

Lecture 4: Influenza virus

Exercise: Optimisation

Given a function, find the maximum (or minimum) value

Find the minimum value of:

$$y = x^4 - 5x^3 - x^2 + 15x + 1$$

Basic Strategy:

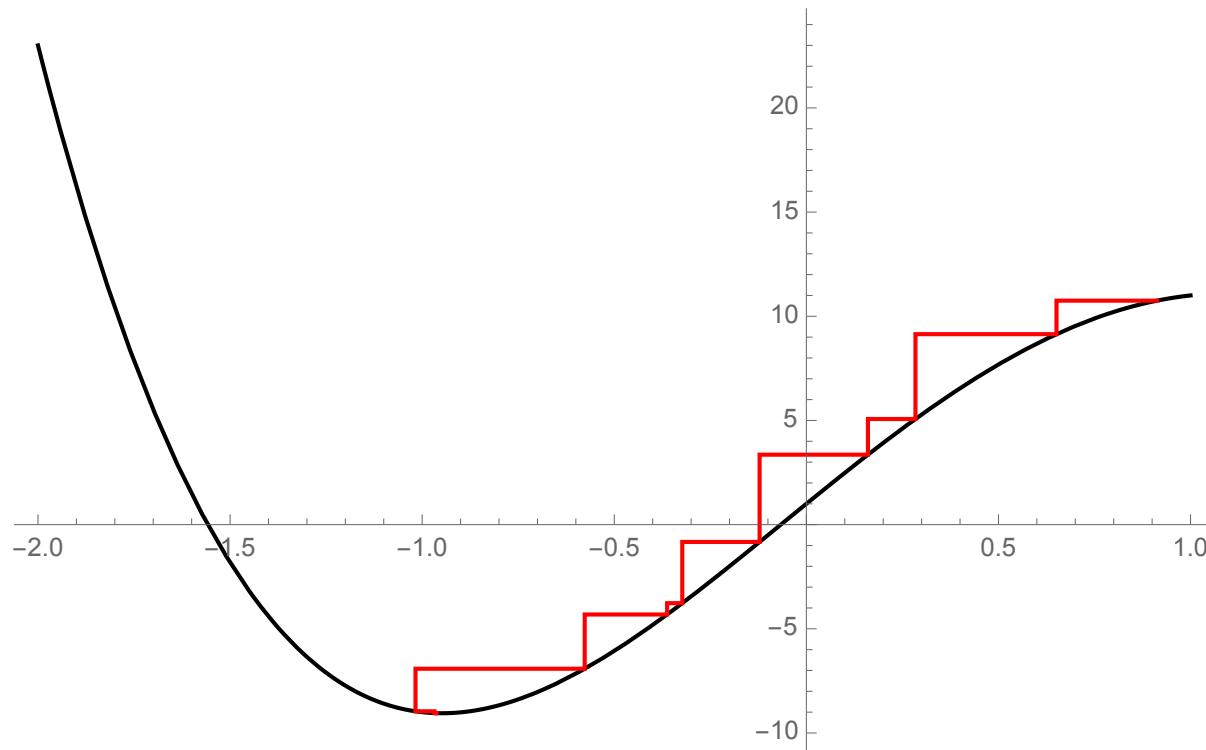
Guess an initial value of x

Calculate y

Change x. Keep the new x if this reduces y.

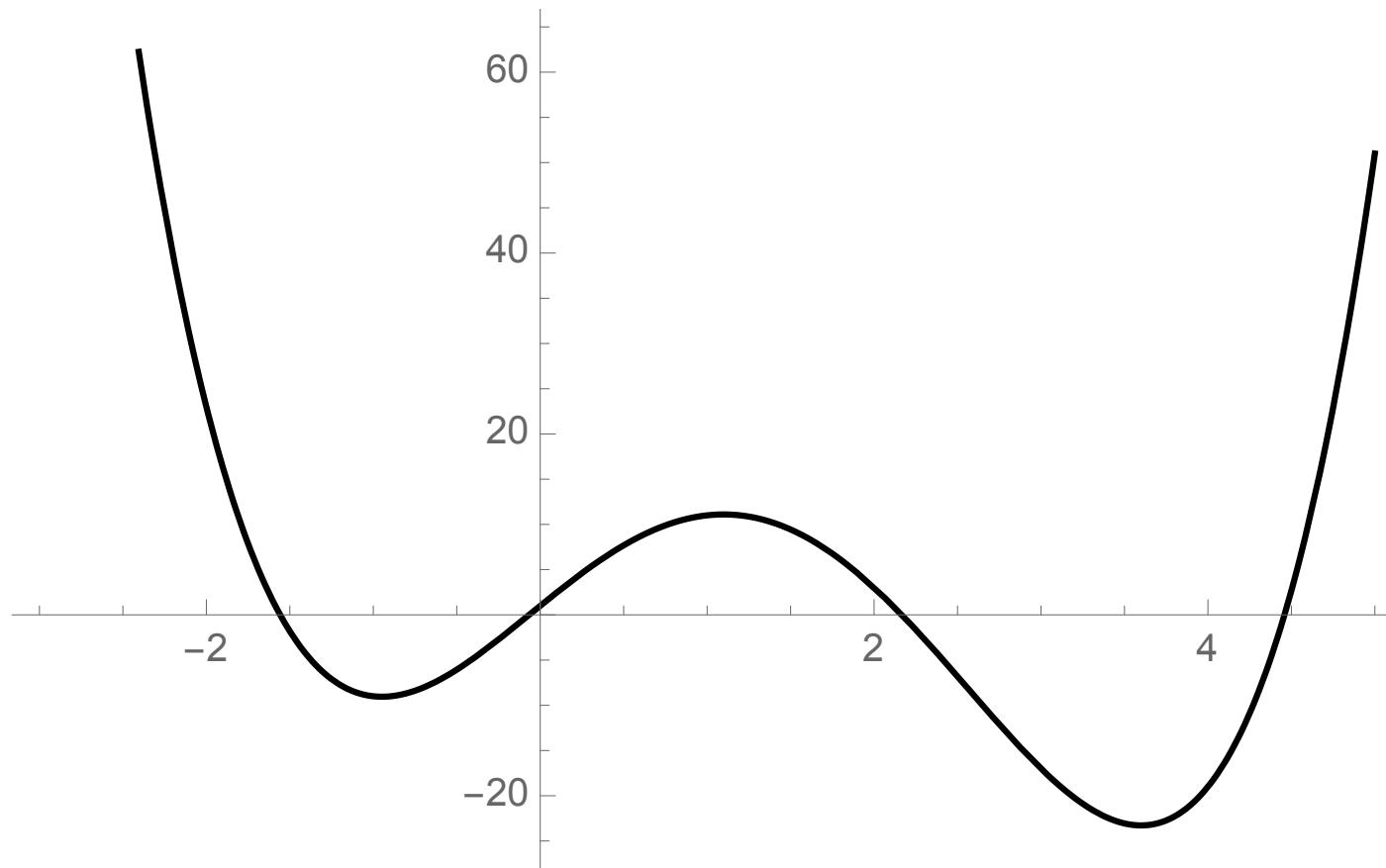
Exercise: Optimisation

Simple changes to value of x:



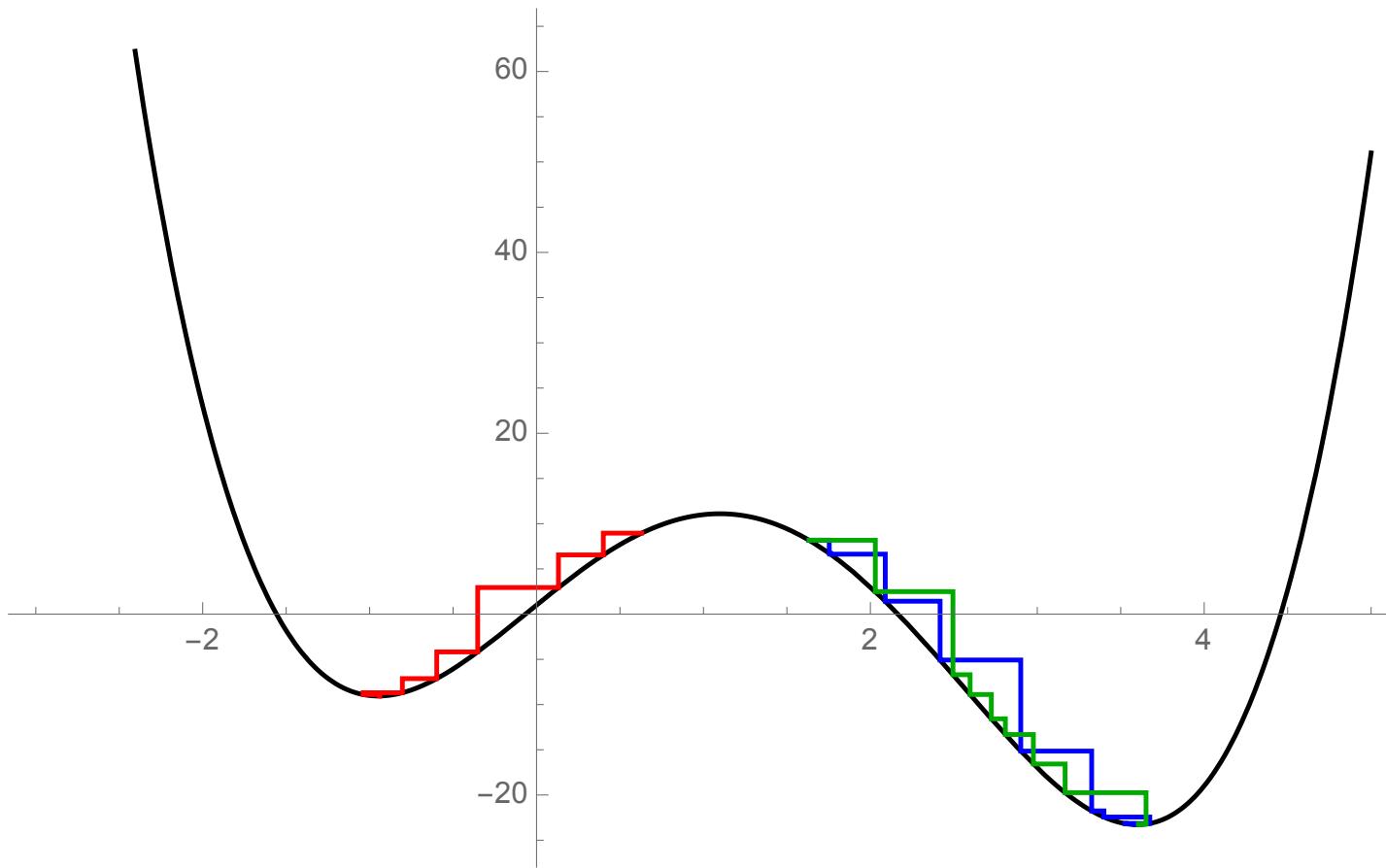
Exercise: Optimisation

May have multiple local minima:



Exercise: Optimisation

Multiple replicate calculations: Look for convergence

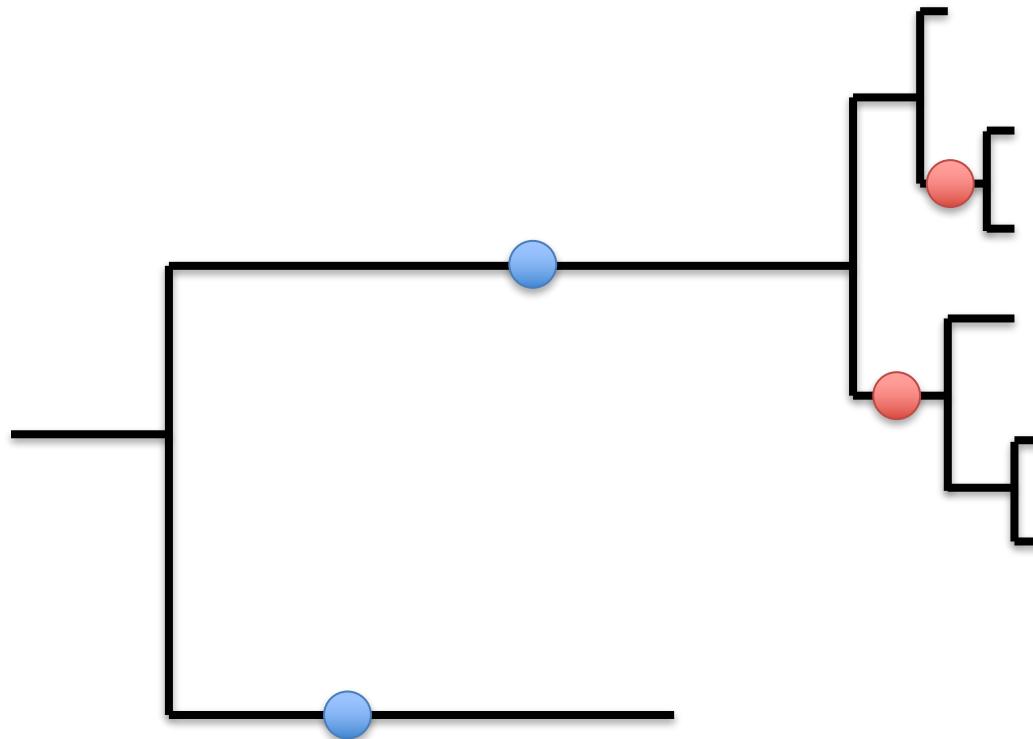


Phylogenetics + Coalescent theory

Phylogenetic models

Given a set of sequences, how are they related?

Aiming to retrieve a phylogeny: sequences connected by ‘relatedness’.

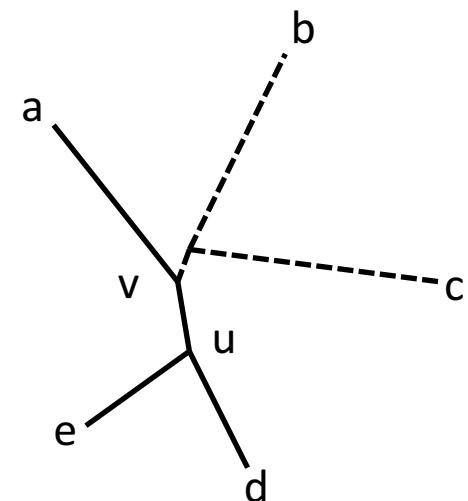
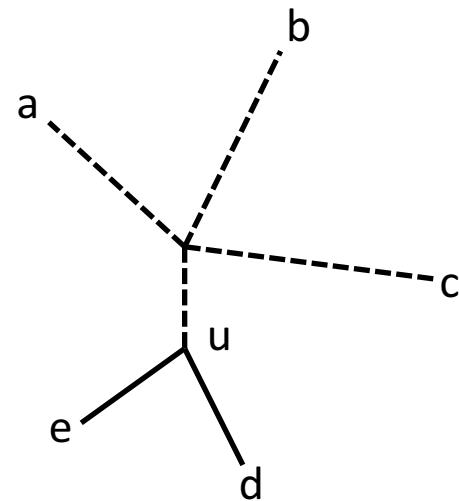
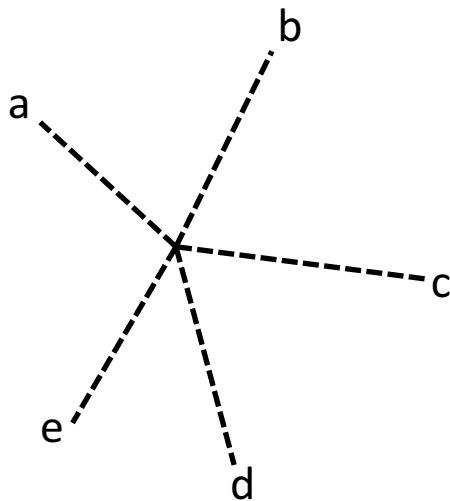


Phylogenetic models

Neighbour joining methods

Calculate a metric of genetic distances between sequences.

Join in a pairwise fashion by creating internal nodes

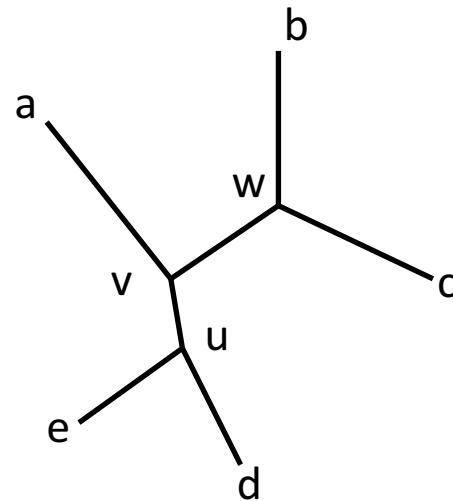
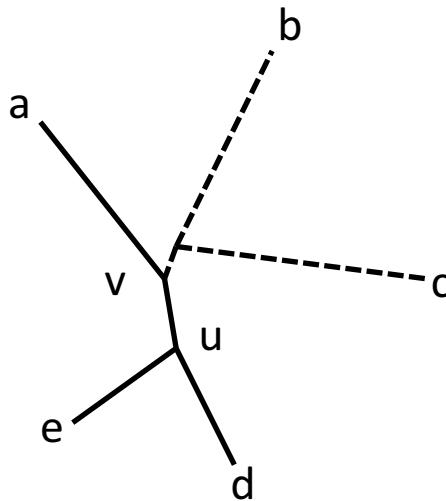


Phylogenetic models

Neighbour joining methods

Calculate a metric of genetic distances between sequences.

Join in a pairwise fashion by creating internal nodes

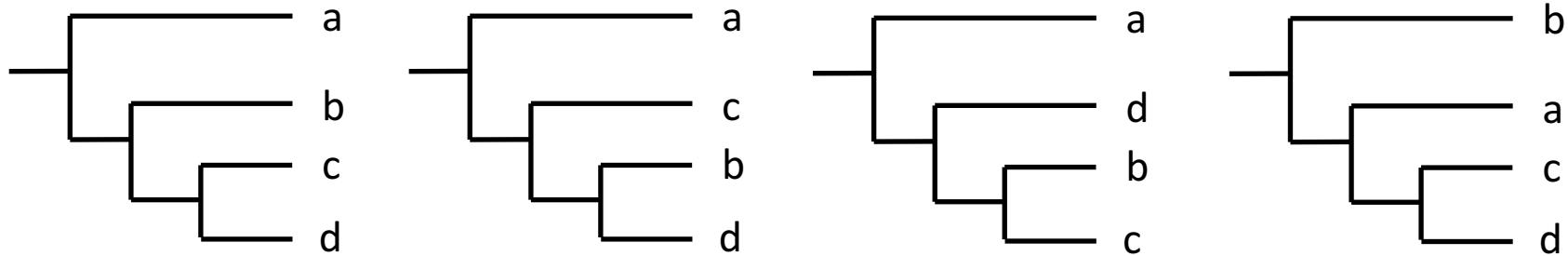
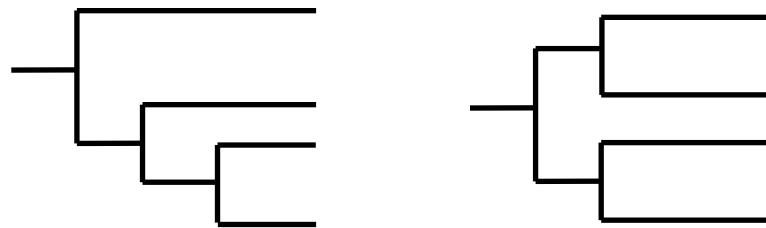


Phylogenetic models

Maximum parsimony method

Identify the network requiring the minimum number of substitutions to produce the observed outcome

Can be done by systematic search where there are few sequences



Phylogenetic models

Define substitution rates with matrices

Transition matrix P : $P_{ij}(t) = \text{Probability of mutating from base } i \text{ to } j \text{ in time } t$

Jukes-Cantor model: All mutations are equal

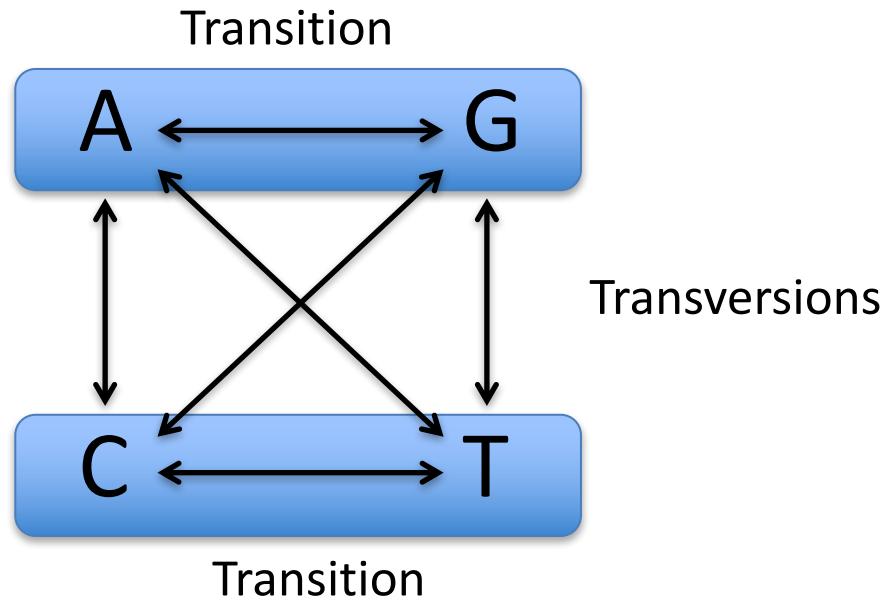
$$\begin{array}{cccc} & A & C & G & T \\ A & -\frac{3}{4}\mu & \frac{1}{4}\mu & \frac{1}{4}\mu & \frac{1}{4}\mu \\ C & \frac{1}{4}\mu & -\frac{3}{4}\mu & \frac{1}{4}\mu & \frac{1}{4}\mu \\ G & \frac{1}{4}\mu & \frac{1}{4}\mu & -\frac{3}{4}\mu & \frac{1}{4}\mu \\ T & \frac{1}{4}\mu & \frac{1}{4}\mu & \frac{1}{4}\mu & -\frac{3}{4}\mu \end{array}$$

Continuous-time Markov chain

Phylogenetic models

Models of substitution define mutation rates

More complex models



Transition / transversion ratio can be a long way from 1:1

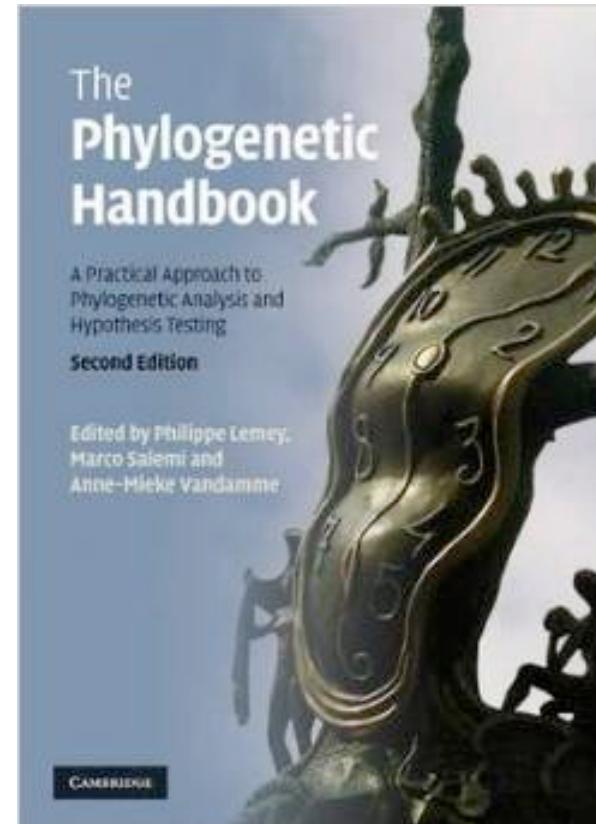
Phylogenetic models

Probability of a given tree

Given a specific initial state, and a mutation model, have a probability of producing a set of sequences at a future time t

Likelihood of a particular tree

Maximum likelihood tree



Influenza virus

Influenza virus

Severe but variable impact on human health

Pandemic influenza:

Arises from a species transfer event.

