	*
		36	G	E	6.69E-13	**
PA	A/WSN/1933(H1N1)	520	Y	F	< 2.20E-16	*
		653	S	P	< 2.20E-16	*
PB1	A/Hong Kong/156/97(H5N1)	400	A	T	< 2.20E-16	*
		644	I	V	< 2.20E-16	*
PB2	A/Viet Nam/1203/2004(H5N1)	108	V	T	3.41E-07	*
		147	T	I	1.25E-08	*
		292	T	I	< 2.20E-16	*
		339	T	K	< 2.20E-16	**
		368	Q	R	9.22E-09	*
		390	N	D	6.13E-10	**
		526	R	K	< 2.20E-16	**
		627	K	E	< 2.20E-16	**
660	R	K	6.66E-09	*		

51 candidate adaptive drivers

# Avian + Human H5 HA Reference Alignment



↑  
140D/S

↑ ↑  
156T 157P

↑ ↑  
172T 178K

↑  
216I

<input checked="" type="checkbox"/>	Influenza A_H5_SF16	Influenza A_H5_sialic-acid-binding-site_98(14)	98,136,153,183,190,193,194,216,221,222,225,226,227,228
-------------------------------------	---------------------	--	--



# Sialic Acid Binding Site

## HIGHLIGHT SEQUENCE FEATURES

Hide Sequence Features

Select a Category

- all
- epitope
- sequence alteration
- functional
- structural

Select a feature. Coordinates highlighted on alignment in dark blue. Only one sequence feature can be highlighted at a time.

	Sequence Feature	Feature Name	Positions
<input type="checkbox"/>	Influenza A_H5_SF15	Influenza A_H5_erythrocyte-binding-site_98(3)	98,183,194
<input checked="" type="checkbox"/>	Influenza A_H5_SF16	Influenza A_H5_sialic-acid-binding-site_98(14)	98,136,153,183,190,193,194,216,221,222,225,226,227,228
<input type="checkbox"/>	Influenza A_H5_SF17	Influenza A_H5_signal-peptide_1(16)	1-16

