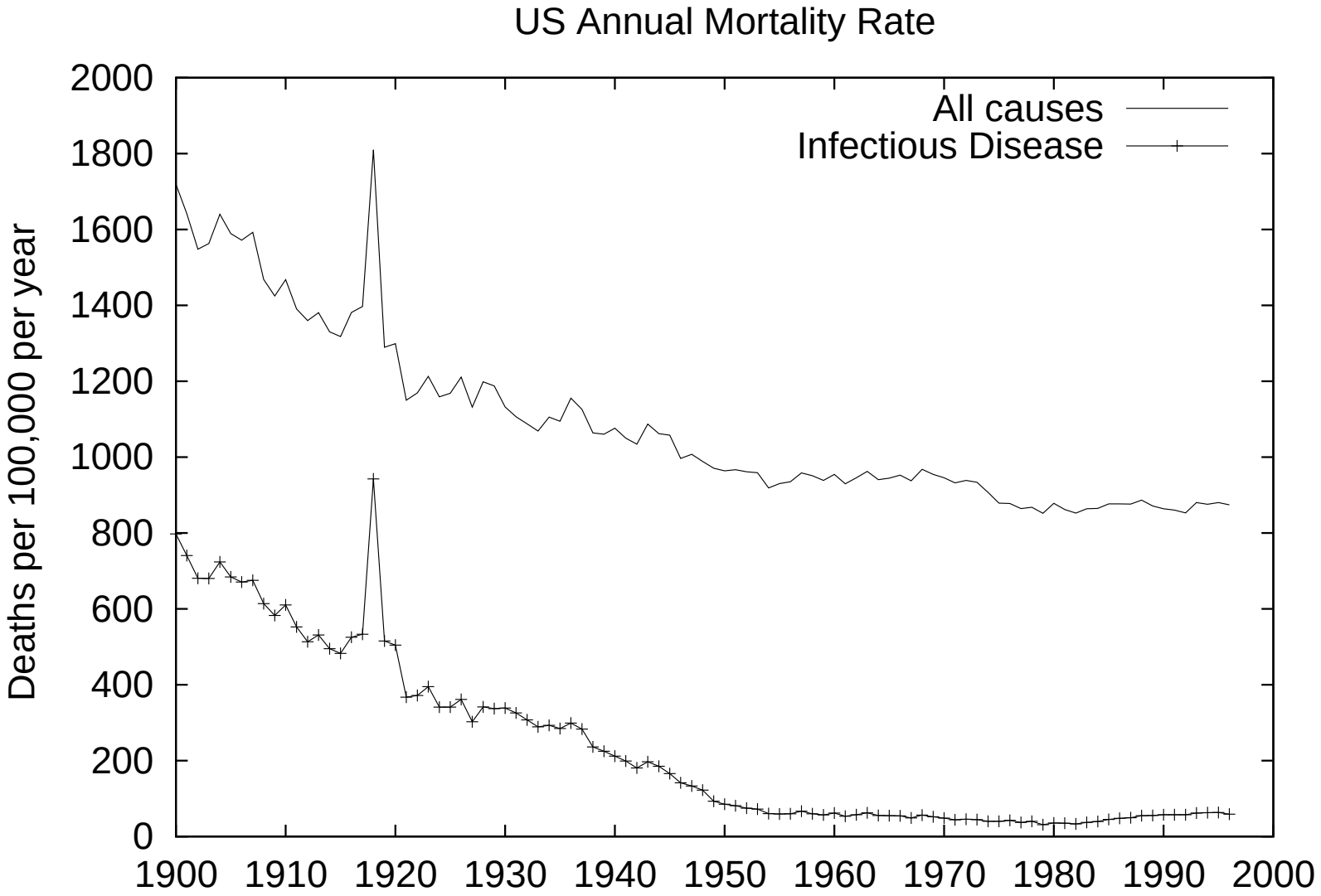
Modeling the Ecology and Evolution of Influenza A Viruses