Little pre-existing immunity in the human population.

Previous events in 1918 (1977), 1957, 1968, 2009

1918 pandemic

Thought to be avian in origin

50-100 million deaths

(3-6% global population)

2009 epidemic ‘swine flu’

Origin in pigs

300-500 thousand deaths



Influenza virus

Influenza and human health

Seasonal influenza:

Continued lifespan of a pandemic strain

Constantly evolves to escape strain-specific immunity

Around 500 000 deaths each year

Strains currently circulating:

A/H1N1 strain : Arose from the 2009 pandemic.

A/H3N2 strain : Arose from the 1968 pandemic.

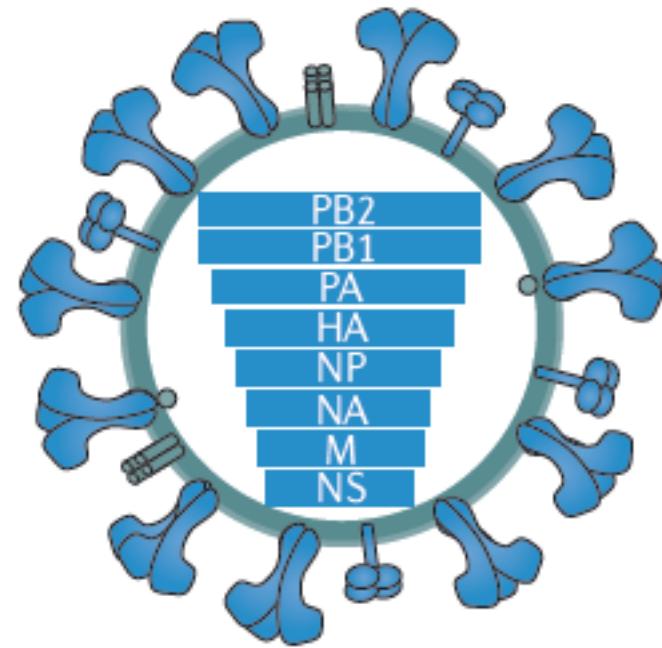
Influenza B : Less common, generally milder.

Influenza virus

Single-stranded, negative sense, RNA

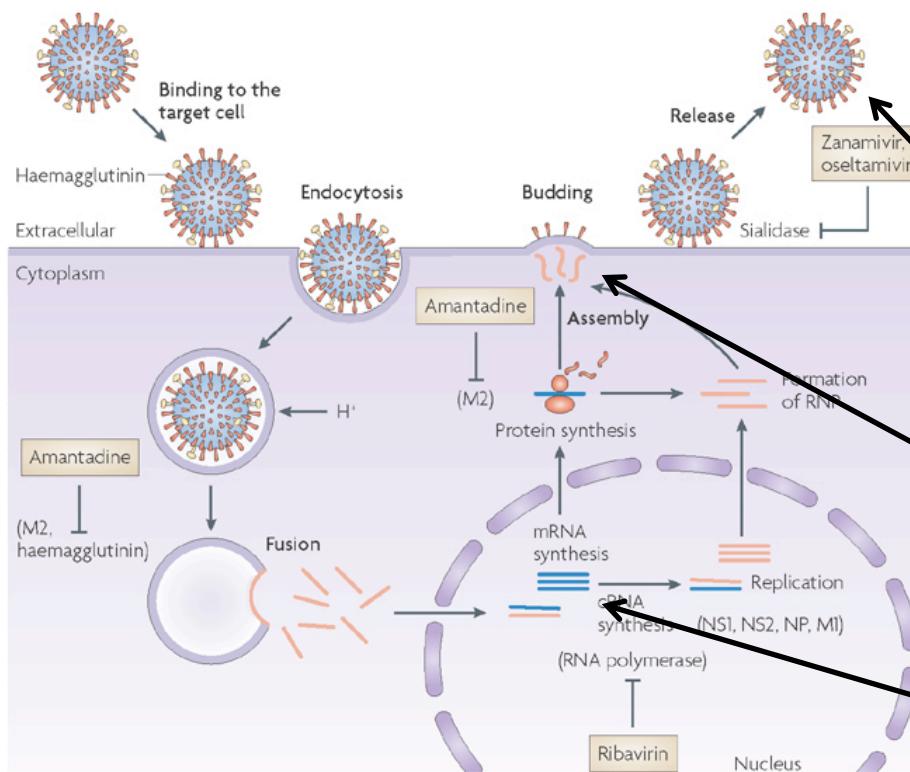
Eight segments

Multiple reading frames:
>12 proteins



Viral evolution

Life cycle of a virus



Viral evolution

Host immune system: Selection
Drug pressure

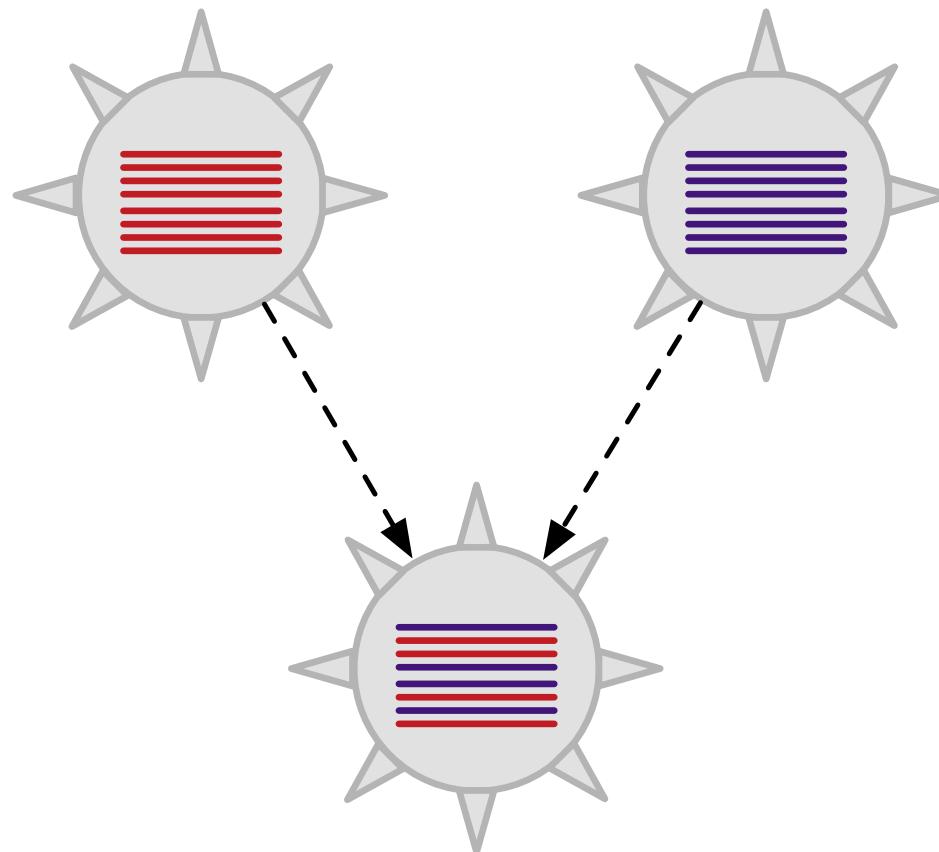
Viral assembly: Reassortment

Polymerase error: Mutation

Reassortment

Swapping of viral segments within a cell

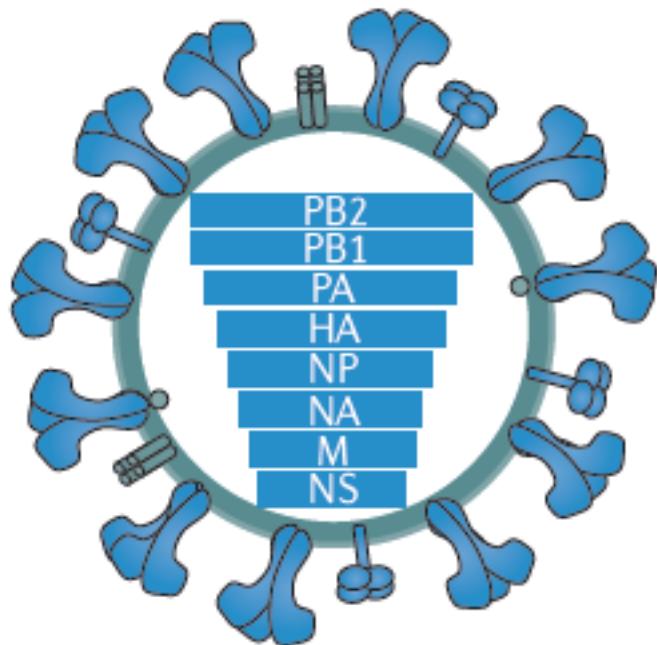
Segments are split then re-form at the cell membrane



Influenza virus

Viral genetics

Genes packaged within viral capsid



HA : Binds to host cell surface

NA : Promotes release from host cell surface

M: Part of capsid structure. Ion channel M2 allows H⁺ ions into capsid: uncoating of capsid

PB1, PB2, PA, NP : Involved in viral replication

PB1-F2, NS1 : Disrupt host immune system: interferon antagonists

Influenza virus

Multiple reading frames encode different proteins

RNA: ... CCU UAC AGC CAU GGG ACA GGA ACA GGA UAC ACC ...
 P P Y S H G T G T Y T ...

... CCU UAC AGC **CAU** **G**GG ACA GGA ACA GGA UAC ACC ...
P P Y S H G T G T Y T ...

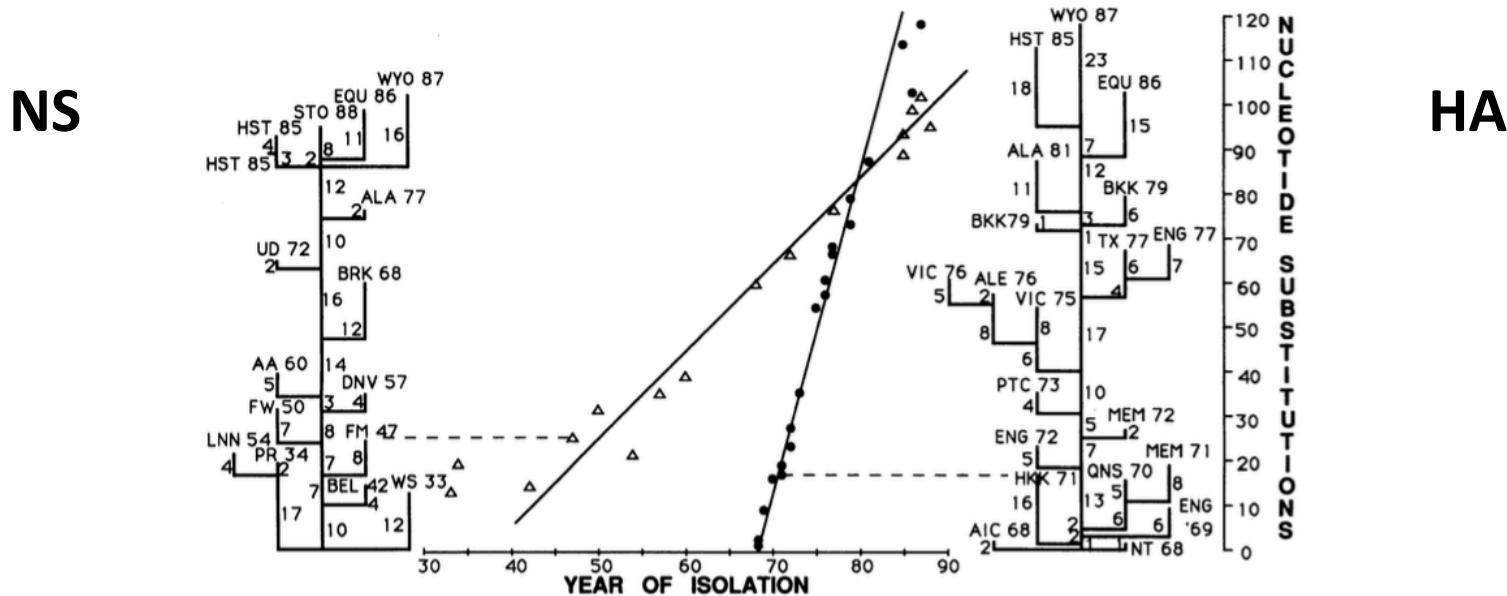
... C CUU ACA GCC **AUG** GGA CAG GAA CAG GAU ACA CC...
M G Q E Q D T P

Two proteins, PB1 and PB1-F2, produced from the same genetic sequence

Places some constraints on evolution...

Phylogenetic studies of influenza

Identification of Darwinian selection



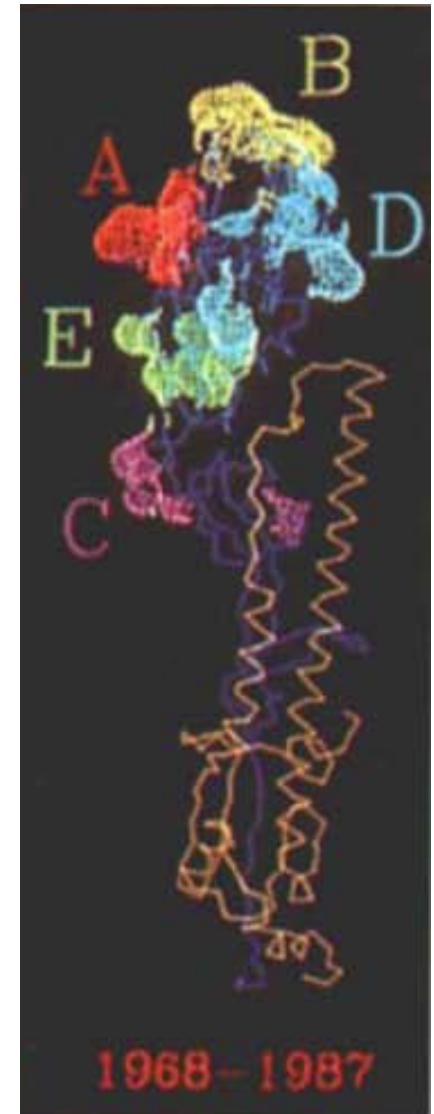
- 1: Shape of the tree suggests continual replacement of the strain
- 2: HA evolves faster than NS
- 3: Changes in HA are often at proposed antigenic sites.

Antigenic sites

Adaptive immunity

Changes in the viral genome between seasons cause changes in the antigenic properties of the virus.

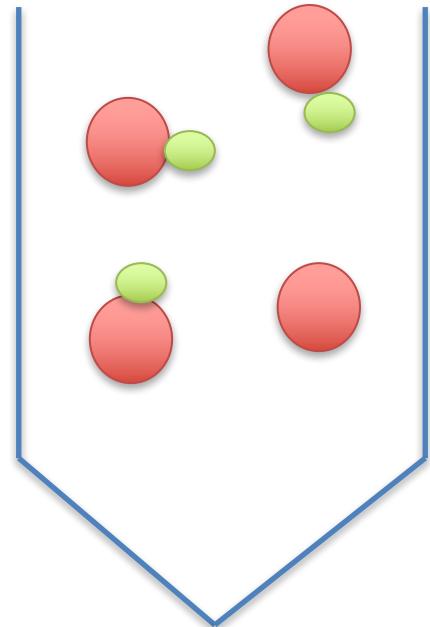
Antigenic changes enable the virus to re-infect people who previously were infected



Measuring vaccine effectiveness

Haemagglutination assay: Test viral concentration

Influenza virus binds red blood cells. Measure number of unbound cells

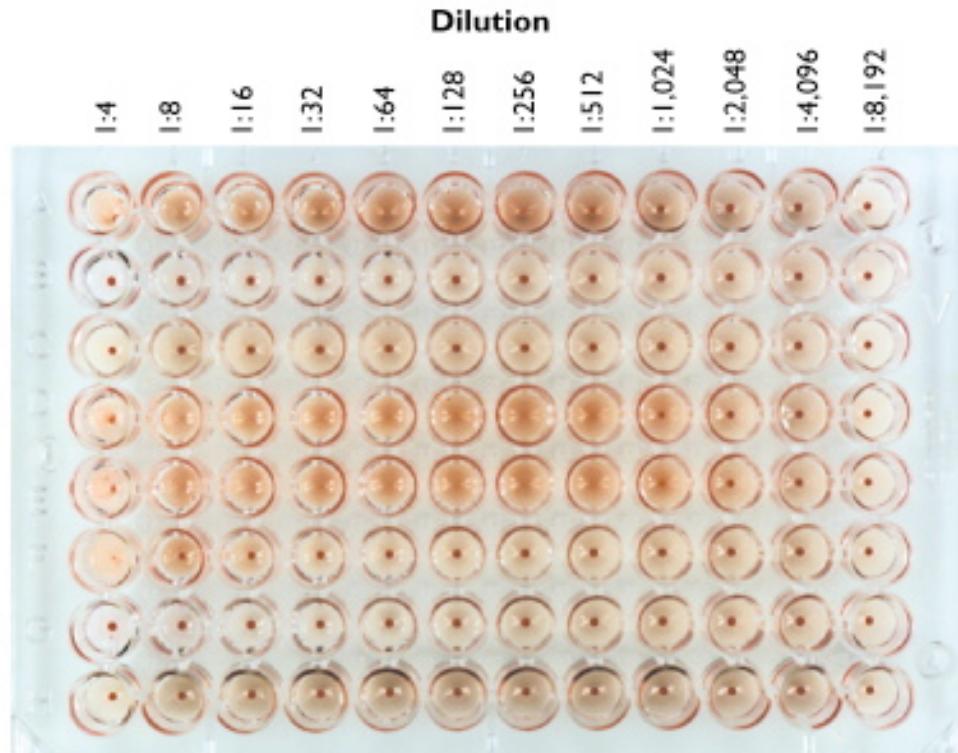


Cells at bottom of well indicate that not all were bound by the virus

Influenza virus

Haemagglutination assay

Influenza virus binds red blood cells. Measure number of unbound cells.