File Edit Select View Format Colour Calculate Help

	140	150	160	170	180	190	200	210	220																																																																											
<i>A/chicken/Indonesia/7/2003</i>	WS	D	E	A	S	S	G	V	S	S	A	C	P	Y	Q	G	K	S	S	F	F	R	N	V	V	W	L	I	K	K	N	S	A	Y	P	T	I	K	R	S	Y	N	N	T	N	Q	E	D	L	L	V	L	W	G	I	H	H	P	N	D	A	A	E	Q	T	R	L	Y	Q	N	P	T	T	Y	I	S	V	G	T	S	T	L	N	
<i>A/chicken/Hubei/wj/1997</i>	WS	N	H	D	A	S	S	G	V	S	S	A	C	P	Y	H	G	R	S	S	F	F	R	N	V	V	W	L	I	K	K	N	S	T	Y	P	T	I	K	R	S	Y	N	N	T	N	Q	E	D	L	L	V	L	W	G	I	H	H	P	N	D	A	A	E	Q	T	K	L	Y	Q	N	P	T	T	Y	I	S	V	G	T	S	T	L	N
<i>A/chicken/Jilin/ha/2003</i>	WS	N	H	E	A	S	S	G	V	S	S	A	C	P	Y	N	G	K	P	S	F	F	R	N	V	V	W	L	I	K	K	N	S	A	Y	P	T	I	K	R	S	Y	N	N	T	N	Q	E	D	L	L	V	L	W	G	I	H	H	P	N	D	A	A	E	Q	T	K	L	Y	Q	N	P	T	T	Y	I	S	V	G	T	S	T	L	N
<i>A/duck/Guangxi/12/2003</i>	WS	N	H	E	A	S	S	G	V	S	S	A	C	P	Y	N	G	K	S	S	F	F	R	N	V	V	W	L	I	K	K	N	S	A	Y	P	T	I	K	R	S	Y	N	N	T	N	Q	E	D	L	L	I	L	W	G	I	H	H	P	N	D	A	A	E	Q	T	K	L	Y	Q	N	P	I	T	Y	I	S	V	G	T	S	T	L	N
<i>A/duck/Zhejiang/bj/2002</i>	WS	N	H	E	A	S	S	G	V	S	S	A	C	P	Y	L	G	K	S	S	F	F	R	N	V	V	W	L	I	K	K	N	S	A	Y	P	T	I	K	R	S	Y	N	N	T	N	Q	E	D	L	M	V	L	W	G	I	H	H	P	N	D	A	A	E	Q	T	K	L	Y	Q	N	P	T	T	Y	I	S	V	G	T	S	T	L	N
<i>A/goose/Fujian/bb/2003</i>	WP	N	H	E	A	S	L	G	V	S	S	A	C	P	Y	Q	G	K	S	S	F	F	R	N	V	V	W	L	I	K	K	N	S	A	Y	P	T	I	K	K	S	Y	N	N	T	N	Q	E	D	L	L	V	L	W	G	I	H	H	P	N	D	A	A	E	Q	I	K	L	Y	Q	N	P	N	T	Y	I	S	V	G	T	S	T	L	N
<i>A/Goose/Guangdong/1/96</i>	WS	N	H	D	A	S	S	G	V	S	S	A	C	P	Y	H	G	R	S	S	F	F	R	N	V	V	W	L	I	K	K	N	S	A	Y	P	T	I	K	R	S	Y	N	N	T	N	Q	E	D	L	L	V	L	W	G	I	H	H	P	N	D	A	A	E	Q	T	K	L	Y	Q	N	P	T	T	Y	I	S	V	G	T	S	T	L	N
<i>A/Indonesia/CDC595/2006</i>	WS	D	H	E	A	S	L	G	V	S	S	A	C	P	Y	L	G	R	S	S	F	F	R	N	V	V	W	L	I	K	K	N	S	T	Y	P	T	I	K	R	S	Y	N	N	T	N	Q	E	D	L	L	V	L	W	G	I	H	H	P	N	D	A	A	E	Q	T	R	L	Y	Q	N	P	T	T	Y	I	S	V	G	T	S	T	L	N
<i>A/Indonesia/CDC644/2006</i>	WS	D	H	E	A	S	S	G	V	S	S	A	C	P	Y	L	G	S	P	S	F	F	R	N	V	V	W	L	I	K	K	N	S	T	Y	P	T	I	K	K	S	Y	N	N	T	N	Q	E	D	L	L	V	L	W	G	I	H	H	P	N	D	A	A	E	Q	T	R	L	Y	Q	N	P	T	T	Y	I	S	V	G	T	S	T	L	N
<i>A/Shanghai/1/2006</i>	WS	D	H	E	A	S	S	G	V	S	S	A	C	P	Y	Q	G	T	P	S	F	F	R	N	V	V	W	L	I	K	K	N	S	T	Y	P	T	I	K	R	S	Y	N	N	T	N	Q	E	D	L	L	I	L	W	G	I	H	H	S	N	N	A	A	E	Q	T	K	L	Y	Q	N	P	T	T	Y	I	S	V	G	T	S	T	L	N
<i>A/Thailand/WRAIR1720H/2004</i>	WS	S	H	E	A	S	L	G	V	S	S	A	C	P	Y	Q	G	K	S	S	F	F	R	N	V	V	W	L	I	K	K	N	S	T	Y	P	T	I	K	R	S	Y	N	N	T	N	Q	E	D	L	L	V	L	W	G	I	H	H	P	N	D	A	A	E	Q	T	K	L	Y	Q	N	P	T	T	Y	I	S	V	G	T	S	T	L	N
<i>A/Xinjiang/1/2009</i>	WS	D	H	E	A	S	S	G	V	S	S	A	C	P	Y	Q	G	T	P	S	F	F	R	N	V	V	W	L	I	K	K	N	S	T	Y	P	T	I	K	R	S	Y	N	N	T	N	Q	E	D	L	L	I	L	W	G	I	H	H	S	N	N	A	A	E	Q	E	K	L	Y	Q	N	P	T	T	Y	I	S	V	G	T	S	T	L	N
<i>A/Viet_Nam/1203/2004</i>	WS	S	H	E	A	S	L	G	V	S	S	A	C	P	Y	Q	G	K	S	S	F	F	R	N	V	V	W	L	I	K	K	N	S	T	Y	P	T	I	K	R	S	Y	N	N	T	N	Q	E	D	L	L	V	L	W	G	I	H	H	P	N	D	A	A	E	Q	T	K	L	Y	Q	N	P	T	T	Y	I	S	V	G	T	S	T	L	N