Row B: No binding : No detectable virus

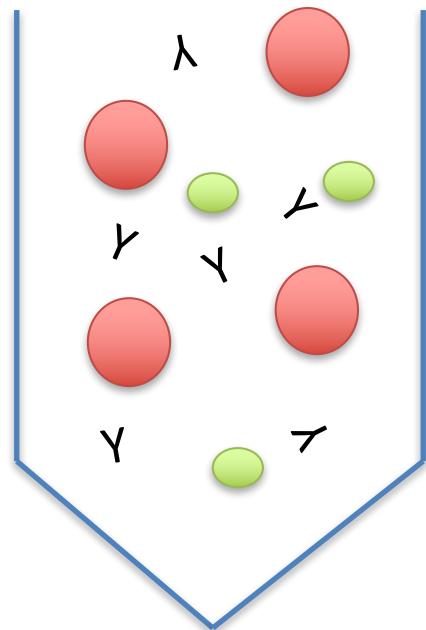
Row D: Binding up to 512-fold dilution of viral protein: lots of virus

Influenza virus

Haemagglutination inhibition assay: Include antibodies

Antibodies against the virus prevent the binding of cells

Test the effectiveness of the antibody



Cells at bottom of well indicate that not all were bound by the virus

Influenza virus

Haemagglutin inhibition assay

Raise antibodies against one virus, test against another

Viruses	Isolation	Haemagglutination inhibition titre ¹									
		Post infection ferret sera									
		A/NC 20/99	A/Eg 96/02	A/Neth 128/04	A/Theis 24/05	A/HK 2637/04	A/Eg 39/05	A/Vic 500/06	A/Mal 100/06	A/Taiw 42/05	A/Tok 6708/05
A/New Caledonia/20/99	9.6.99	Ex	320	320	640	320	1280	1280	640	640	80
A/Egypt/96/02 (H1N2)	25.1.02	Ex	320	640	640	320	640	640	320	320	40
A/Netherlands/128/04	14.9.04	MDCK1 \4	160	320	1280	640	1280	640	640	640	160
A/Thessaloniki/24/05	3.2.05	E2 \3	160	320	1280	640	1280	640	640	640	160
A/Hong Kong/2637/2004	24.6.04	MDCK2 \4	80	160	320	160	320	640	160	320	40
A/Egypt/39/2005	25.11.05	E1 \4	320	320	640	320	1280	1280	640	640	80
A/Victoria/500/2006	19.4.06	E3 \1	160	160	1280	640	640	640	1280	1280	320
A/Malaysia/100/2006	unknown	E3 \1	160	160	1280	640	1280	640	640	1280	160
A/Taiwan/42/2005	28.1.06	C4 \1	<	<	<	<	160	80	40	40	640
A/Tokyo/6708/2005	9.9.06	E1/1 \1	80	40	80	80	320	640	80	160	1280
											640

Identify changes in the antigenic properties of the virus

Influenza virus

Protein structural studies

Find the location of genetic changes that cause antigenic change

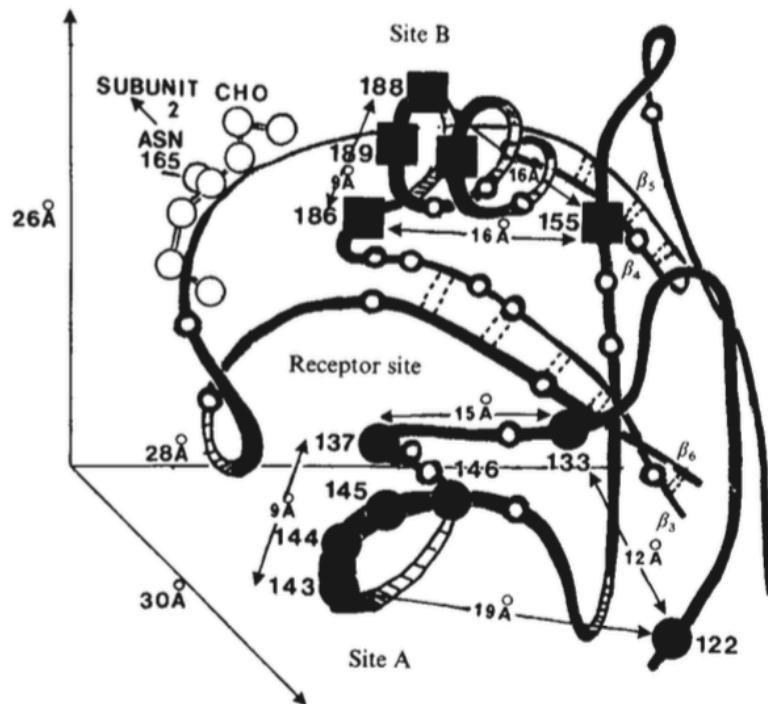
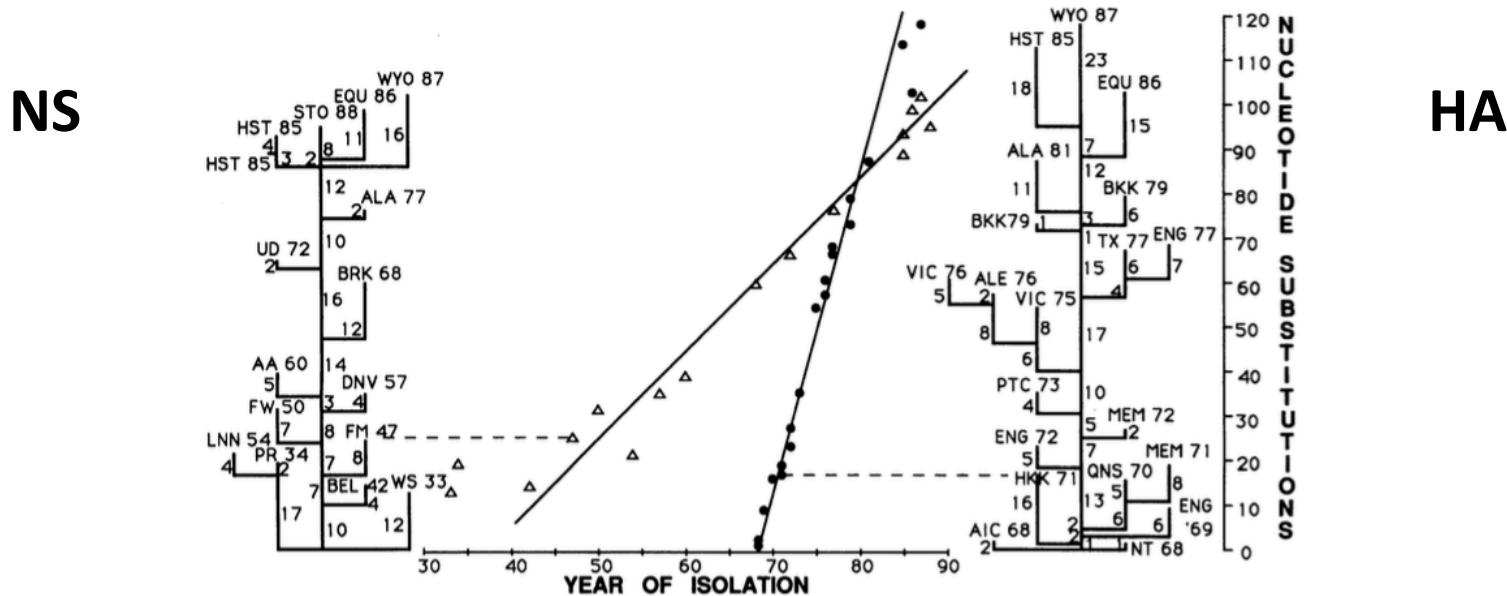


Fig. 3 Schematic drawing showing the size and shape of two of the proposed antigenic sites: ●, A and ■, B (see Fig. 2 and text). Their relationship to the pocket of conserved residues tentatively identified as the host sialic-acid-containing receptor binding site is shown.

Phylogenetic studies of influenza

Identification of Darwinian selection



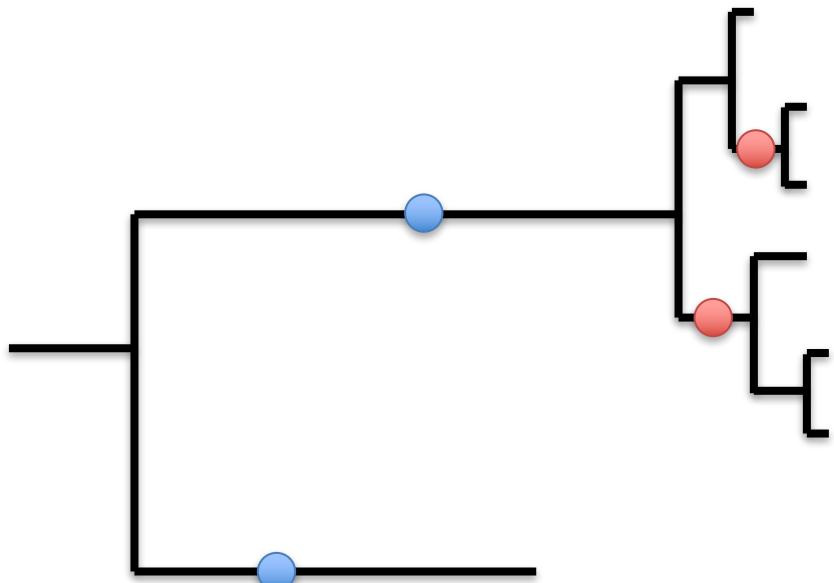
1. Shape of the tree suggests continual replacement of the strain
- 2: HA evolves faster than NS
- 3: Changes in HA are often at proposed antigenic sites.

Modelling selection in a phylogeny

dN/dS

Assumption that synonymous mutations are neutral.

Compare non-synonymous and synonymous mutation rates



Tree gives numbers of synonymous
and non-synonymous substitutions

Calculate opportunity for
synonymous and non-synonymous
mutations

Rate = #substitutions / opportunity

Modelling selection in a phylogeny

dN/dS

Interpretation

$dN/dS < 1$: Indicates purifying selection

Selection to keep the current set of amino acids
e.g. Functionally-important amino acids

$dN/dS > 1$: Indicates positive selection

Selection to replace amino acids
e.g. Time-dependent fitness

Can also identify variation in dN/dS in a subsection of sites within a sequence

Reality check : dN/dS is very rarely > 1 . Even under strong selection there may be very few potentially beneficial mutations

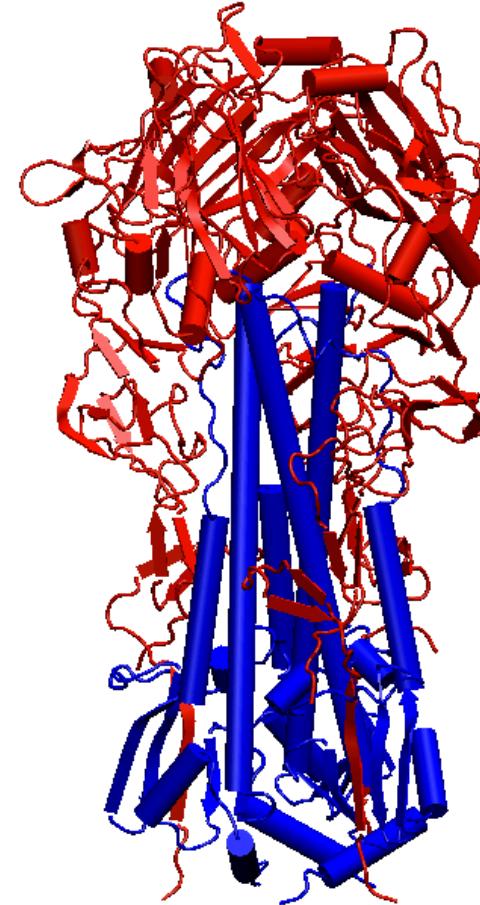
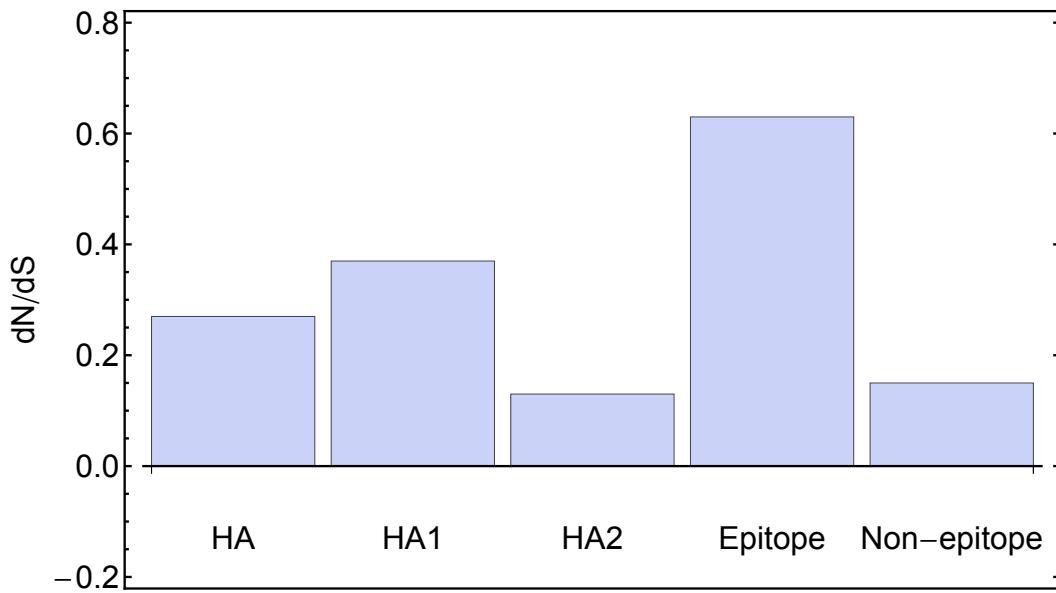
Selection in the influenza virus

Measurement via dN/dS

HA1 : Red

HA2: Blue

Epitope region known to interact with human immune system



Antigenic change is discrete

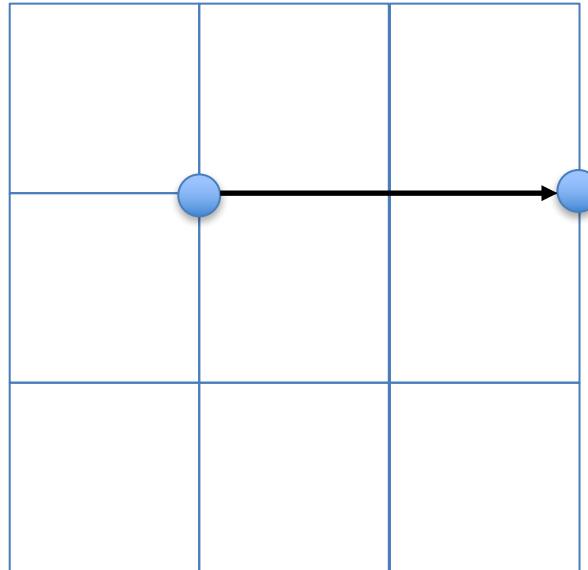
Antigenic data gives distances in ‘antigenic space’

Viruses	Isolation	Haemagglutination inhibition titre ¹										
		Post infection ferret sera										
		A/NC	A/Eg	A/Neth	A/Theis	A/HK	A/Eg	A/Vic	A/Mal	A/Taiw	A/Tok	
	Date	20/99	96/02	128/04	24/05	2637/04	39/05	500/06	100/06	42/05	6708/05	
A/New Caledonia/20/99	9.6.99	Ex	320	320	640	320	1280	1280	640	640	80	40
A/Egypt/96/02 (H1N2)	25.1.02	Ex	320	640	640	320	640	640	320	320	40	<
A/Netherlands/128/04	14.9.04	MDCK1/4	160	320	1280	640	1280	640	640	640	160	40
A/Thessaloniki/24/05	3.2.05	E2/3	160	320	1280	640	1280	640	640	640	160	40
A/Hong Kong/2637/2004	24.6.04	MDCK2/4	80	160	320	160	320	640	160	320	40	<
A/Egypt/39/2005	25.11.05	E1/4	320	320	640	320	1280	1280	640	640	80	40
A/Victoria/500/2006	19.4.06	E3/1	160	160	1280	640	640	640	1280	1280	320	40
A/Malaysia/100/2006	unknown	E3/1	160	160	1280	640	1280	640	640	1280	160	40
A/Taiwan/42/2005	28.1.06	C4/1	<	<	<	<	160	80	40	40	640	80
A/Tokyo/6708/2005	9.9.06	E1/1/1	80	40	80	80	320	640	80	160	1280	640

Antigenic change is discrete

Viral strain mapping

Concentrations describe an ‘antigenic distance’ between viruses



Measure many strain-strain distances

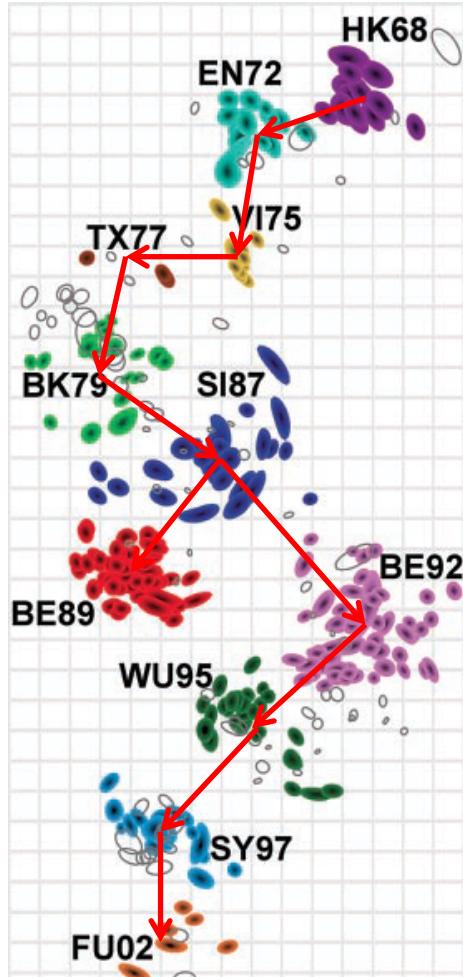
Project distances onto a two-dimensional surface

$$D = \begin{pmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & \ddots & & d_{2n} \\ \vdots & & \ddots & \vdots \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{pmatrix} \quad D_{\mathcal{R}^2} = \begin{pmatrix} d'_{11} & d'_{12} & \dots & d'_{1n} \\ d'_{21} & \ddots & & d'_{2n} \\ \vdots & & \ddots & \vdots \\ d'_{n1} & d'_{n2} & \dots & d'_{nn} \end{pmatrix}$$

Aim to reproduce the ‘antigenic distances’ as distances on the page

Antigenic change is discrete

Viral strain mapping



Observe “antigenic clusters”

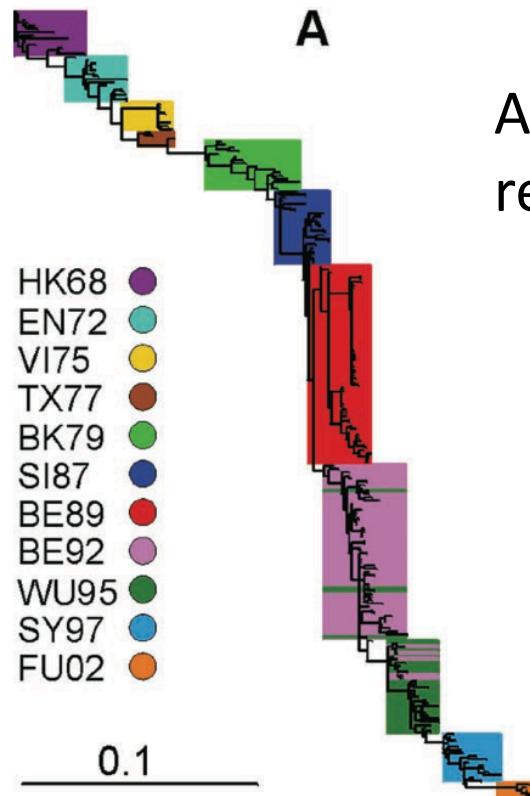
One square ≈ two-fold reduction
in effectiveness of vaccine (assay)

Discrete jumps occur in the
antigenic properties of the virus

Antigenic change is discrete

Viral strain mapping

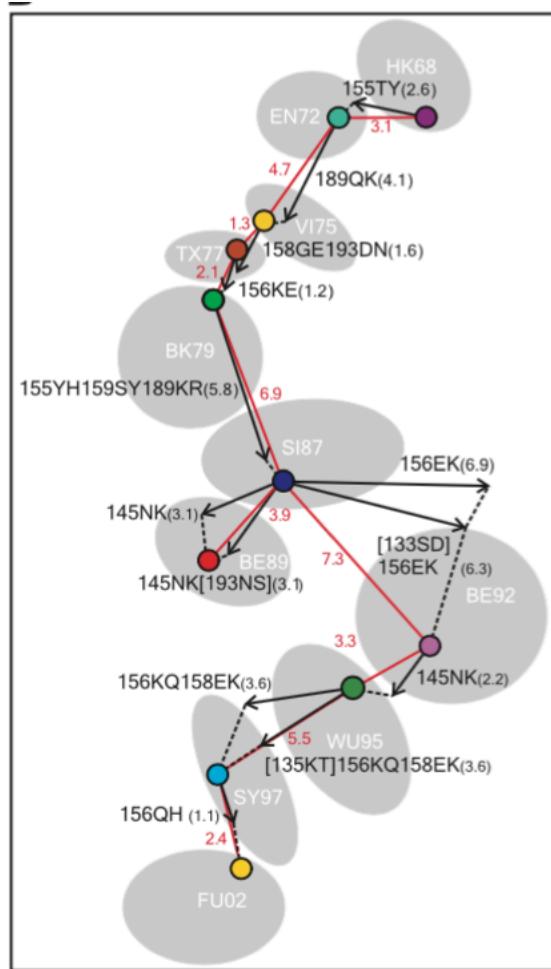
Phylogeny of influenza (H3N2)



Antigenic changes correspond quite well to
replacements of the global population

Antigenic change is discrete

Small numbers of changes underlie antigenic change



Multiple genetic changes occur between antigenic jumps, but only a few of these are required for antigenic change

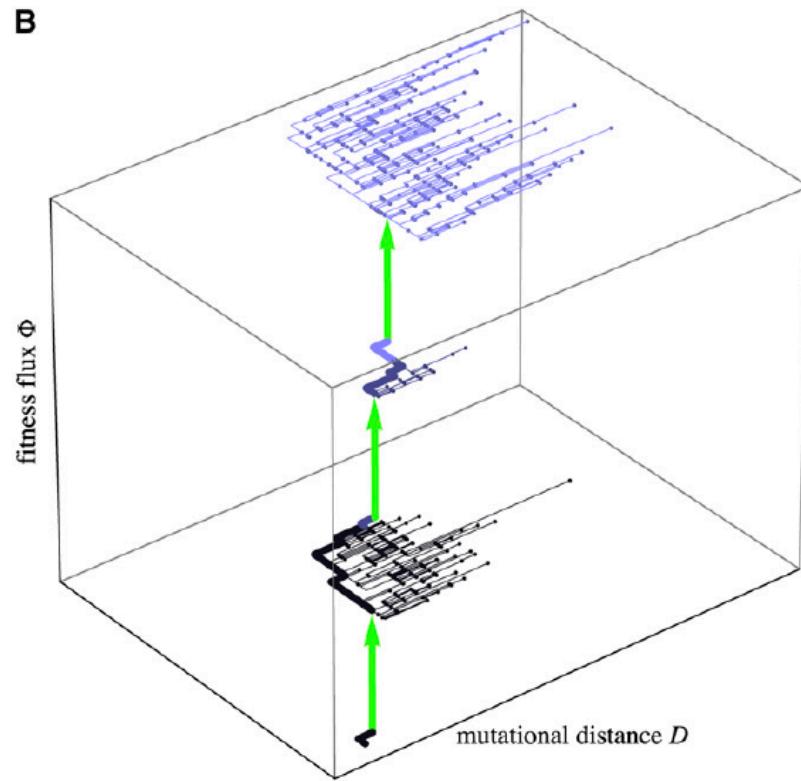
Why are there more substitutions than needed for antigenic change?

Why don't substitutions happen more quickly?

Clonal interference and influenza virus

Model 1: Occasional sweeps

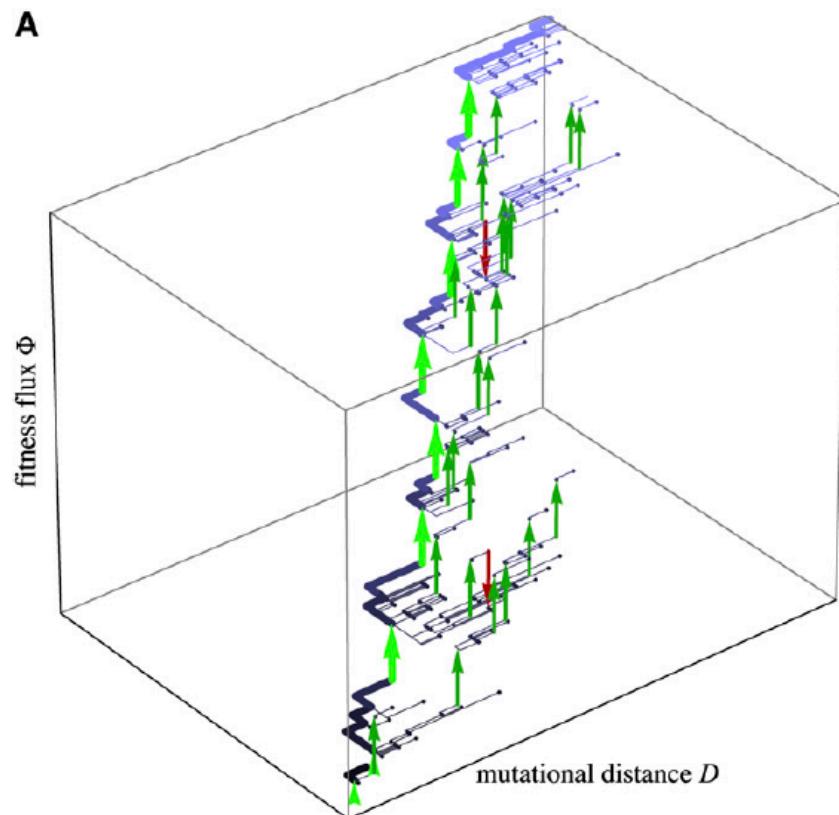
Largely neutral evolution is interrupted by occasional, strongly selected mutations



Clonal interference and influenza virus

Model 2: Clonal interference

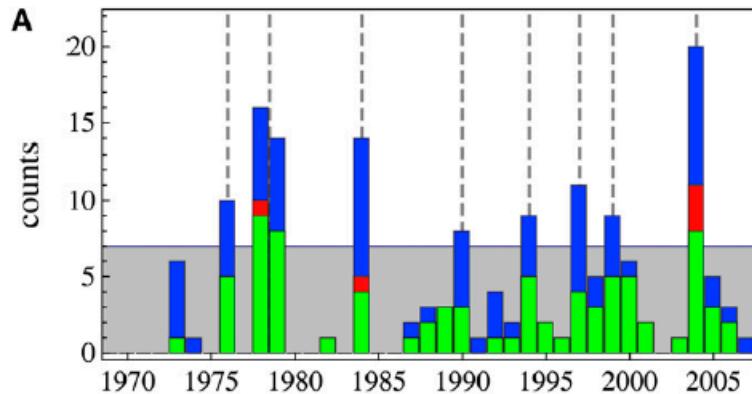
Multiple competing beneficial mutations are present at the same time



Clonal interference and influenza virus

Evidence for selective sweeps

Most fixations occur in clusters, with many in a single year



NS Epitope substitutions

NS non-epitope substitutions

Synonymous substitutions

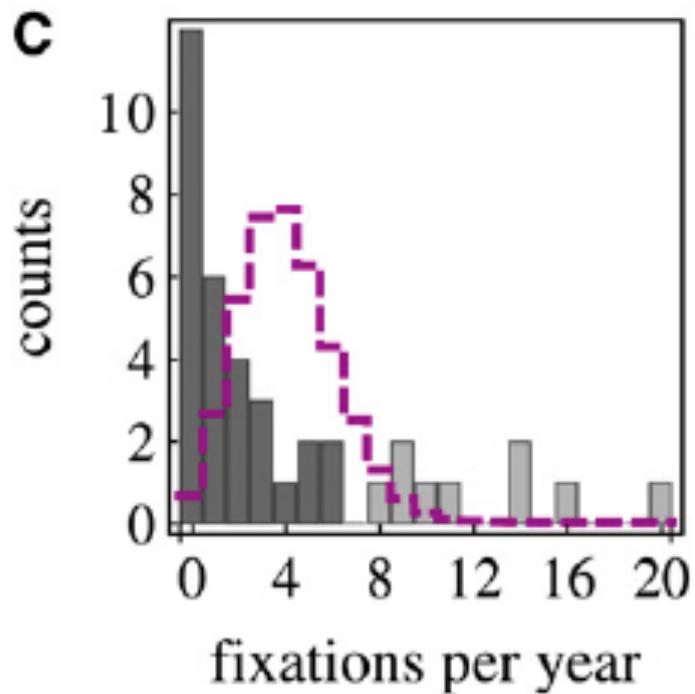
More than expected NS epitope substitutions: excess of around 45 mutations

Mean fixation time c. 3 years implies multiple substitutions in population at once

Clonal interference and influenza virus

Evidence for selective sweeps

Inhomogeneous process of fixations



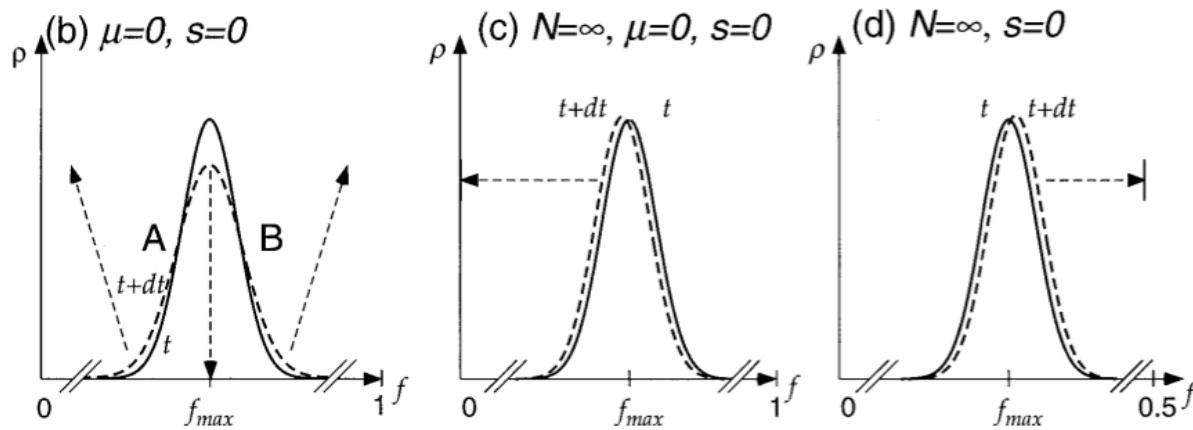
Non-Poisson distribution of number of fixations per year

Fixation events are not independent of one another

Mutational load and influenza virus

Muller's ratchet: Accumulation of deleterious mutations

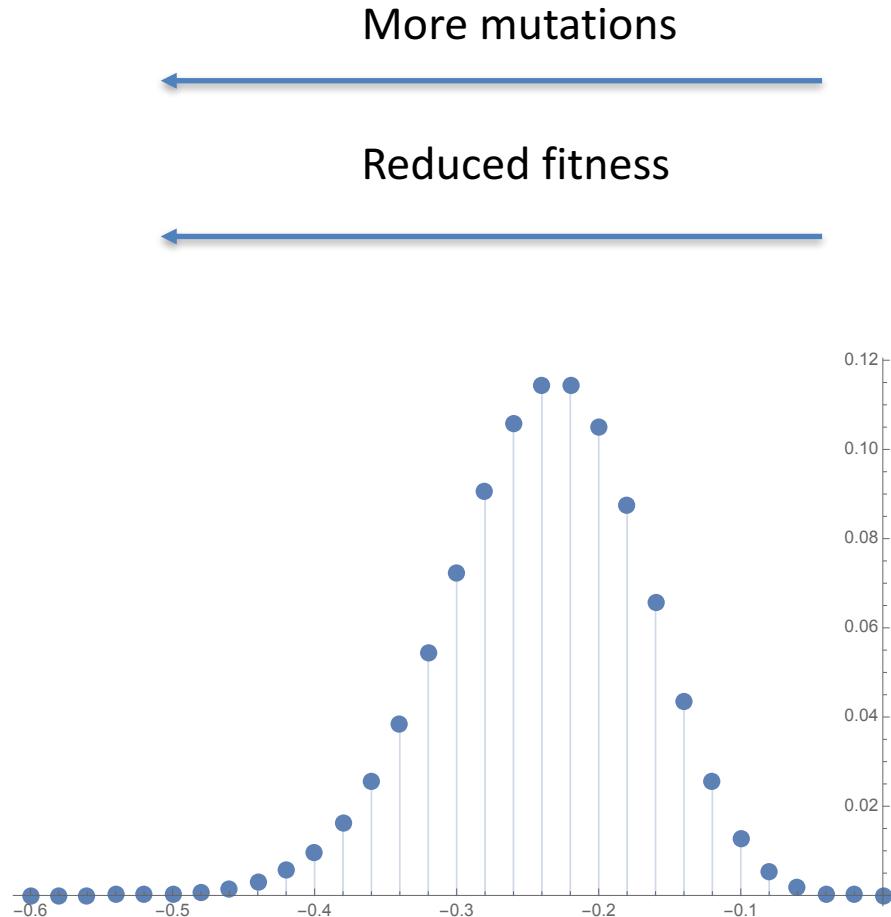
Mutations removed by selection: Mutation-selection balance



Mutation increases variant frequencies, selection returns them to zero

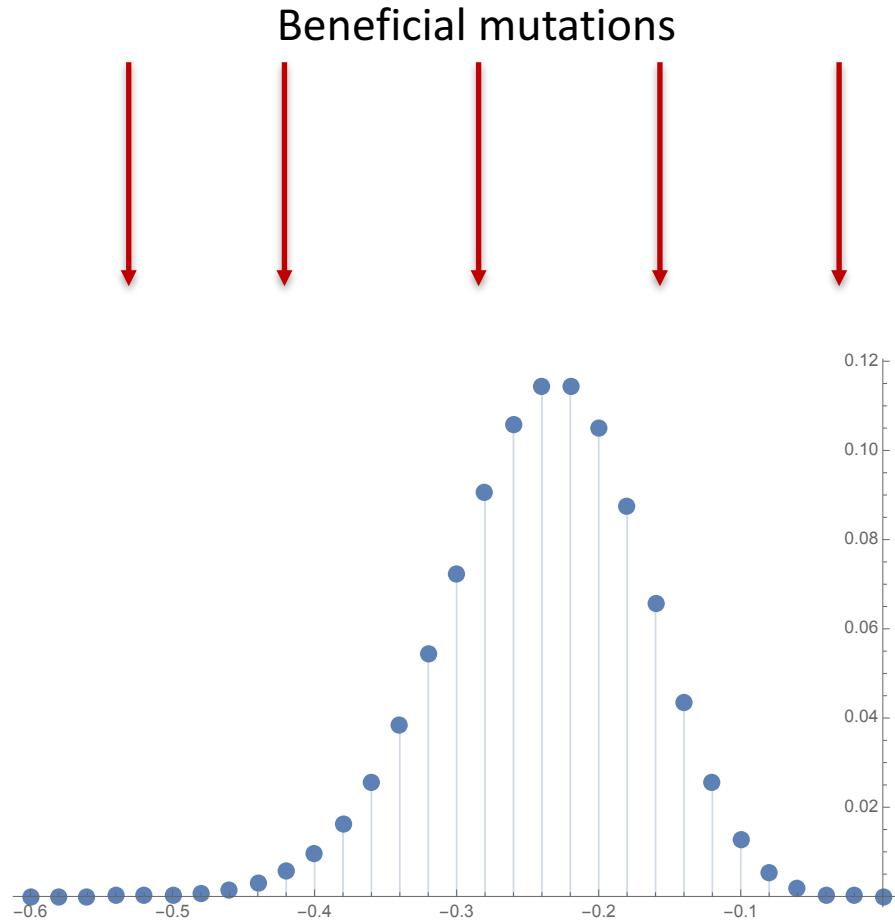
Mutational load and influenza virus

Distribution of relative fitness arising from mutational load



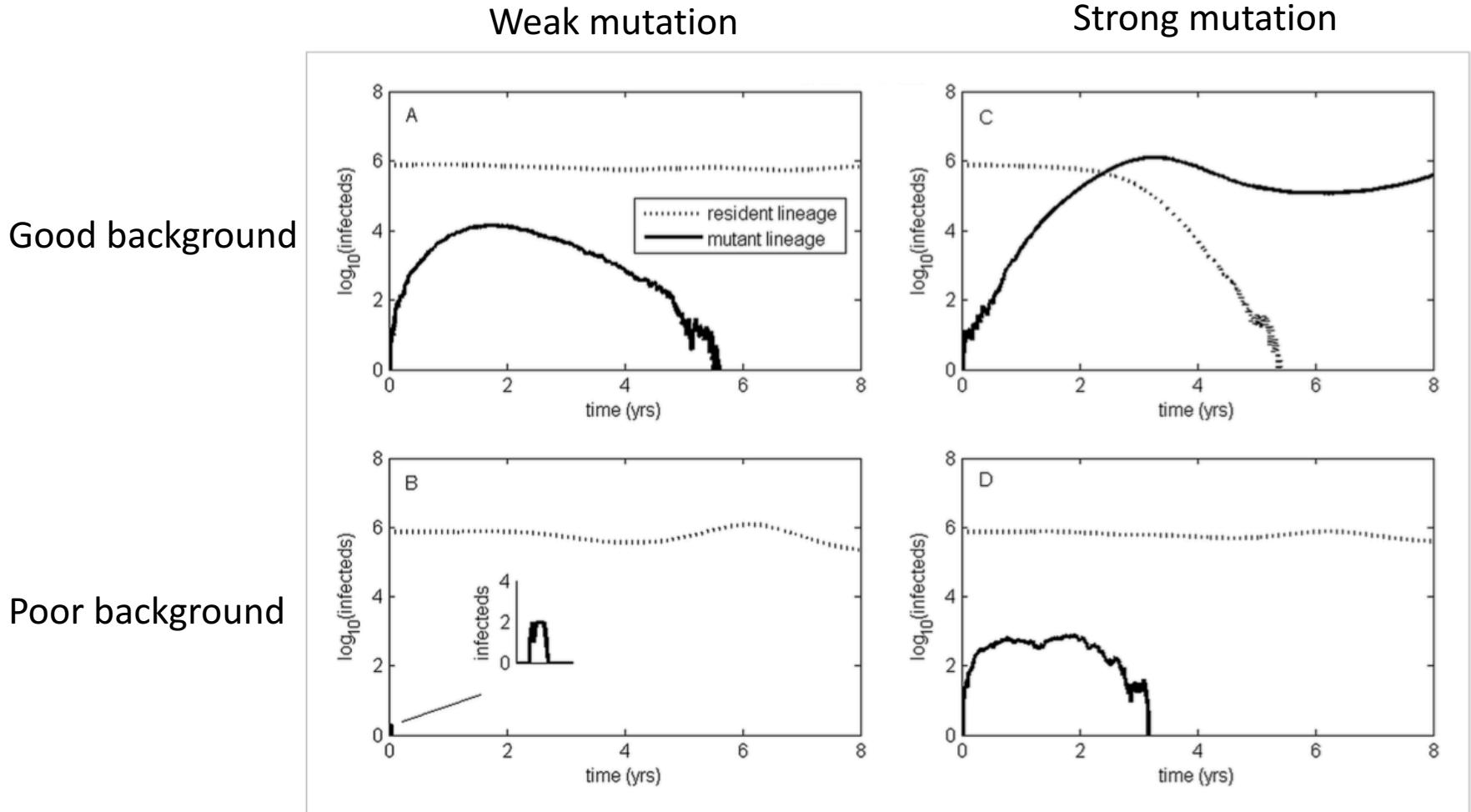
Mutational load and influenza virus

Distribution of relative fitness arising from mutational load



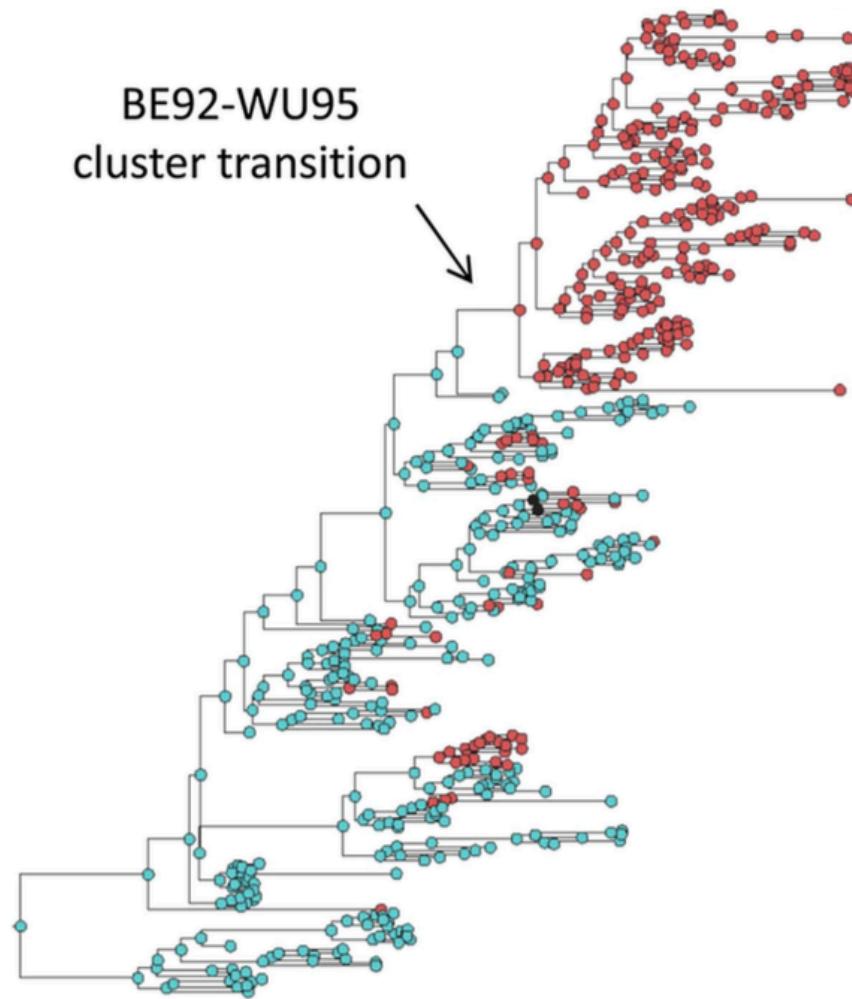
Mutational load and influenza virus

Fate of beneficial mutations



Mutational load and influenza virus

Example: N145K mutation



Summary

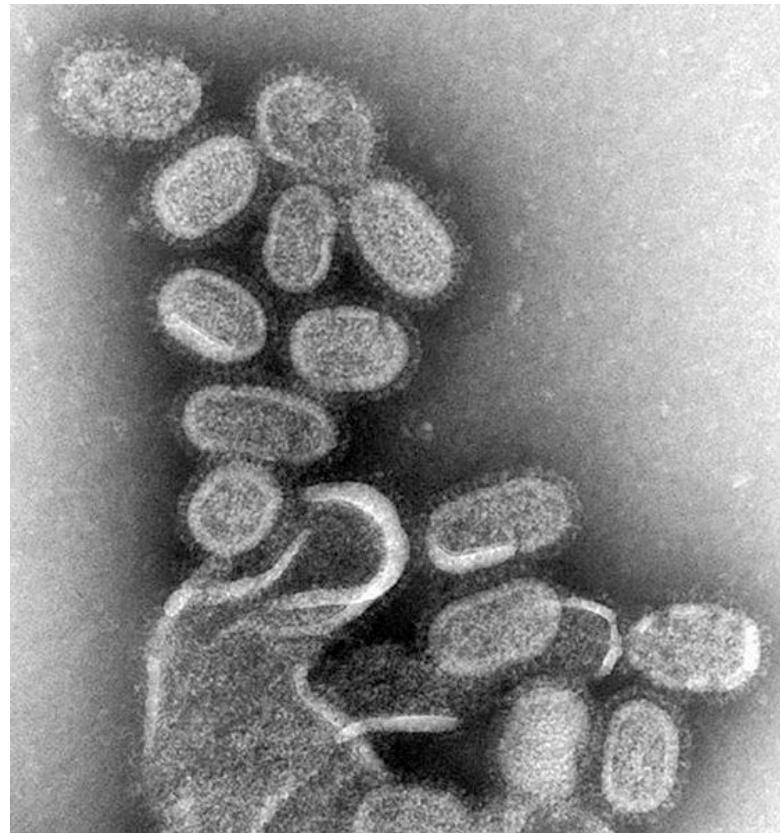
Long history of models describing influenza evolution

Rapid adaptation under strong selection; immune pressure

Regular strain replacement

Interesting evolutionary system + severe consequences for human health

Forecasting evolution



Forces and evolutionary change

Time for an object to fall under gravity (Galileo, 1589)



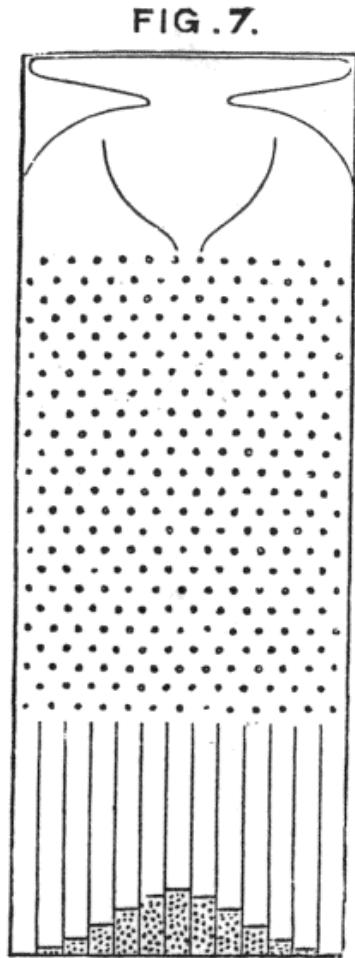
Equation of motion:

$$\frac{d^2x}{dt} = g$$

Acceleration 9.8 ms^{-2}

Predictability of evolution

Galton's bean machine (Francis Galton, 1889)