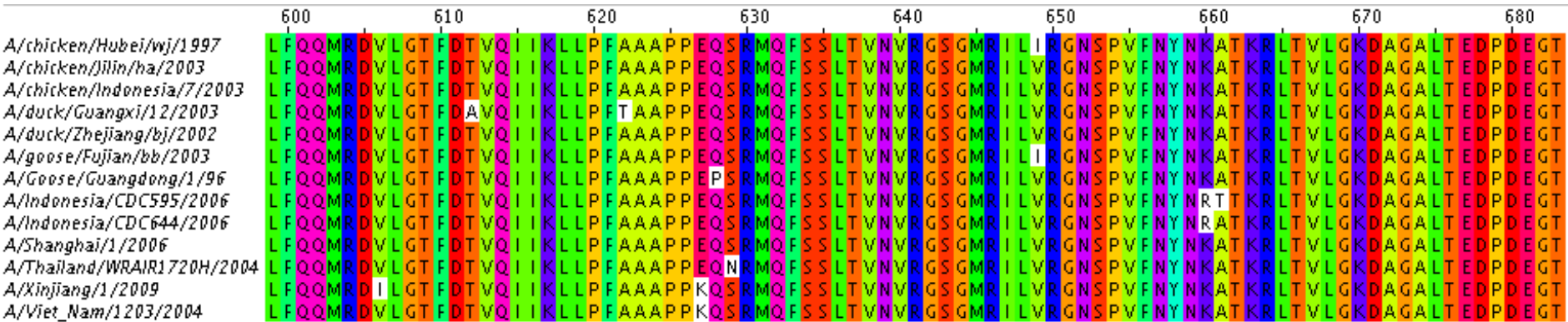
A	B	C	D	E	F	G
Protein	Reference strain	Position	Human Residue	Avian Residue	p-value	Confidence level
		102	T	A	< 2.20E-16	*
			V		2.04E-14	*

<b>PB2</b>	<b>A/Viet Nam/1203/2004(H5N1)</b>	<b>108</b>	<b>V</b>	<b>T</b>	<b>3.41E-07</b>	<b>*</b>
		<b>147</b>	<b>T</b>	<b>I</b>	<b>1.25E-08</b>	<b>*</b>
		<b>292</b>	<b>T</b>	<b>I</b>	<b>&lt; 2.20E-16</b>	<b>*</b>
		<b>339</b>	<b>T</b>	<b>K</b>	<b>&lt; 2.20E-16</b>	<b>**</b>
		<b>368</b>	<b>Q</b>	<b>R</b>	<b>9.22E-09</b>	<b>*</b>
		<b>390</b>	<b>N</b>	<b>D</b>	<b>6.13E-10</b>	<b>**</b>
		<b>526</b>	<b>R</b>	<b>K</b>	<b>&lt; 2.20E-16</b>	<b>**</b>
		<b>627</b>	<b>K</b>	<b>E</b>	<b>&lt; 2.20E-16</b>	<b>**</b>
		<b>660</b>	<b>R</b>	<b>K</b>	<b>6.66E-09</b>	<b>*</b>

<b>NA</b>	<b>A/California/07/2009(H1N1)</b>	48	S	P	< 2.20E-16	*
		74	P	F	< 2.20E-16	*
		78	N	K	1.19E-10	*
		84	K	T	< 2.20E-16	*
		95	N	S	< 2.20E-16	**
			R		< 2.20E-16	*
		253	H	Y	< 2.20E-16	*
		340	S	P	2.64E-08	*
		382	E	G	2.43E-14	*
		454	S	G	< 2.20E-16	**
<b>NP</b>	<b>A/WSN/1933(H1N1)</b>	400	K	R	< 2.20E-16	**
<b>NS1</b>	<b>A/Udorn/8/1972(H3N2)</b>	7	L	S	< 2.20E-16	*
		79	K	M	< 2.20E-16	*
		214	F	L	< 2.20E-16	*
<b>NS2</b>	<b>A/Udorn/8/1972(H3N2)</b>	7	L	S	< 2.20E-16	*
		36	G	E	6.69E-13	**
<b>PA</b>	<b>A/WSN/1933(H1N1)</b>	520	Y	F	< 2.20E-16	*
		653	S	P	< 2.20E-16	*
<b>PB1</b>	<b>A/Hong Kong/156/97(H5N1)</b>	400	A	T	< 2.20E-16	*
		644	I	V	< 2.20E-16	*
<b>PB2</b>	<b>A/Viet Nam/1203/2004(H5N1)</b>	108	V	T	3.41E-07	*
		147	T	I	1.25E-08	*
		292	T	I	< 2.20E-16	*
		339	T	K	< 2.20E-16	**
		368	Q	R	9.22E-09	*
		390	N	D	6.13E-10	**
		526	R	K	< 2.20E-16	**
		627	K	E	< 2.20E-16	**
660	R	K	6.66E-09	*		

51 candidate adaptive drivers

# Avian + Human H5 PB2 Reference Alignment



↑  
627K

↑  
660R

# E627K and species jump

Proc Natl Acad Sci U S A. 2005 Dec 20;102(51):18590-5. Epub 2005 Dec 8.

## **The viral polymerase mediates adaptation of an avian influenza virus to a mammalian host.**

Gabriel G, Dauber B, Wolff T, Planz O, Klenk HD, Stech J.

Institut für Virologie, Universitätsklinikum Giessen und Marburg, D-35033 Marburg, Germany.

J Gen Virol. 2007 Feb;88(Pt 2):547-53.

## **Adaptation of an H7N7 equine influenza A virus in mice.**

Shinya K, Watanabe S, Ito T, Kasai N, Kawaoka Y.

Department of Pathobiological Sciences, University of Wisconsin-Madison, WI 53706, USA.

# Gain-of-Function vs. Loss-of-Function

- Loss-of-Function
  - Start with a common human backbone that carries the human-enriched variants
  - Revert one or more residues back to avian residue
  - Test for decreased infection, transmission and/or pathogenesis in ferret model
- Gain-of-Function
  - Start with a common avian backbone
  - Introduce one or more of human-enriched sequence variants
  - Test for altered protein activity in relevant in vitro biochemical assays (e.g. HA binding in glycan arrays)

A	B	C	D	E	F	G	L	M	N	O	P	Q	
A	B	C	D	E	F	G	L	M					
Protein	Reference strain	Position	Human Residue	Avian Residue	p-value	Confidence level	A/Indonesia/CDC595/2006	A/Indonesia/CDC644/2006					
HA	A/Viet Nam/1203/2004(H5N1)	102	T	A	< 2.20E-16	*		X					
			V		2.04E-14	*							
		140	D	N	< 2.20E-16	**	X	X					
			S		4.70E-06	*							
		156	T	K/R	3.75E-13	**							
		157	P	S	< 2.20E-16	**		X					
		172	T	A	< 2.20E-16	*	X	X					
		178	K	R	< 2.20E-16	*		X					
		216	I	V	< 2.20E-16	*		X					
		228	R	K	< 2.20E-16	*							
		498	I	V	< 2.20E-16	*	X	X					
		504	N	D	< 2.20E-16	*	X	X					
549	I	V	5.41E-16	*		X							
	M		9.38E-07	*	X								
		95	R	S	<2.20E-16	*	x					x	
		253	H	Y	<2.20E-16	*				x	x		
		340	S	P	2.64E-08	*				x	x		
		382	E	G	2.43E-14	*				x	x	x	
		454	S	G	<2.20E-16	**	x		x			x	
NP	A/WSN/1933(H1N1)	400	K	R	<2.20E-16	**	x	x					
NS1	A/Udorn/8/1972(H3N2)	7	L	S	<2.20E-16	*	x	x					
		79	K	M	<2.20E-16	*	x	x					
		214	F	L	<2.20E-16	*	x	x					
NS2	A/Udorn/8/1972(H3N2)	7	L	S	<2.20E-16	*	x	x					
		36	G	E	6.69E-13	**		x				x	
PA	A/WSN/1933(H1N1)	520	Y	F	<2.20E-16	*	x	x					
		653	S	P	<2.20E-16	*	x	x					
PB1	A/Hong Kong/156/97(H5N1)	400	A	T	<2.20E-16	*	x	x			x		
		644	I	V	<2.20E-16	*						x	
PB2	A/Viet Nam/1203/2004(H5N1)	108	V	T	3.41E-07	*	x						
		147	T	I	1.25E-08	*			x				
		292	T	I	<2.20E-16	*	x	x					
		339	T	K	<2.20E-16	**		x	x				
		368	Q	R	9.22E-09	*			x				
		390	N	D	6.13E-10	**			x				
		526	R	K	<2.20E-16	**		x					
		627	K	E	<2.20E-16	**			x	x			
660	R	K	6.66E-09	*	x	x							