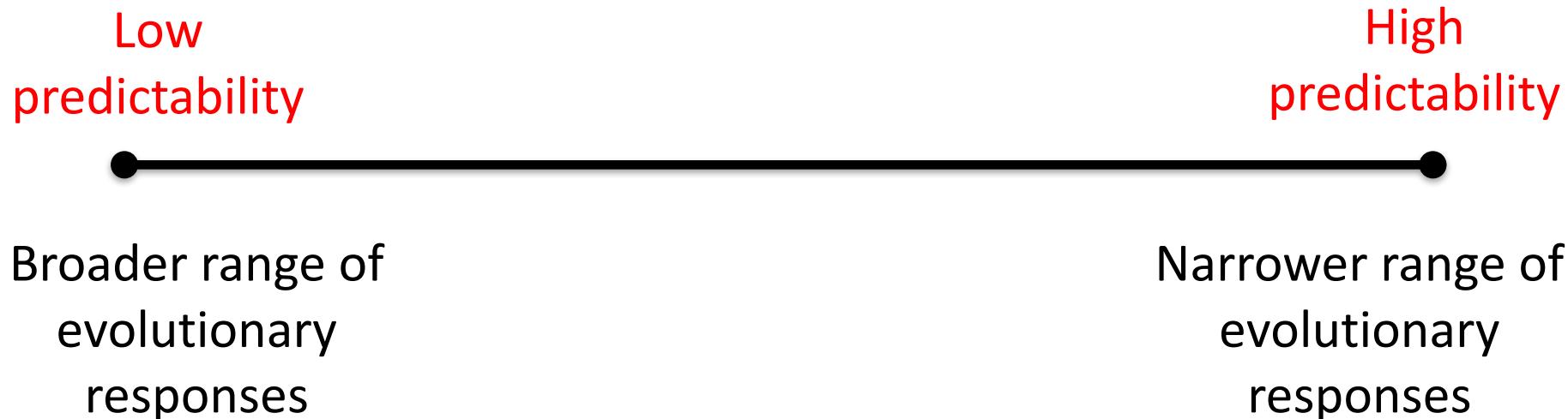


Stochastic
process

Outcome approximates a
binomial distribution

Predictability of evolution

Genetic change driven by natural selection



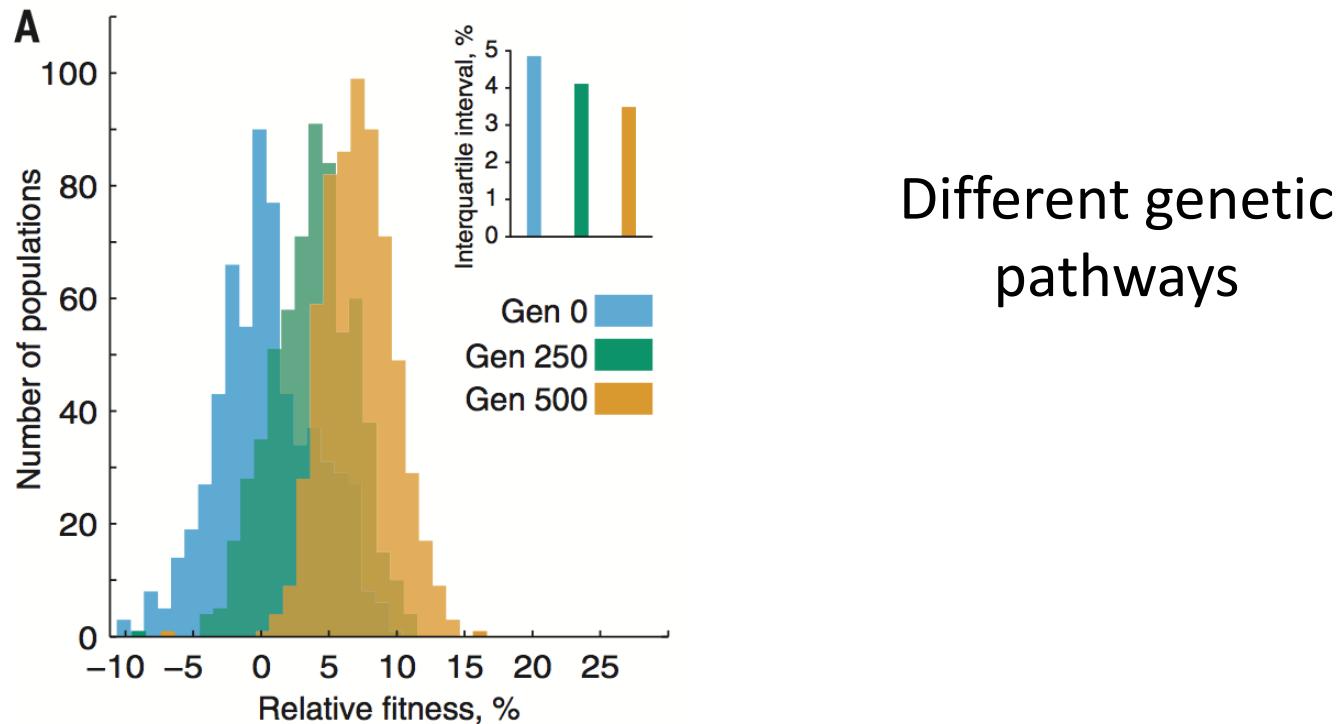
Example: Yeast given environmental change
Kryazhimskiy et al, 2014

Example: Bacteria given extreme drug pressure
Toprak et al, 2011

Genetic predictability of evolution

Global epistasis makes adaptation predictable despite sequence-level stochasticity

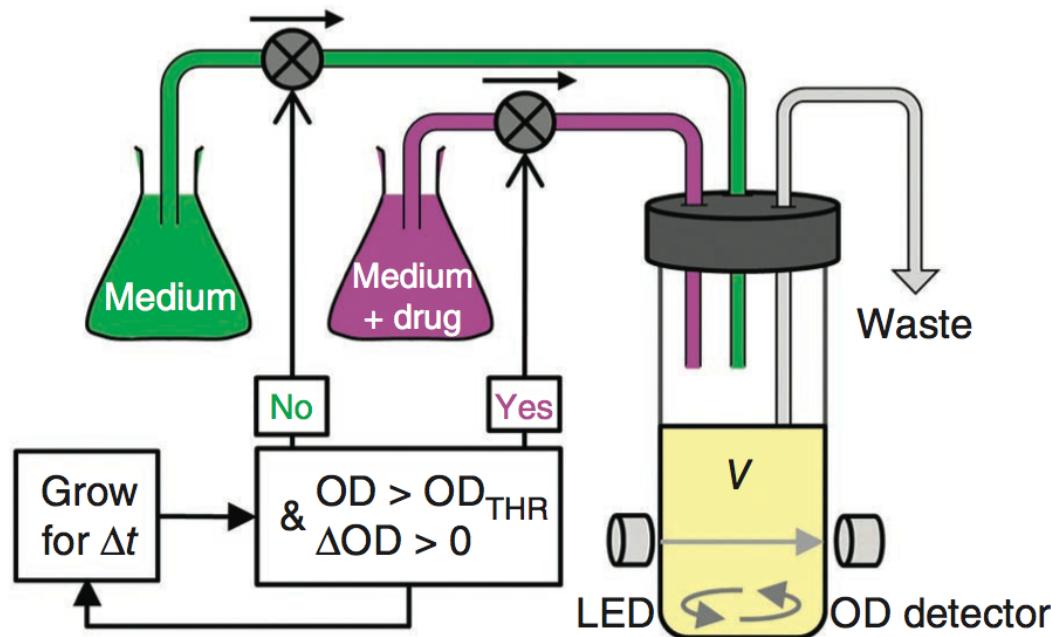
Sergey Kryazhimskiy,^{1,3*}† Daniel P. Rice,^{1,3*} Elizabeth R. Jerison,^{2,3} Michael M. Desai^{1,2,3†}



Genetic predictability of evolution

Evolutionary paths to antibiotic resistance under dynamically sustained drug selection

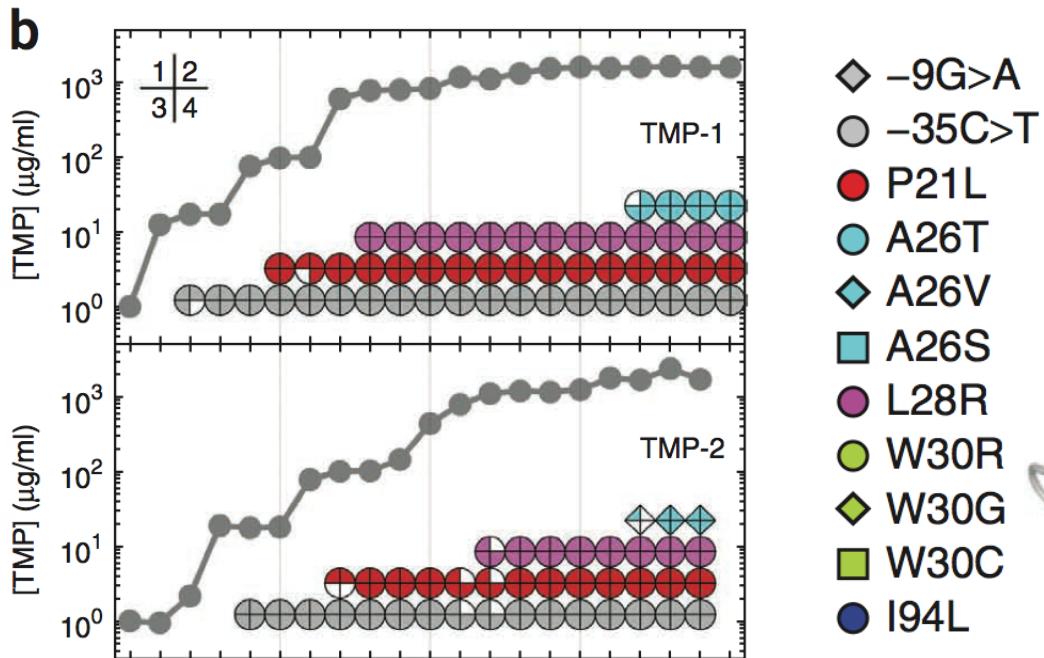
Erdal Toprak^{1,6}, Adrian Veres^{2,6}, Jean-Baptiste Michel^{1,3}, Remy Chait¹, Daniel L Hartl⁴ & Roy Kishony^{1,5}



Genetic predictability of evolution

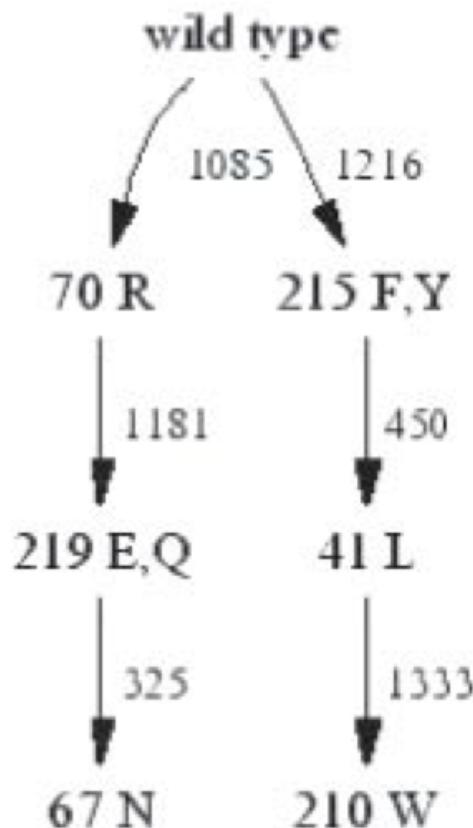
Evolutionary paths to antibiotic resistance under dynamically sustained drug selection

Erdal Toprak^{1,6}, Adrian Veres^{2,6}, Jean-Baptiste Michel^{1,3}, Remy Chait¹, Daniel L Hartl⁴ & Roy Kishony^{1,5}



Predictability of evolution

HIV adaptation: repeated pathways to drug resistance



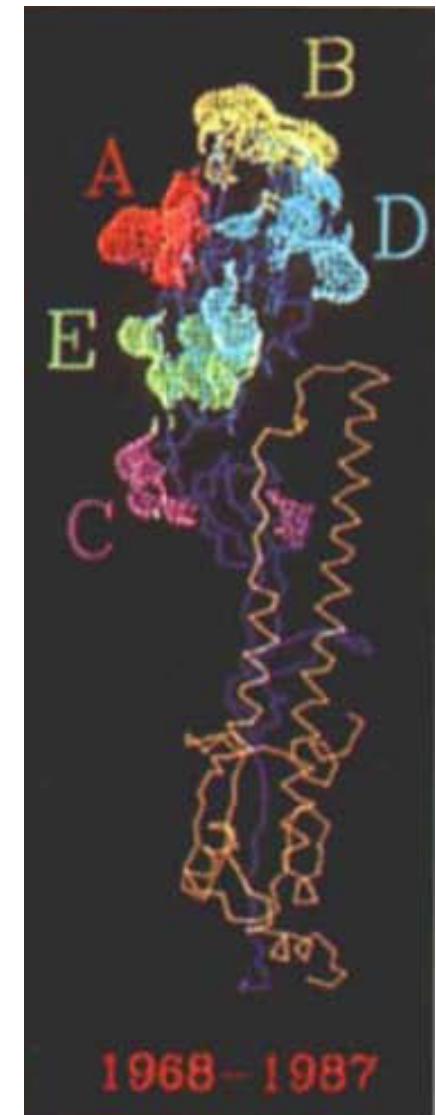
Immunity and vaccination

**Human immunity to influenza is long-lived,
but strain-specific**

Vaccination provides immunity to the current strain but vaccines need updating each year

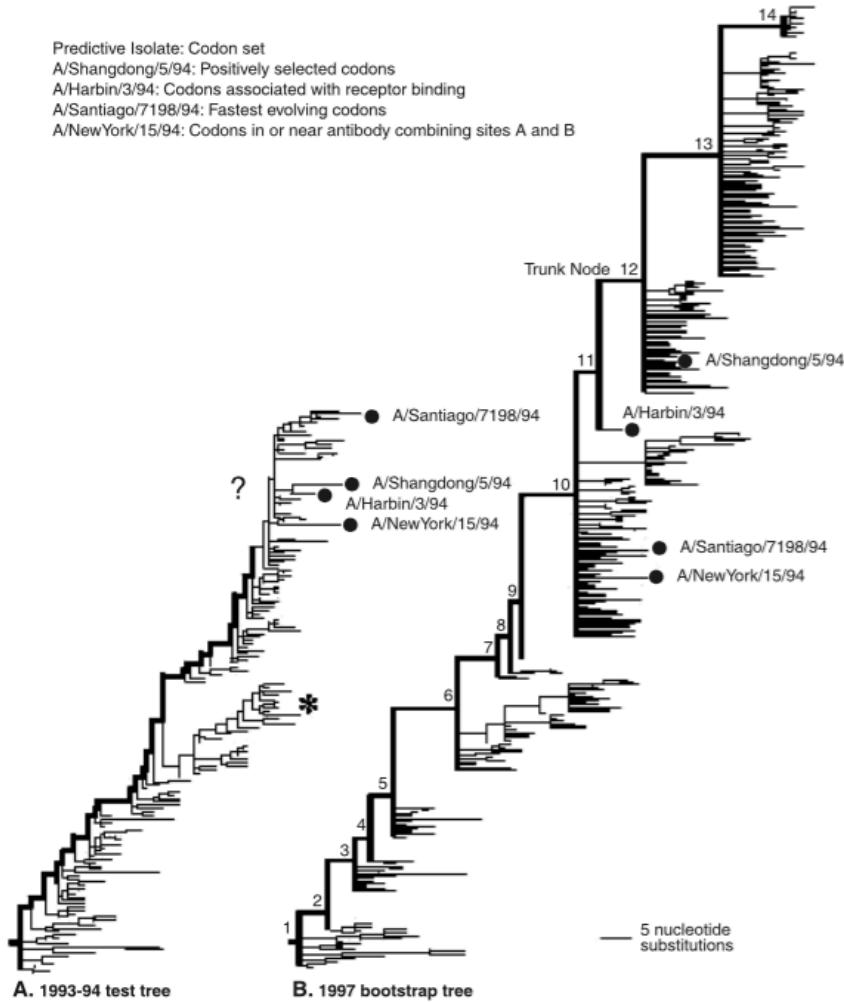
Vaccines require several months for production

Can the next strain be predicted?



Prediction from sequence data

Search for positive selection in the influenza genome

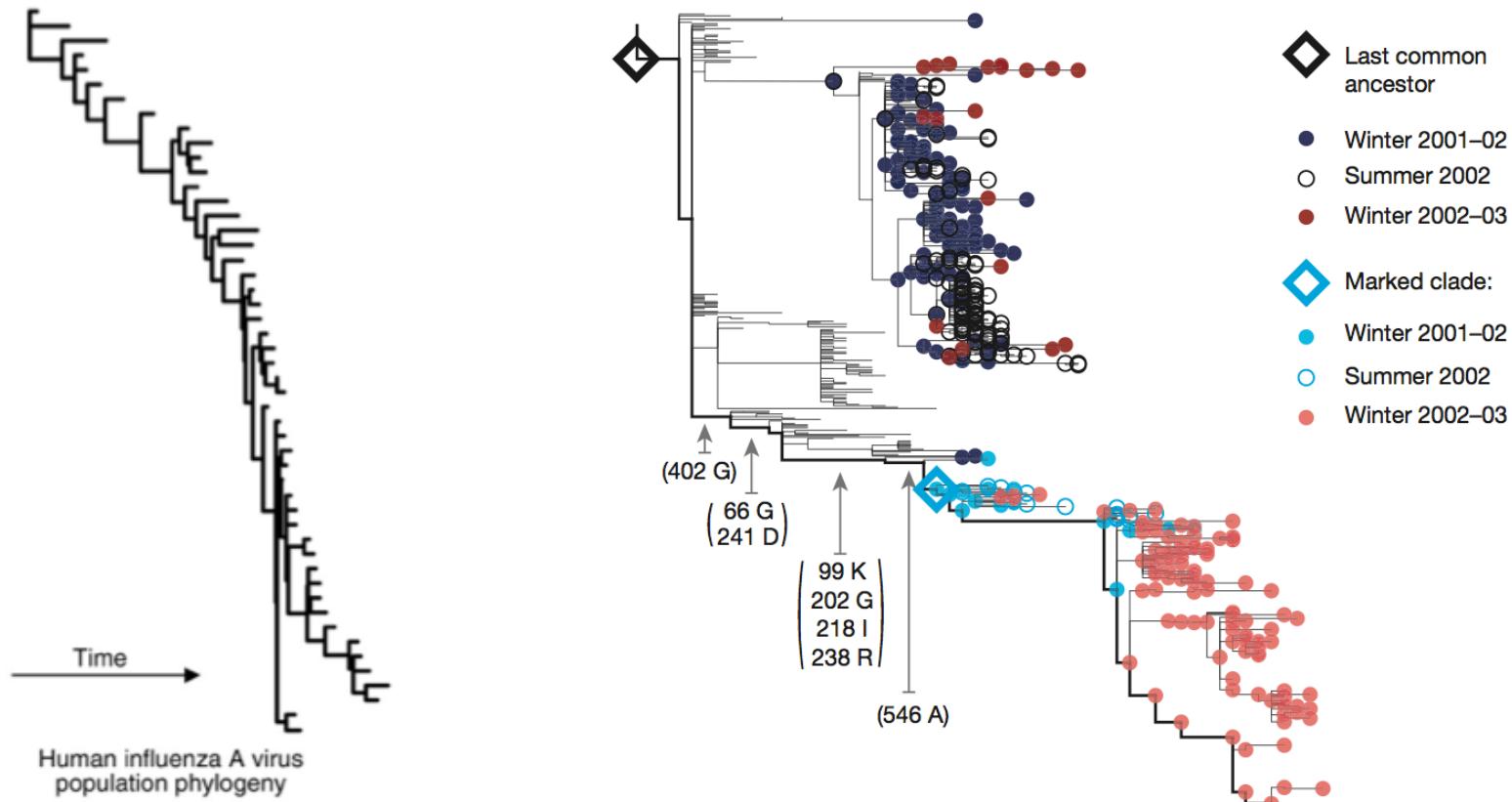


Identify positive selection at 18 sites in the HA gene

Proposition: Can predict the next strain by the sequence that has the greatest number of changes in these amino acid positions

Phylogeny of influenza

Competition between clades



Phylogeny of influenza

Growth of clade based upon strain fitness

$$X_\nu(t) = \sum_i x_i$$

Clade frequency: Sum of frequencies of strains in a clade

$$\hat{X}_\nu(t+1) = \sum_i x_i \exp(f_i)$$

Predicted clade frequency: proportional to exponential growth by fitness

Mean calculated over multiple phylogenetic trees

Note: Correct inference of fitness implies predictive knowledge of short-term evolution

Phylogeny of influenza

Strain fitness function:

$$f_i = f_0 - \mathcal{L}(\mathbf{a}_i) - \sum_{j: t_j < t_i} x_j \mathcal{C}(\mathbf{a}_i, \mathbf{a}_j)$$

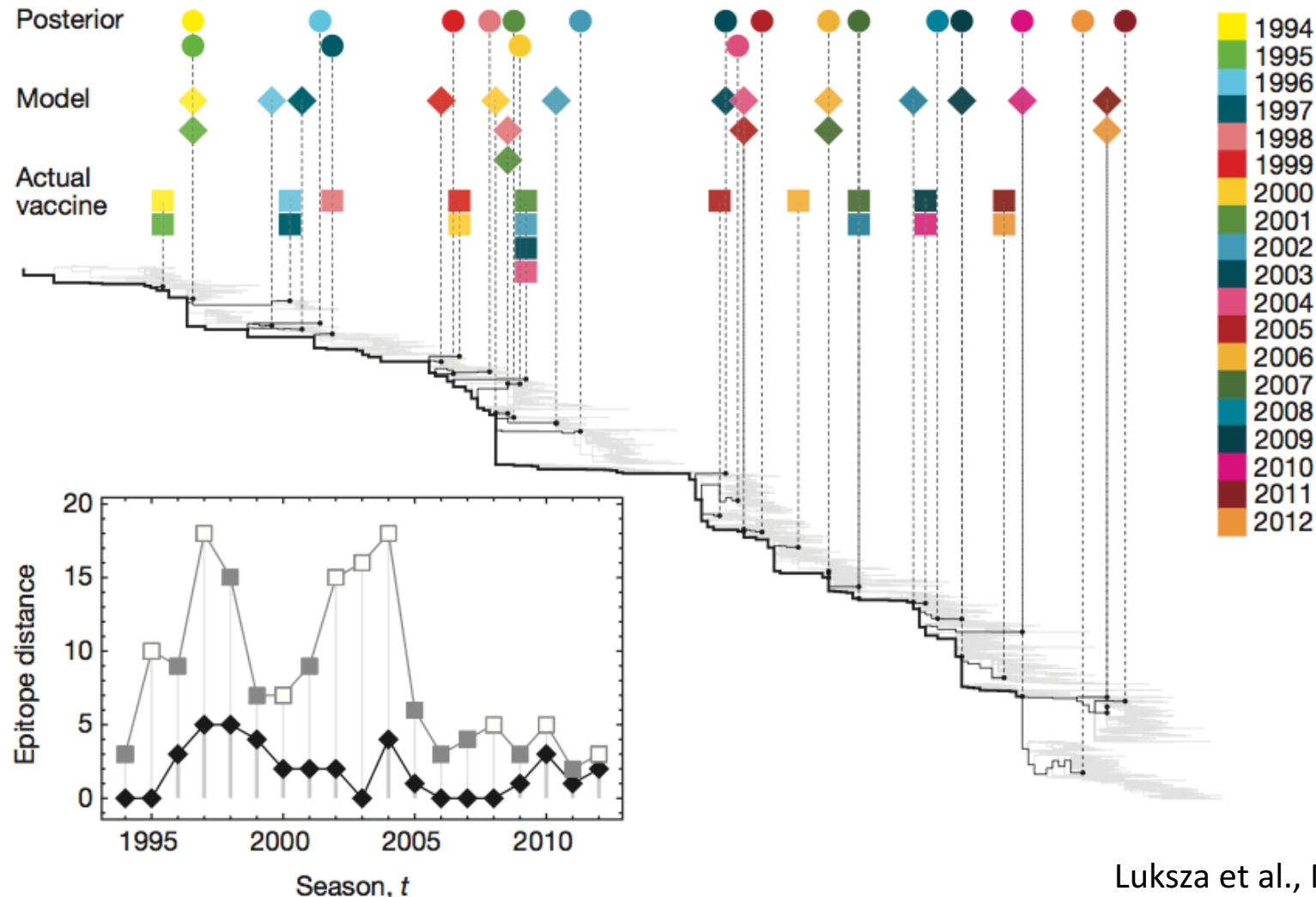
f_0 Base fitness: normalisation

$\mathcal{L}(\mathbf{a}_i)$ Cost of non-synonymous non-epitope mutations

$\mathcal{C}(\mathbf{a}_i, \mathbf{a}_j)$ Cross-immunity: Based upon past existence of strain
Non-synonymous epitope mutations increase distance
and decrease cost of cross-immunity