# Summary

- Human influenza pandemics are initiated by species jump events followed by sustained human to human transmission ( $R_0 > 1$ )
- Multiple independent occurrences of the same mutation during stuttering transmission is evidence of convergent evolution of adaptive drivers
- A combination of statistical and phylogenetic analysis can reveal candidate adaptive drivers – *hypotheses for experimental testing*
- Surveillance for adaptive drivers in reservoir species could help anticipate the next pandemic

# Acknowledgements

- U.T. Southwestern
  - Richard Scheuermann (PI)
  - Burke Squires
  - Jyothi Noronha
  - Shubhada Godbole
  - **Brett Pickett**
  - Yun Zhang
  - **Haizhou Liu**
- MSSM
  - Adolfo Garcia-Sastre
  - Eric Bortz
  - Gina Conenello
  - Peter Palese
- Vecna
  - Chris Larsen
  - Al Ramsey
- LANL
  - Catherine Macken
  - Mira Dimitrijevic
- U.C. Davis
  - Nicole Baumgarth
- Northrop Grumman
  - Ed Klem
  - Mike Atassi
  - Kevin Biersack
  - Jon Dietrich
  - Wenjie Hua
  - Wei Jen
  - Sanjeev Kumar
  - Xiaomei Li
  - Zaigang Liu
  - Jason Lucas
  - Michelle Lu
  - Bruce Quesenberry
  - Barbara Rotchford
  - Hongbo Su
  - Bryan Walters
  - Jianjun Wang
  - Sam Zaremba
  - Liwei Zhou
- IRD SWG
  - Gillian Air, OMRF
  - Carol Cardona, Univ. Minnesota
  - Adolfo Garcia-Sastre, Mt Sinai
  - Elodie Ghedin, Univ. Pittsburgh
  - Martha Nelson, Fogarty
  - Daniel Perez, Univ. Maryland
  - Gavin Smith, Duke Singapore
  - David Spiro, JCVI
  - Dave Stallknecht, Univ. Georgia
  - David Topham, Rochester
  - Richard Webby, St Jude
- USDA
  - David Suarez
- Sage Analytica
  - Robert Taylor
  - Lone Simonsen
- CEIRS Centers

N01AI40041



National Institute of Allergy and Infectious Diseases

Leading research to understand, treat, and prevent infectious, immunologic, and allergic diseases.



SEARCH DATA

ANALYZE & VISUALIZE

WORKBENCH

SUBMIT DATA

## Search

Search our comprehensive database for:

- ▲ Influenza segment and protein sequences
- ▲ Avian and non-human mammalian surveillance data
- ▲ Virus phenotypic characteristics
- ▲ Host Factor Data (Prototype)
- ▲ Immune epitope data
- ▲ 3D protein structures

[Browse All Search Types](#)

## Analyze

Analyze data online:

- ▲ Align sequences
- ▲ Identify similar sequences (BLAST)
- ▲ Identify short peptides in flu proteins
- ▲ Identify point mutations in flu proteins
- ▲ Analyze Sequence Variation (SNP)
- ▲ Generate a phylogenetic tree

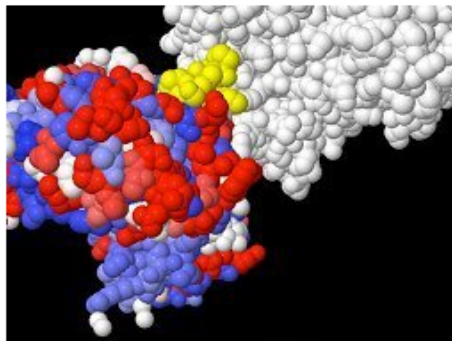
[Browse All Tools](#)

## Save to Workbench

Use your workbench to:

- ▲ Store sequences or other data in working sets for future analysis
- ▲ Combine working sets
- ▲ Integrate IRD data with your laboratory data
- ▲ Store analysis results
- ▲ Share results

### Highlights



#### 3D Protein Structure

Visualize protein structures in 3D. Users can display sequence conservation score on a structure and highlight experimentally determined epitopes as well.