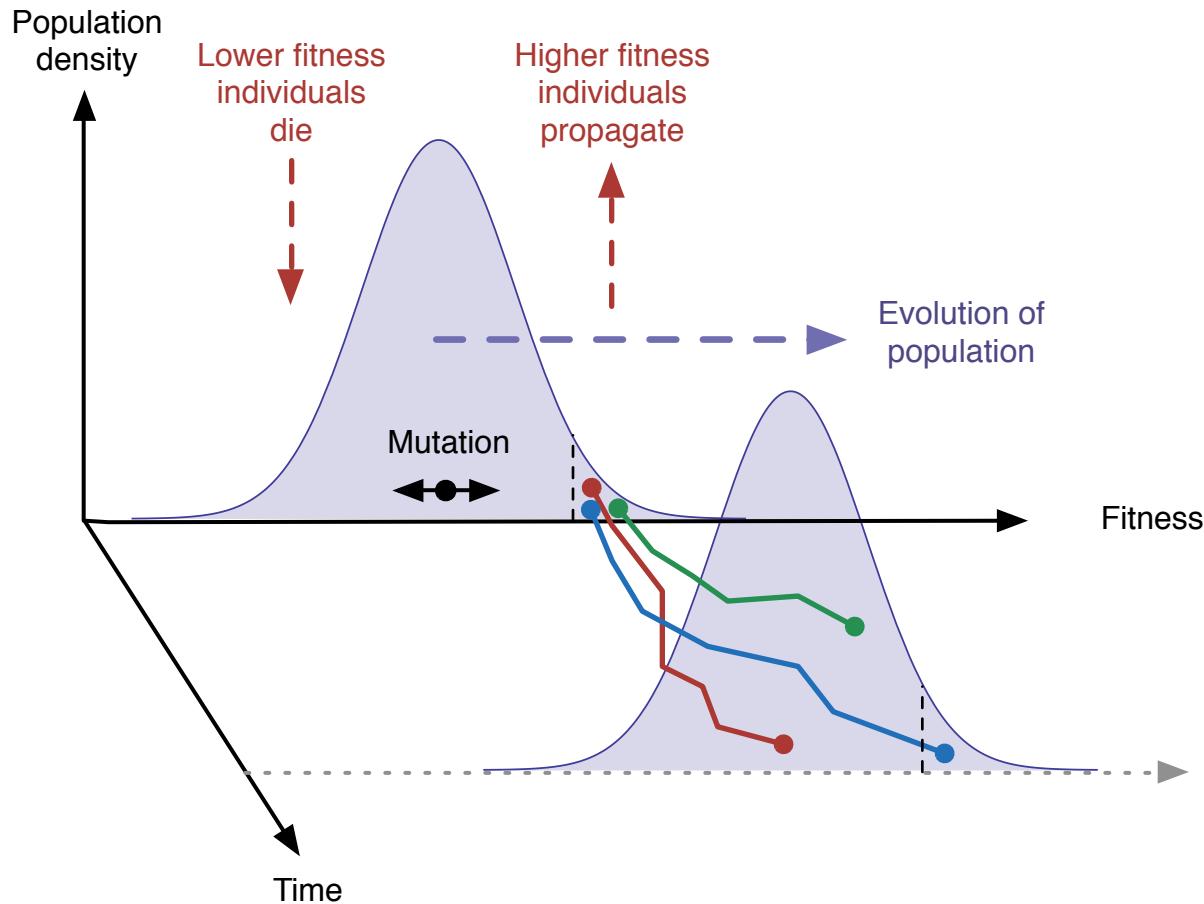
Predictive efficacy

Predicted versus actual vaccine strain



Prediction from phylogenetic shape

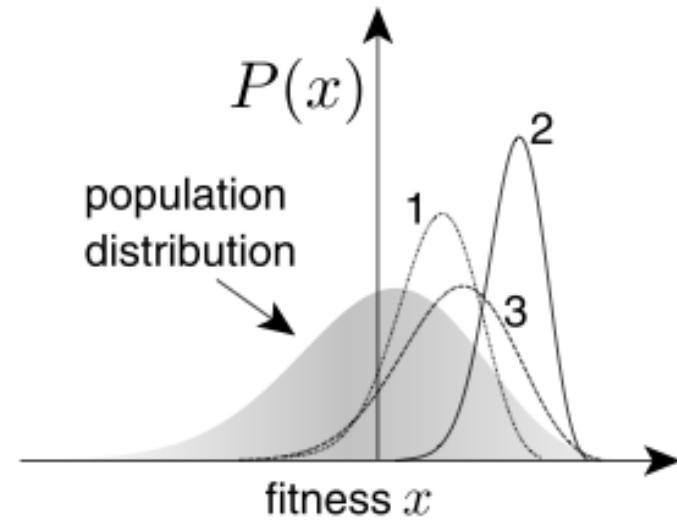
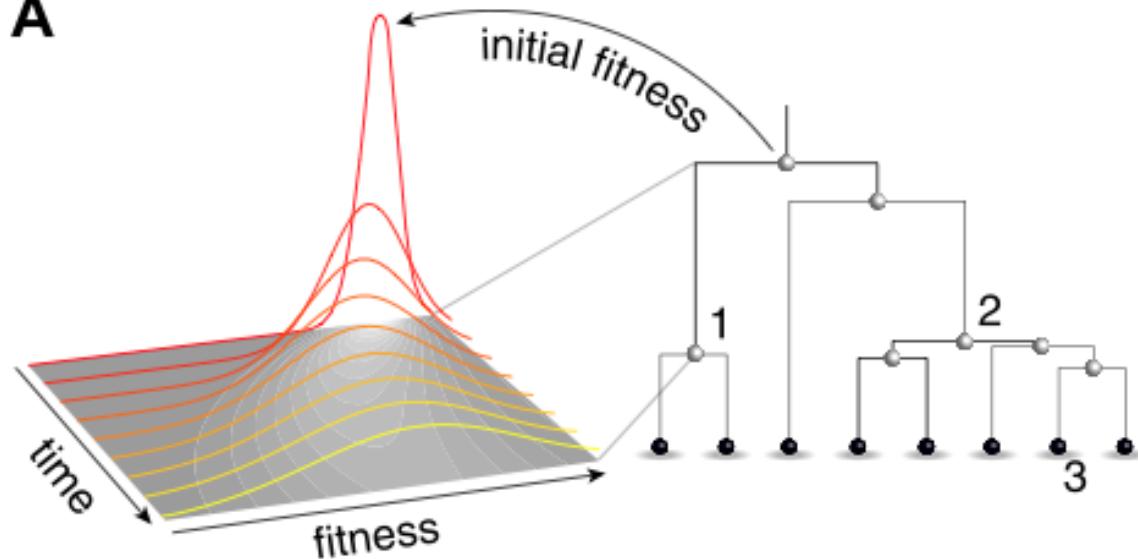
Inflow of beneficial mutations: Travelling wave model



Prediction from phylogenetic shape

Inflow of beneficial mutations: Travelling wave model

A



Greater number of offspring suggests higher fitness:
Offspring more likely to continue to arise from this clade

Methods of prediction

Methods of prediction

Phenotype-based prediction

Look for change in antigenic type

Other phenotypes e.g. protein stability can dictate fitness

Genotype-based prediction

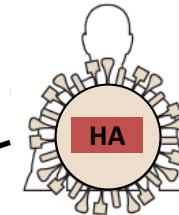
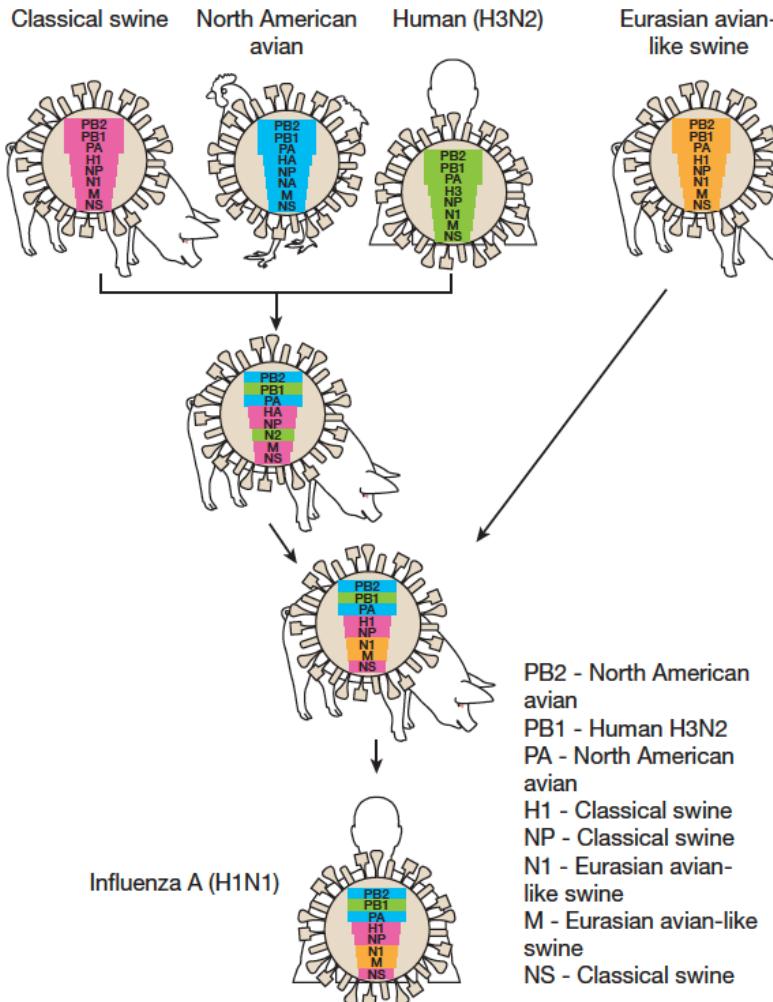
Genome contains additional information

Harder to recover information from genome – obscure encoding

Predicting the next influenza pandemic

Reassortment?

Origin of the 2009 pandemic

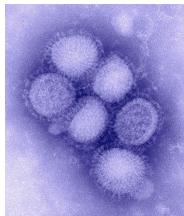
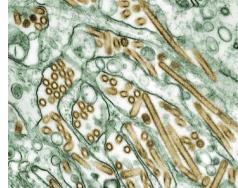
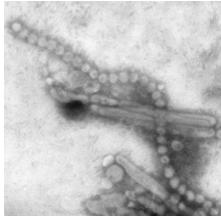


Genes came from multiple species

Multiple reassortant events over time

HA gene of classical swine flu had its origin in the 1918 pandemic: more elderly people had greater immunity

Or just mutation/selection?

Strain	Transmission	Mortality
H1N1	  → 	500,000 infections (UK 2009) Death rate: 1 in 4000
H5N1	  → 	640 infections (World 2003-13) Death rate: 1 in 2
H7N9	  → 	145 infections (World 2013) Death rate: 1 in 4

Key differential: Mode of transmission

Experimental evolution

Aim: evolve a transmissible H5N1 virus in the lab

Ferrets infected with mutant form of the virus kept in cages a short distance apart

Select for more transmissible viruses across multiple generations

With five mutations, the virus became airborne-transmissible: suggests that a pandemic may be fairly likely



Response to the H5N1 experiments

Press coverage and moratorium

“Probably one of the most dangerous viruses you can make”
– *Ron Fouchier, Senior Author*

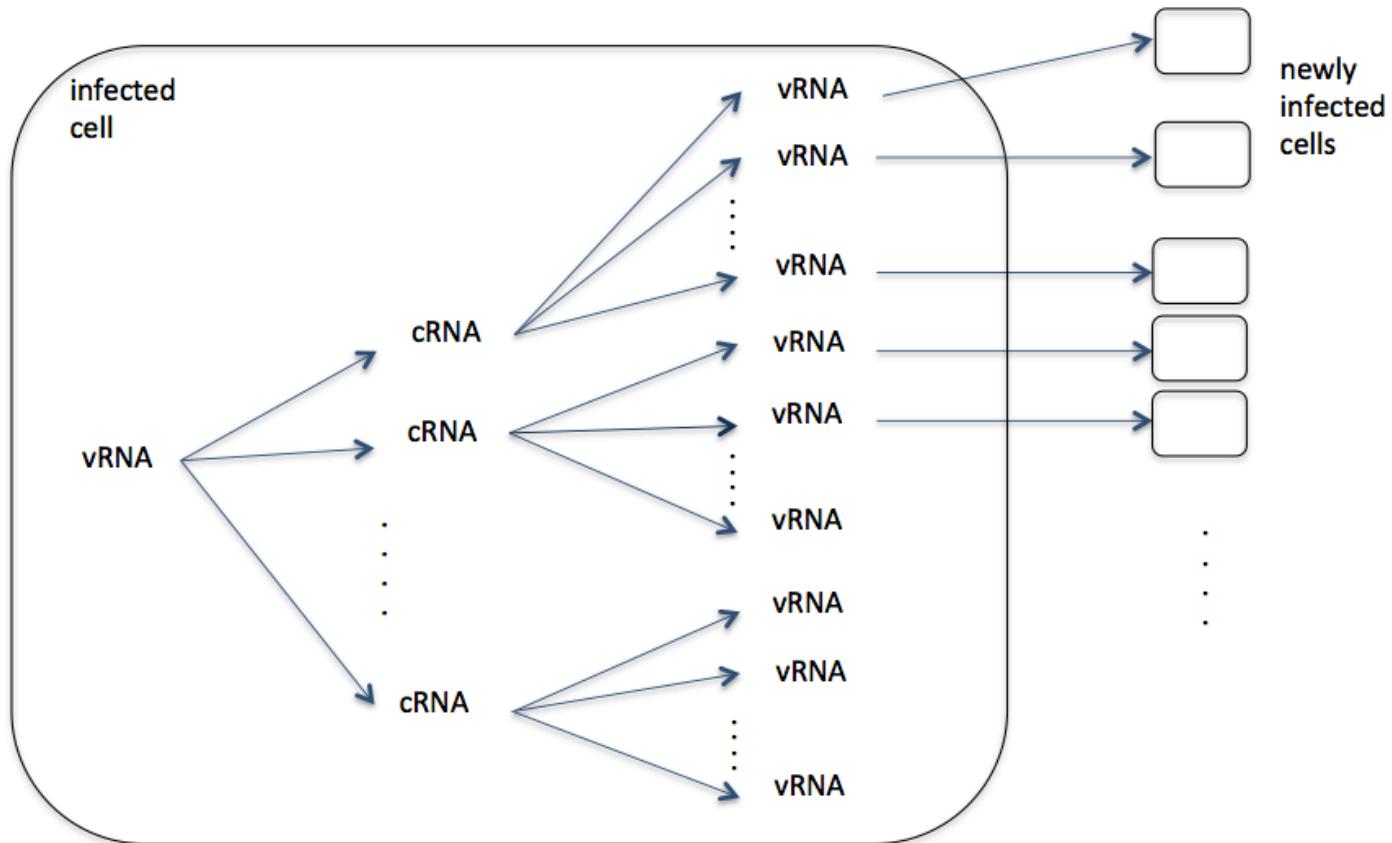
“I don't think anthrax is scary at all compared to this”
– *Paul Keim, National Science Advisory Board for Biosecurity*

“A man-made flu virus that could change world history if it were ever set free.”
– *Science Magazine*

Influenza virus

Simulation studies

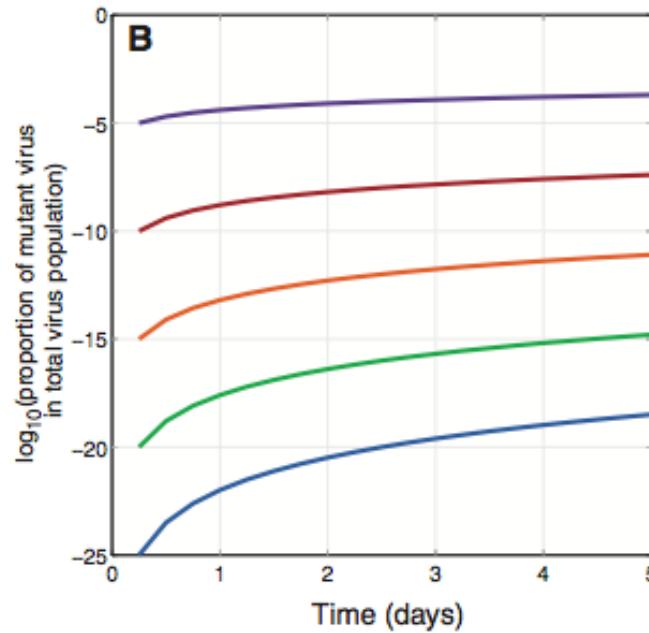
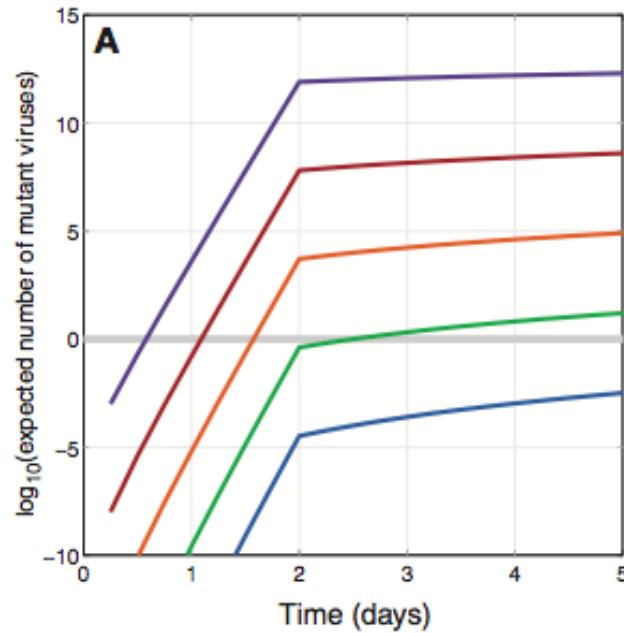
How likely is it to get 5 mutations within the course of a single infection?



Influenza virus

Simulation studies: Genetic drift plus mutation

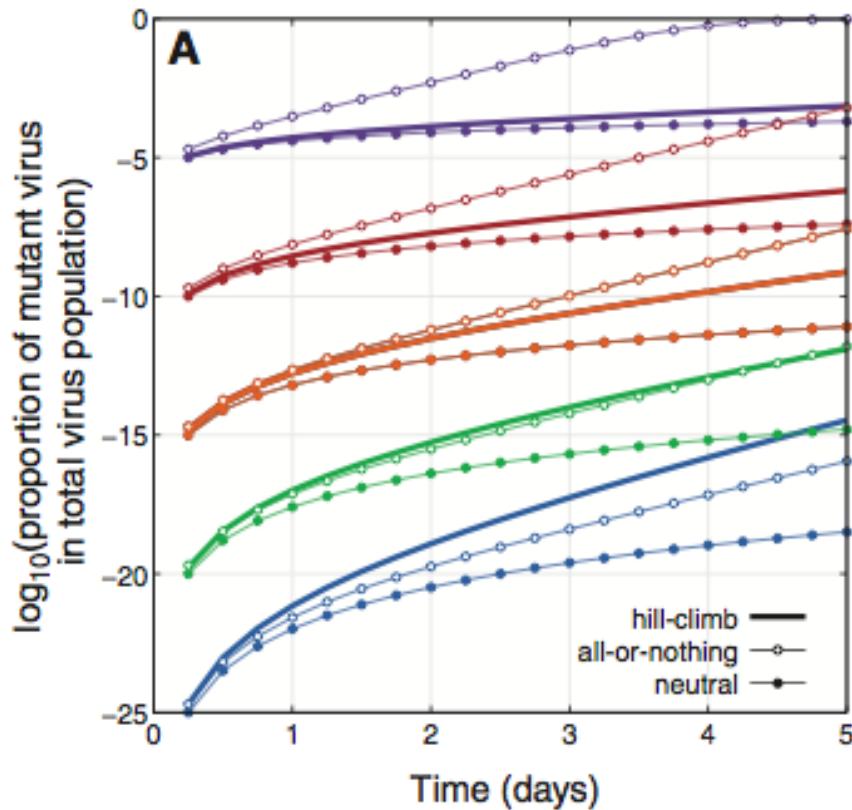
How likely is it to get 5 mutations within the course of a single infection?



Influenza virus

Simulation studies

Different selection models



Three fitness landscapes:

Neutral

All-or-nothing: With all n mutants,
get a fitness bonus

Hill-climb: Each mutation grants a
partial benefit

Example V: Selection in within-host influenza

Wish to measure fitness effects within a single host

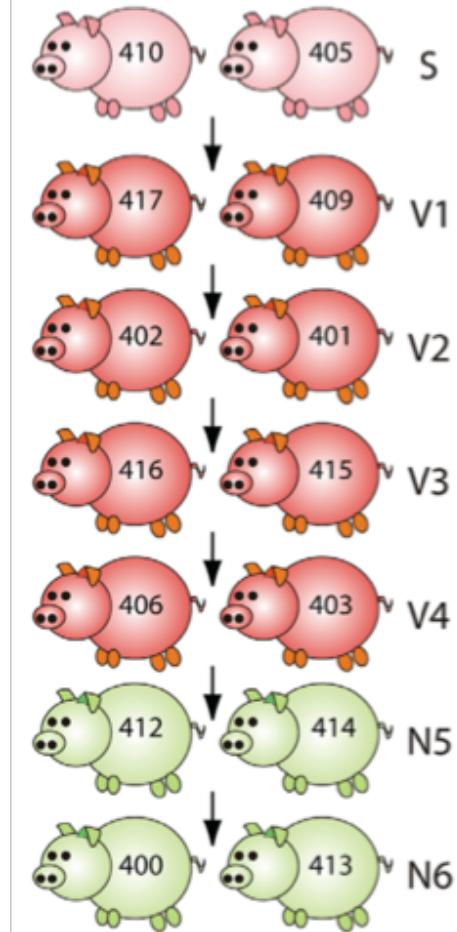
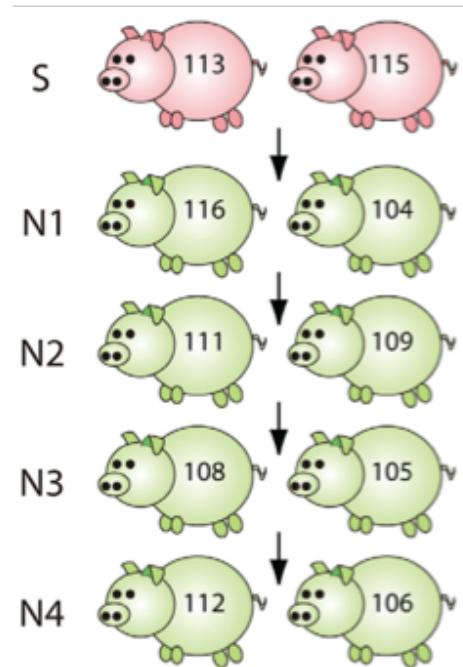
Collect genetic data from single infections

Seeder pigs

Naive pigs

Vaccinated pigs

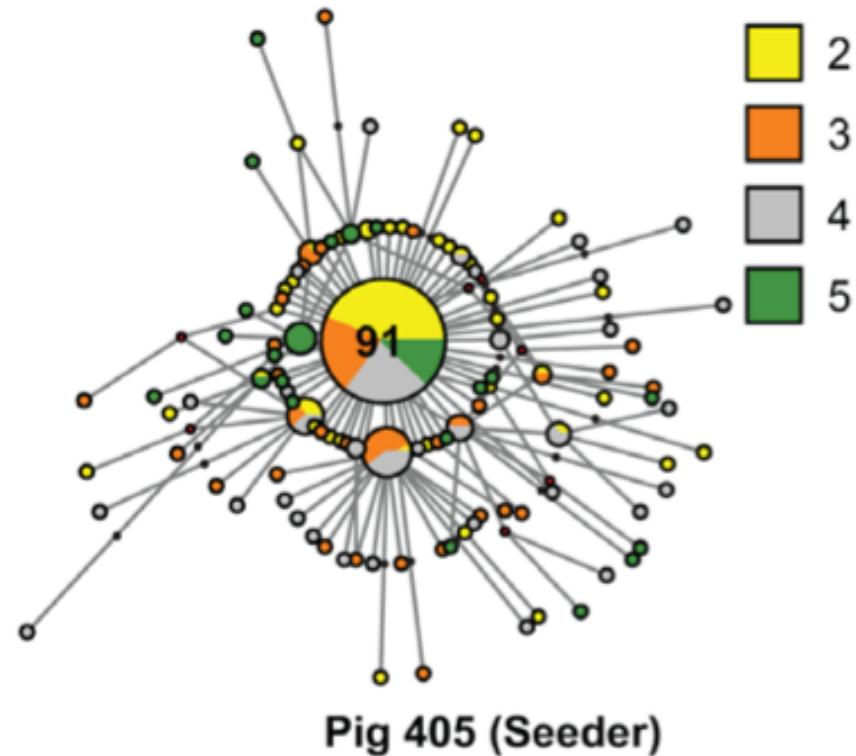
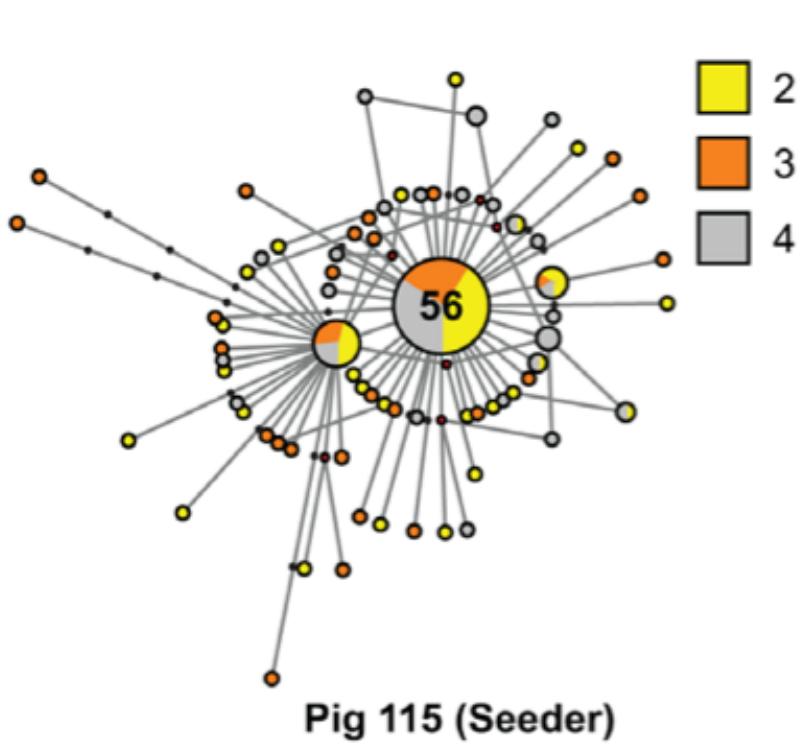
Sanger sequence data from
HA1 region of the virus
collected over time



Images: Murcia et al, 2012

Within-host sequence data

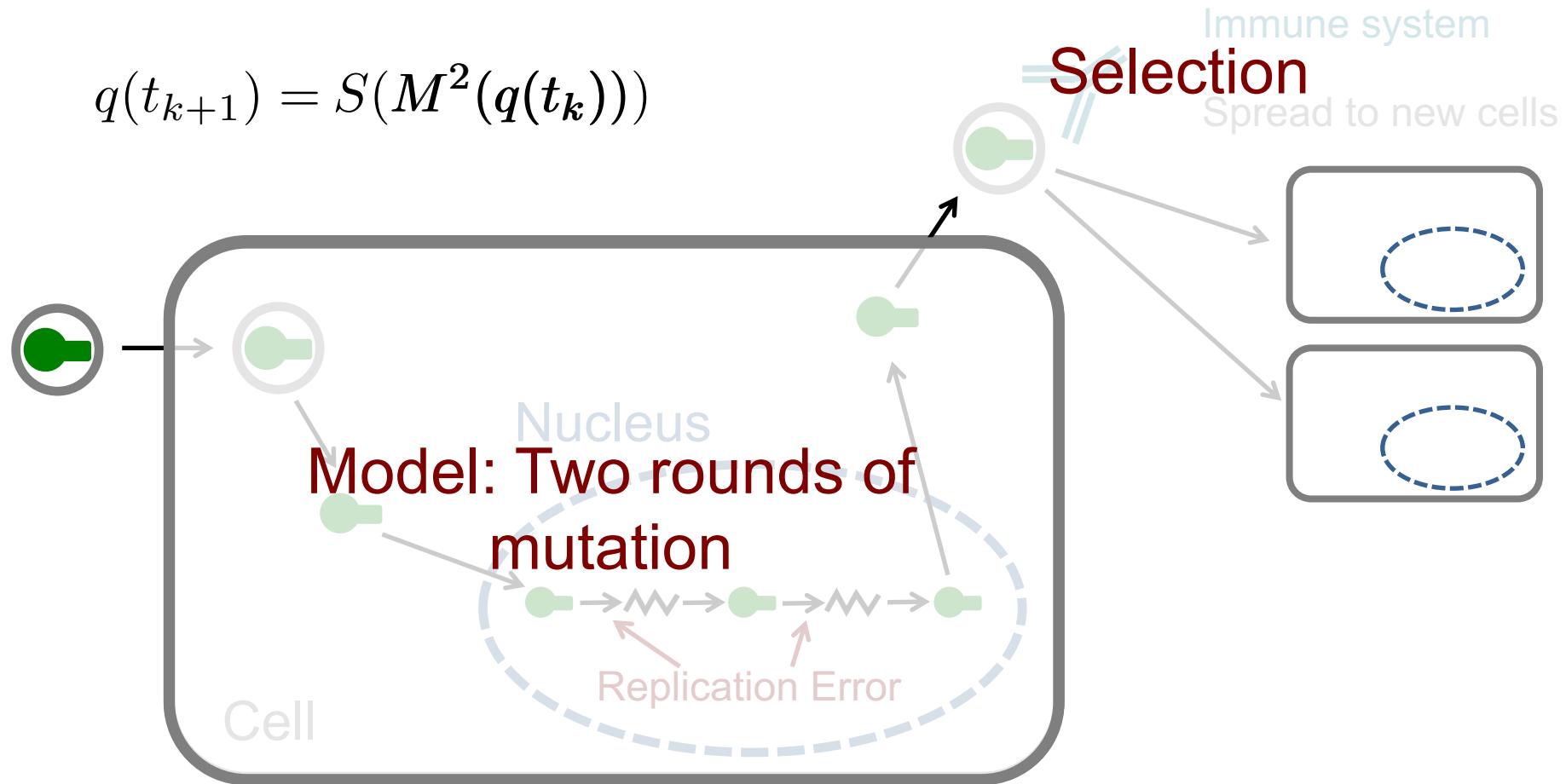
Diversity observed within individual animals: time-resolved structure



Model of viral evolution

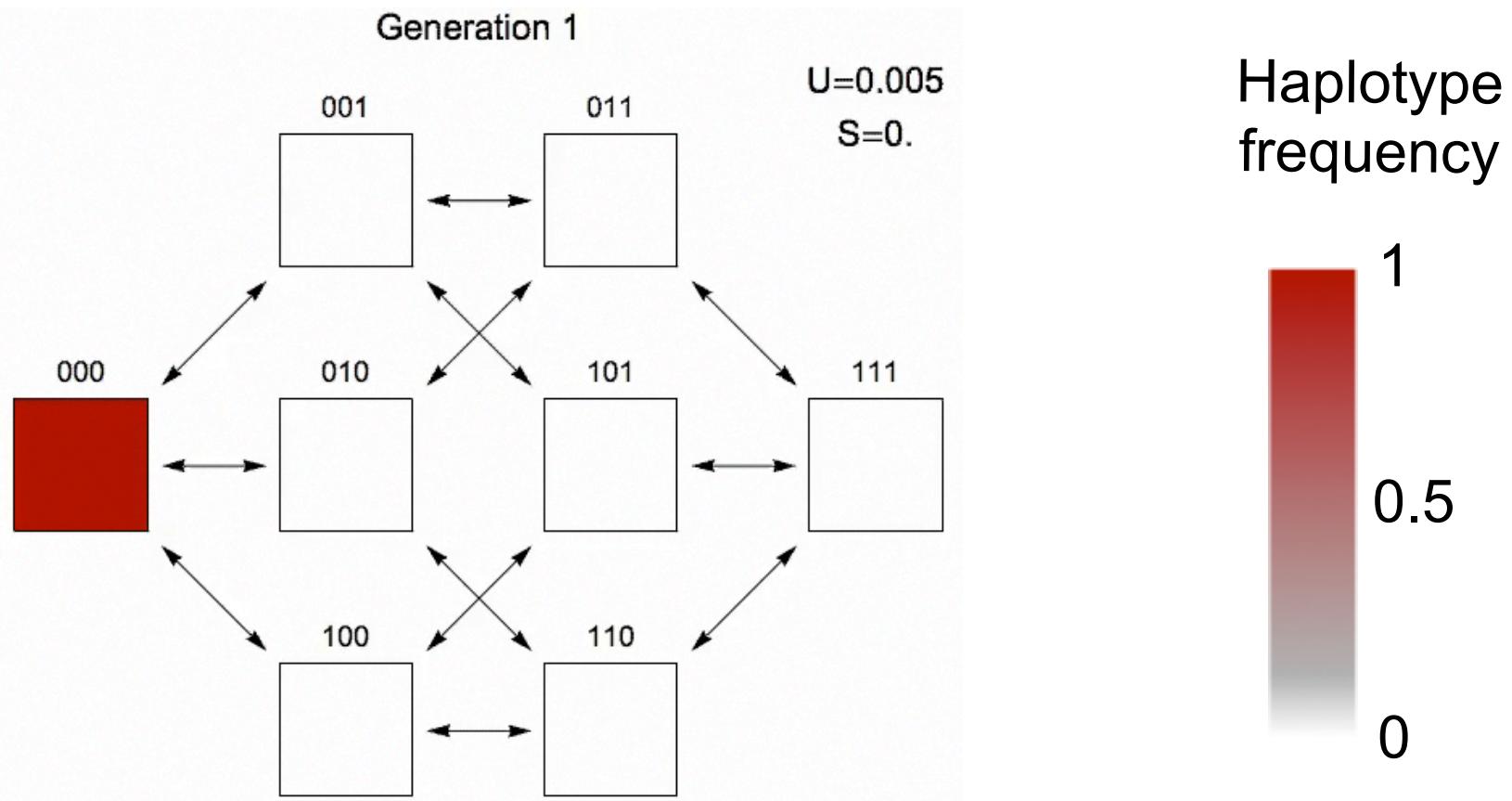
Inference from sequence data?

$$q(t_{k+1}) = S(M^2(q(t_k)))$$



Model of viral evolution

Mutation + selection model: three loci, two alleles



Inference of selection in within-host influenza

Challenges:

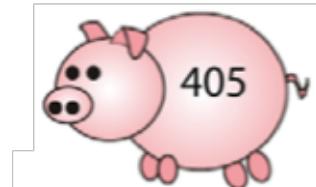
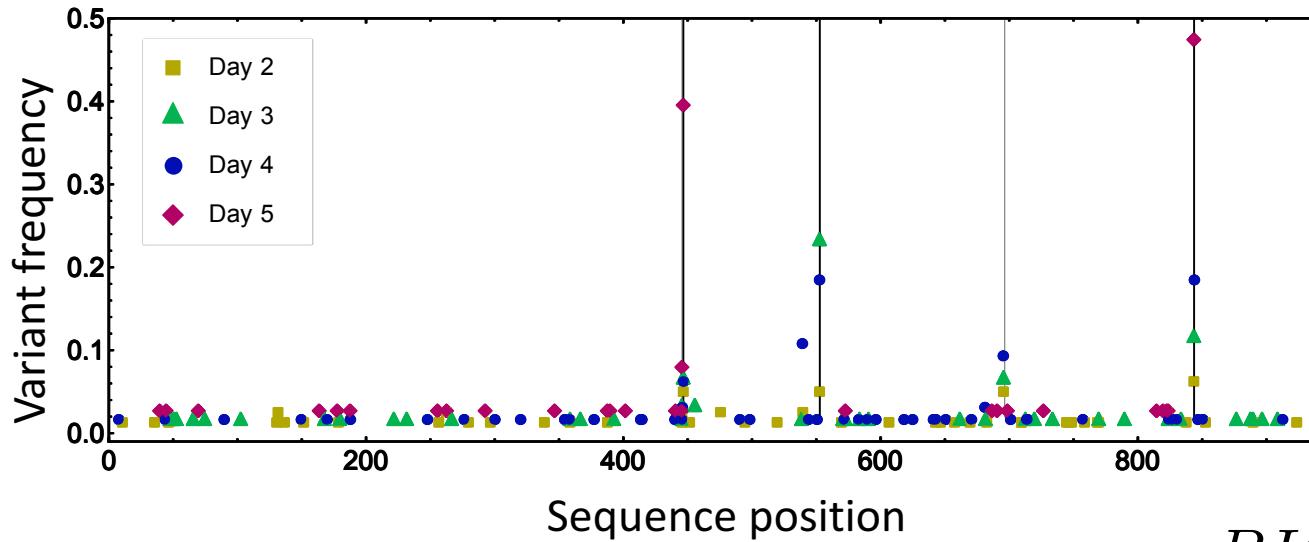
Genotype space is large: 939 nucleotides in HA1:

2×10^{565} possible sequences

Consider loci at which some change in allele frequency is observed

Allele frequency data

Genetic variants in a single animal



Identify sites for which a model of selection outperforms the neutral model:
Three sites fit this criterion

“Error” “Penalty”

$$BIC = -2L + k \log n$$

Likelihood

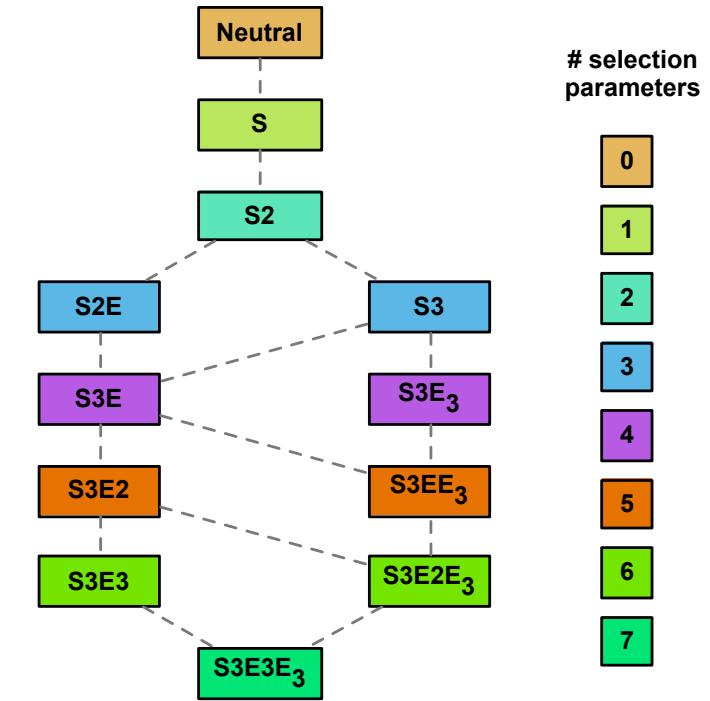
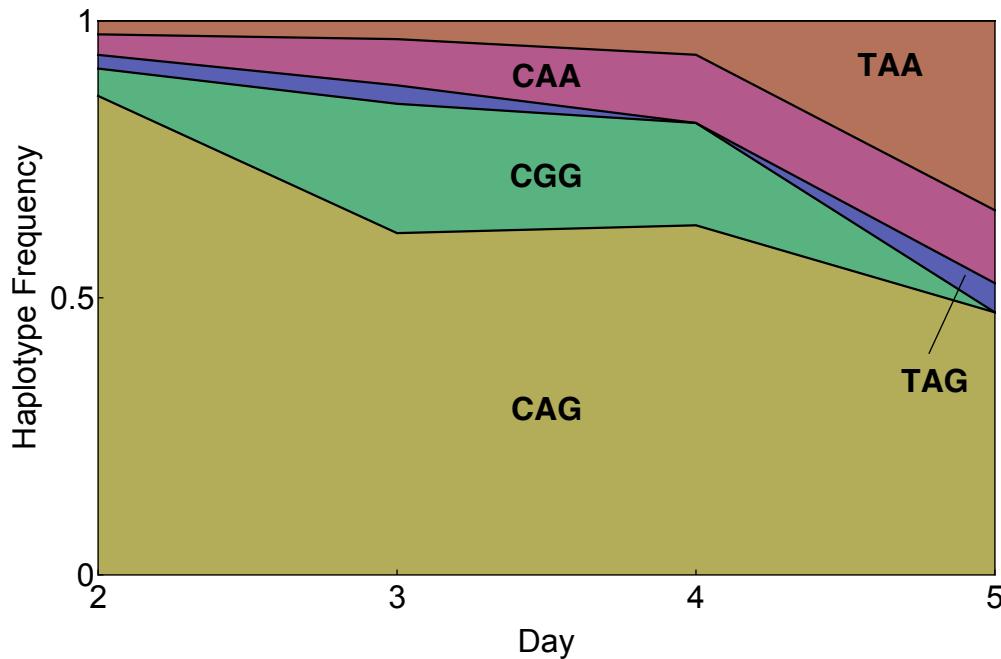
parameters

observations

Multi-locus model of evolution

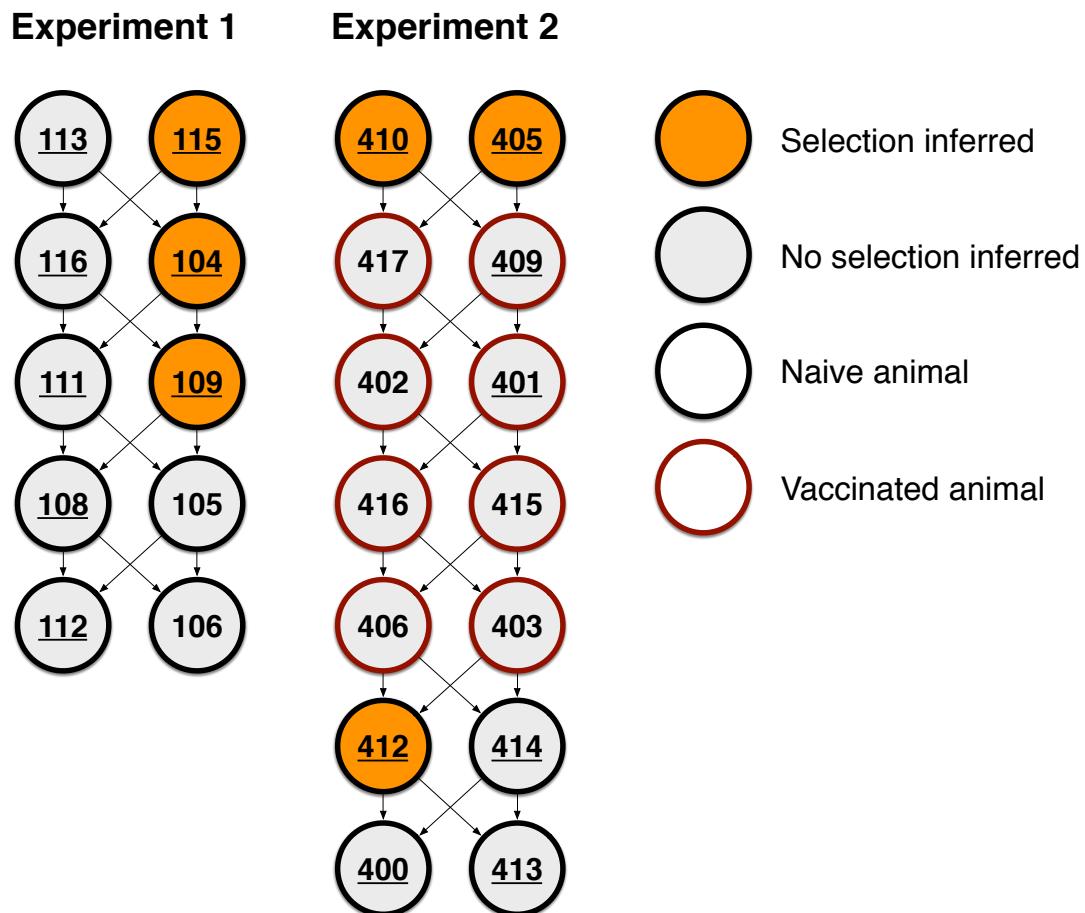
Collect multi-locus sequence data - haplotypes