##### Key Highlights:

- Visualize protein structure in 3D
- Display sequence conservation heat map on the structure
- Highlight sequence features (epitopes, etc.)
- Download highlighted protein structure image

[View 3D Structure](#)

[Start Search](#)

[Tutorial](#)

### Comparative Analysis of MERS-CoV Sequences

We have recently completed a [comparative genomics analysis](#) of Middle East Respiratory Syndrome Coronavirus (MERS-CoV) whole genome sequences, with implications for viral evolution, performed using the suite of bioinformatics tools available in VIPR. Also see a digest of recent events concerning MERS-CoV [here](#).

### What's New with Flu

An article in *Cell* says that a single amino acid change enhances H7N9 binding to lung receptors, but no new human H7N9 cases reported in a month. WHO implements a new 4-phase pandemic alert system, and issues a new H7N9 risk assessment. Find these and other flu developments in the latest [IRD Influenza Digest \(View Archive\)](#).

And be sure to see a [sequence analysis](#) with implications for H7N9 evolution, carried out by IRD scientists using IRD comparative genomics analysis tools.

▲ [Announcements](#)

▲ [Community Spotlight](#)

[View Archive](#)

• [St. Jude Children's Research Hospital CEIRS](#)

We would like to recognize the contribution of all our



# Outline

- Comparative genomics of H5N1 human adaptation
- Hands on workshop
  - NIAID Bioinformatics Resource Centers
  - Virus Pathogen Resource (ViPR)/Influenza Research Database (IRD) overview
  - Human adaptation of 2013 avian H7N9
  - Sequence annotation/submission/recombination
  - PlasmODB example
- T-shirt contest

# \* NIAID/DMID Genomics Program

## Sequencing

Genomic Sequencing Centers

## Functional Genomics

Functional Genomic Research Centers

## Proteomics

Clinical Proteomics Centers

## Structural Genomics

Structural Genomics Centers

## Systems Biology

Systems Biology Centers

Bioinformatics Resource Centers

## Genomic Research Resources

Genomic/Omics Data Sets, Databases, Bioinformatics Tools, Biomarkers, 3D Structures, Protein Clones, Predictive Models

To address key questions in microbiology and infectious disease

To identify new targets and develop new strategies for vaccines, diagnostics and therapeutics

# Bioinformatics Resource Centers (BRCs)

Welcome To VectorBase!

An NIAID Bioinformatics Resource Center for  
Invertebrate Vectors of Human Pathogens.



[www.vectorbase.org](http://www.vectorbase.org)



[www.eupathdb.org](http://www.eupathdb.org)



[www.patricbrc.org](http://www.patricbrc.org)



[www.viprbrc.org](http://www.viprbrc.org)



[www.fludb.org](http://www.fludb.org)

# Live Demo

# T-Shirt Question #1

- What is the case mortality rate (CMR) of H7N9 as of May 2013?

# T-Shirt Question #1

- What is the case mortality rate (CMR) of H7N9 as of May 2013?
- 28%

# T-Shirt Question #2

- What statistical test is used by meta-CATS to identify significant sequence variation differences between groups of sequences?

# T-Shirt Question #2

- What statistical test is used by meta-CATS to identify significant sequence variation differences between groups of sequences?
- Chi-squared test



# T-Shirt Question #3

- From what animal species did the initial H7N9 spillover event occur?

# T-Shirt Question #3

- From what animal species did the initial H7N9 spillover event occur?
- duck

# T-Shirt Question #4

- 2 part question – How many human H7N9 complete genomes are currently represented in GenBank? How many in IRD?

# T-Shirt Question #4

- 2 part question – How many human H7N9 complete genomes are currently represented in GenBank? How many in IRD?
- 11

# T-Shirt Question #5

- Name two different virus genome annotation bioinformatics tools?

# T-Shirt Question #5

- Name two different virus genome annotation bioinformatics tools?
- VIGOR and GATU

# T-Shirt Question #6

- What type of sequence feature is affected by variations in H7 positions 195 & 198?

# T-Shirt Question #6

- What type of sequence feature is affected by variations in H7 positions 195 & 198?
- Determinants of receptor binding