Hierarchical network of models of selection



Inference of selection in within-host influenza

Evidence for selection in 6 of 16 animals



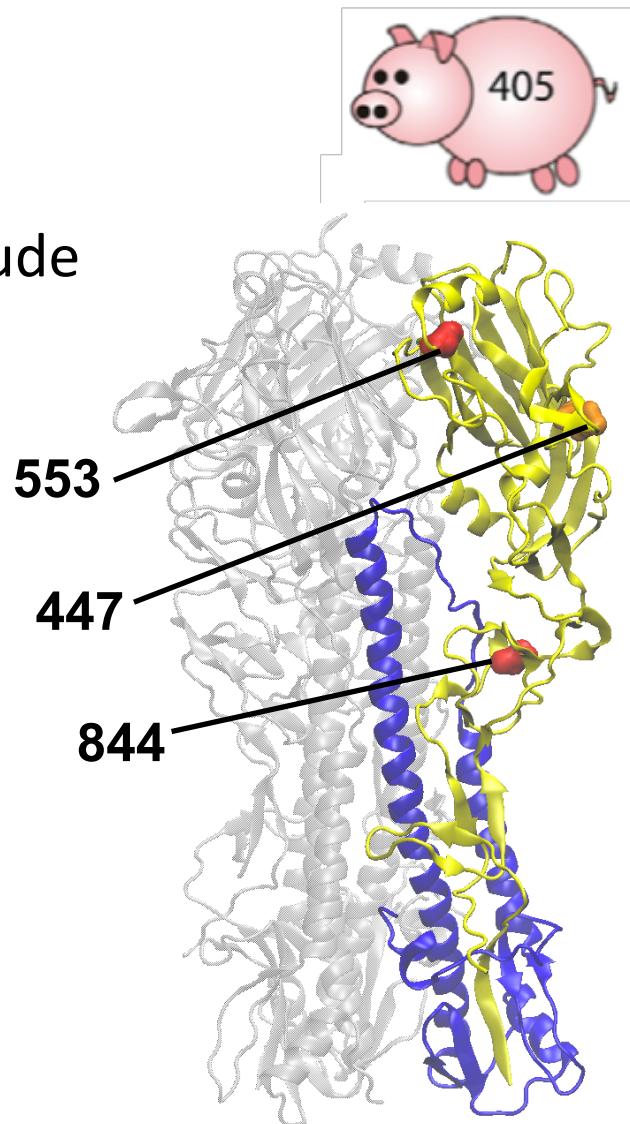
Inference of selection in within-host influenza

Evidence for selection at two loci

Positive selection for A at locus 844 (magnitude +0.4): amino acid change V282I

Variable selection for G at locus 553 (+ve to -ve) : amino acid change N185D

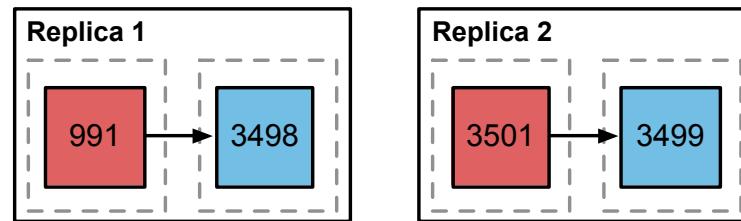
No selection inferred for synonymous change at locus 447



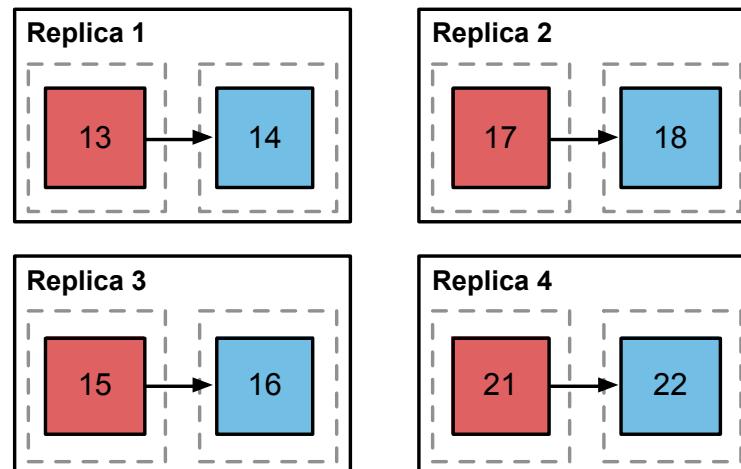
Application to reassortant A/H5N1

Sequence data from Wilker et al, Nature Communications 2013

Experiment 1: VN1203-HA(3)-CA04

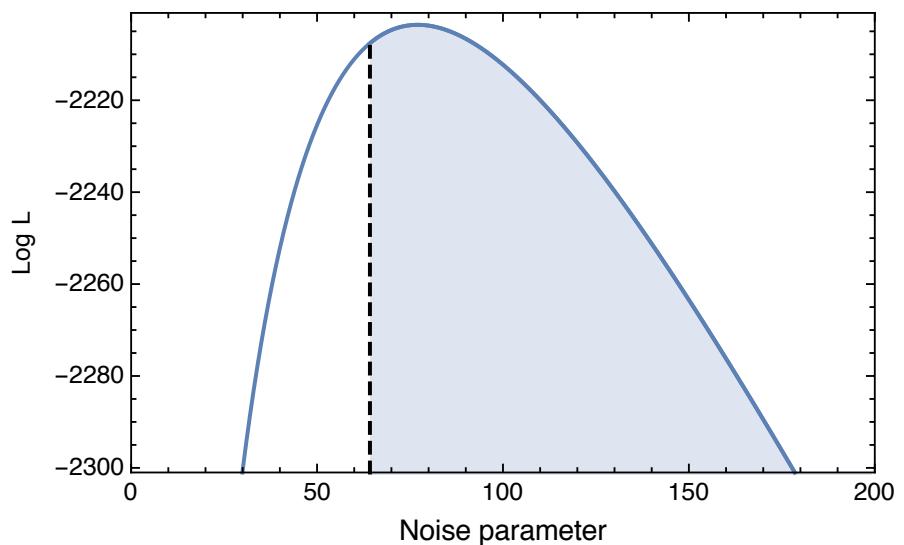
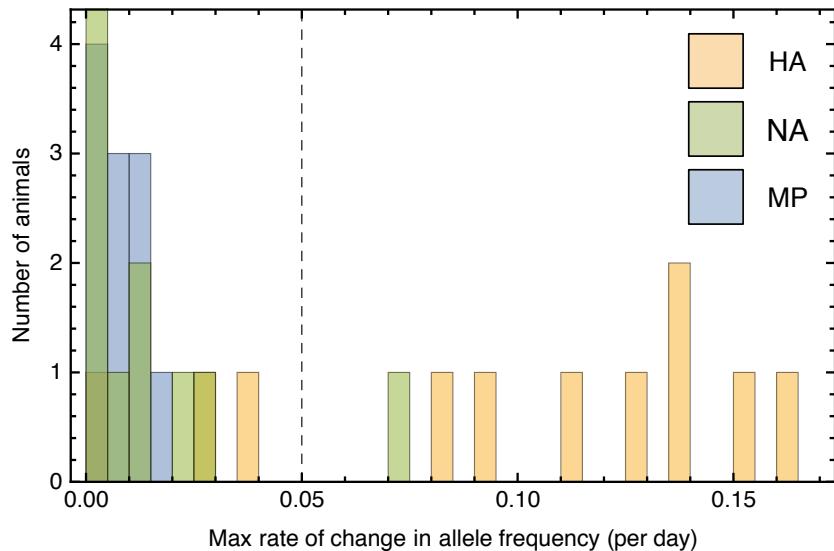
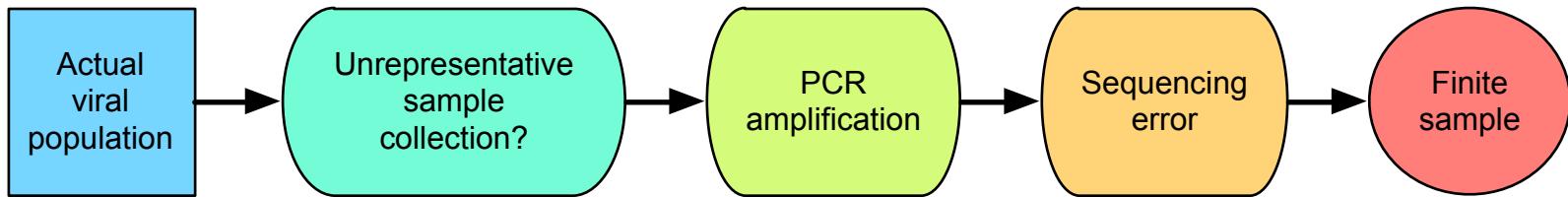


Experiment 2: VN1203-HA(4)-CA04



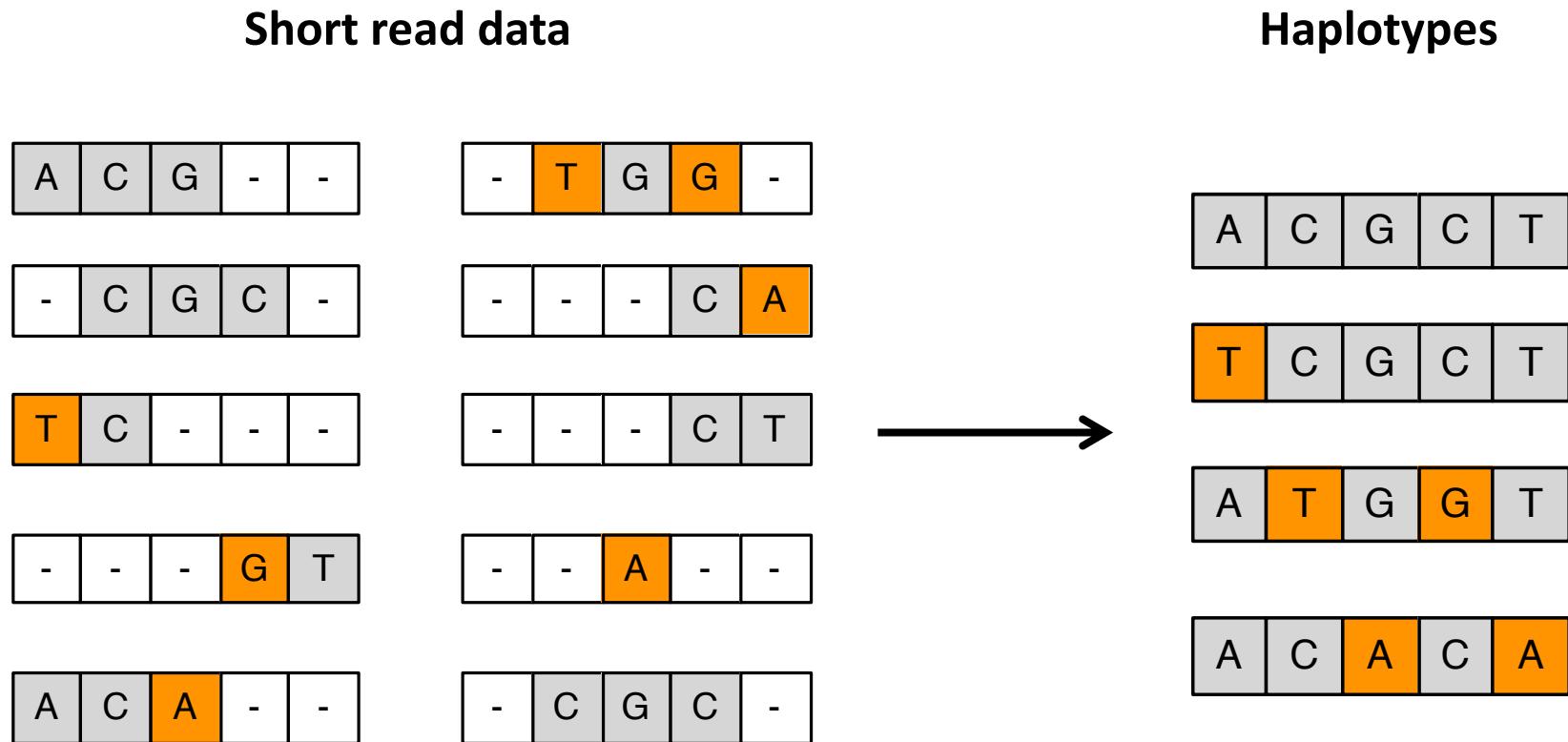
Analysing next-generation data

Quantify the extent of noise in the data



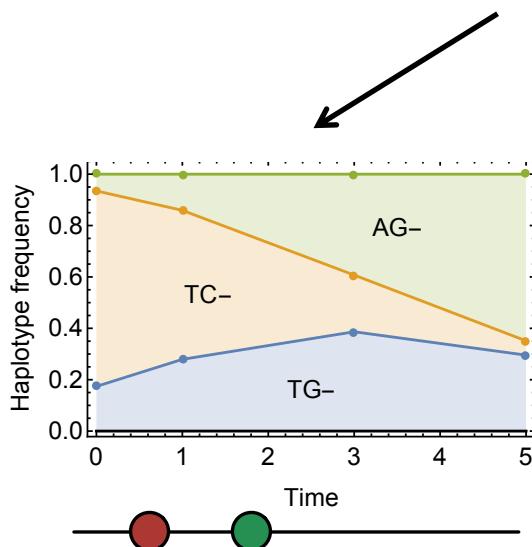
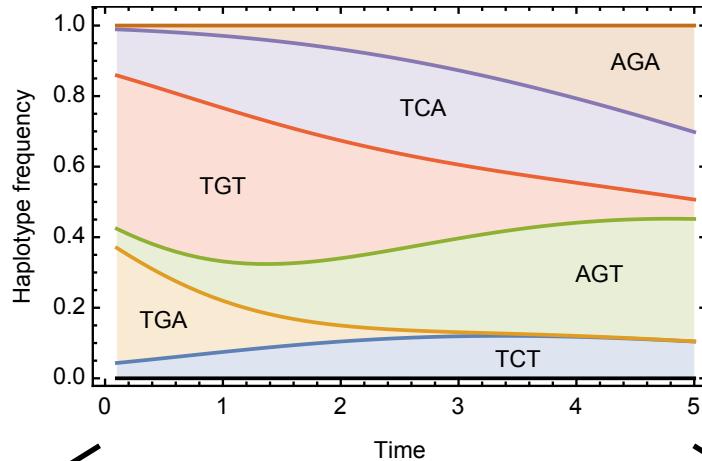
Analysing next-generation data

Reconstruct haplotypes from short sequence reads

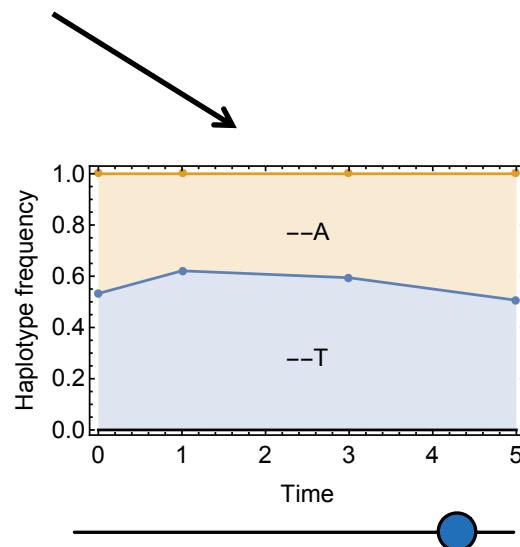


Analysing next-generation data

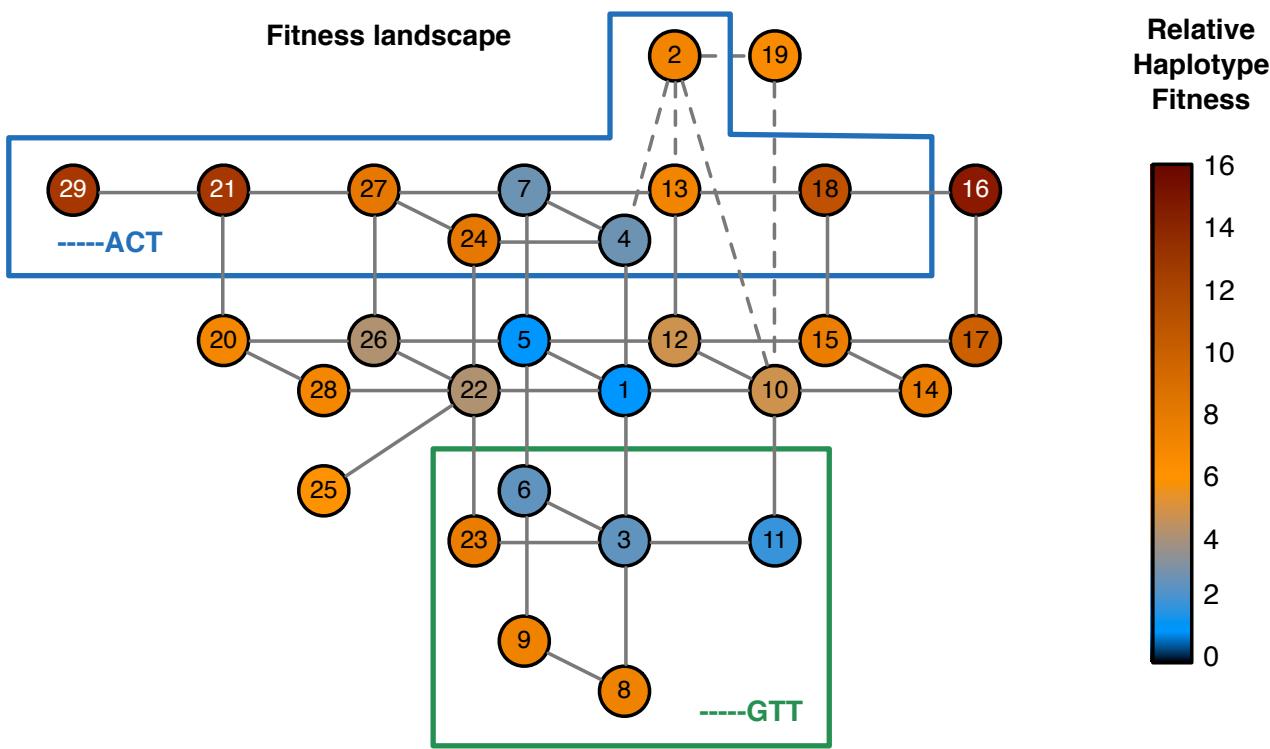
Fit a full sequence model to the partial sequence data



Observations



Multiple pathways to adaptation



1

A G A G G G G C T

2 G G C G G G A C T

3 A G A G G G G T T

4 A G A G G G A C T

5 A G A G T G C T

6 A G A G T G T T

7 A G A G T A C T

8 A G A A G G T T

9

A G A A T G T T

10 G G A G G G G C T

11 G G A G G G G T T

12 G G A G T G C T

13 G G A G T A C T

14 G G A A G G C T

15 G G A A T G C T

16

G G A A T A C C

17 G G A A T G C C

18 G G A A T A C T

19 G G C G G G G C C

20 A T A A T G C T

21 A T A A T A C T

22 A T A G G G C T

23

A T A G G G T T

24 A T A G G A C T

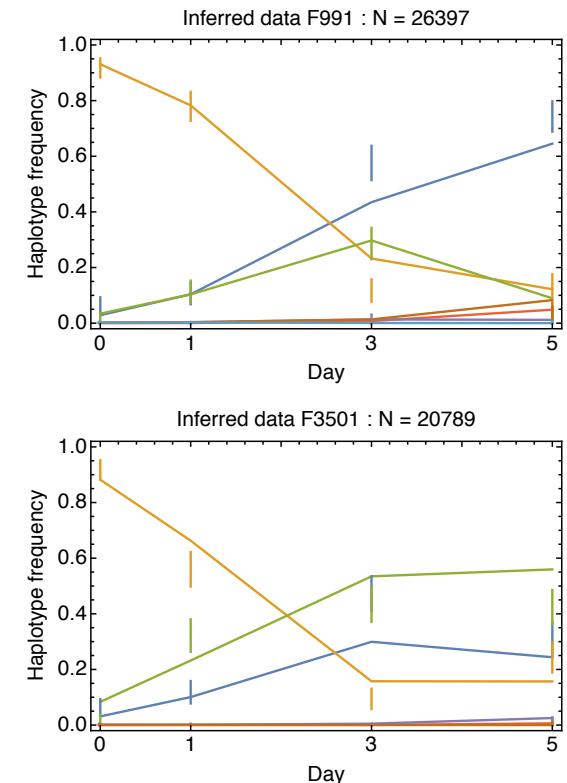
25 A T A G G G C C

26 A T A G T G C T

27 A T A G T A C T

28 A T A A G G C T

29 A T C A T A C T



Summary

Prediction of influenza evolution

Short-term prediction is possible but difficult: Incorporated into vaccine development process

Pandemic prediction is difficult