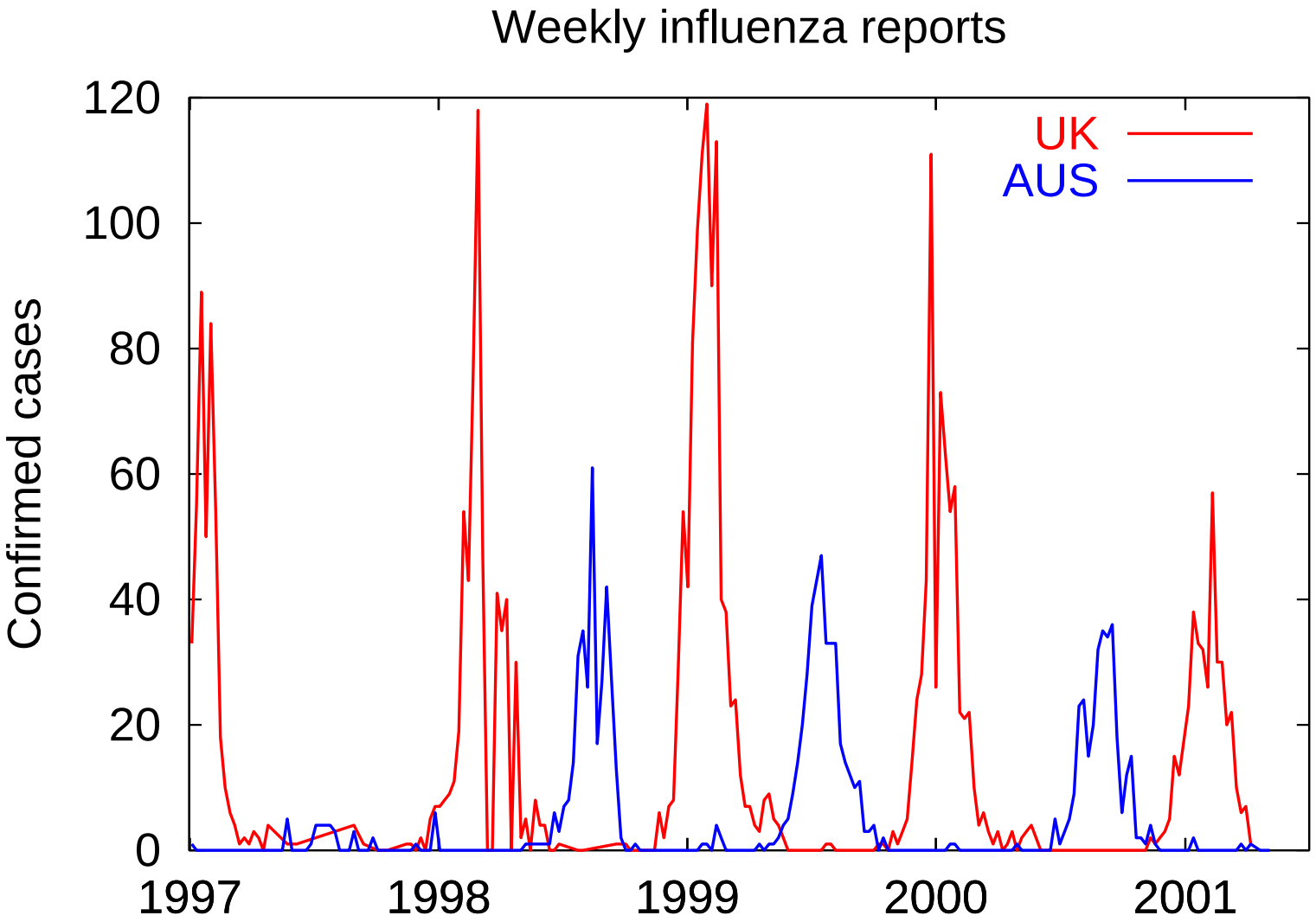
Jonathan Dushoff

Disease Ecology Symposium 2005

eware of mathematics and all who make empty prophecies. The danger already exists that mathematicians have made a covenant with the devil to darken the spirit and confine man to the bonds of Hell.

—St. Augustine

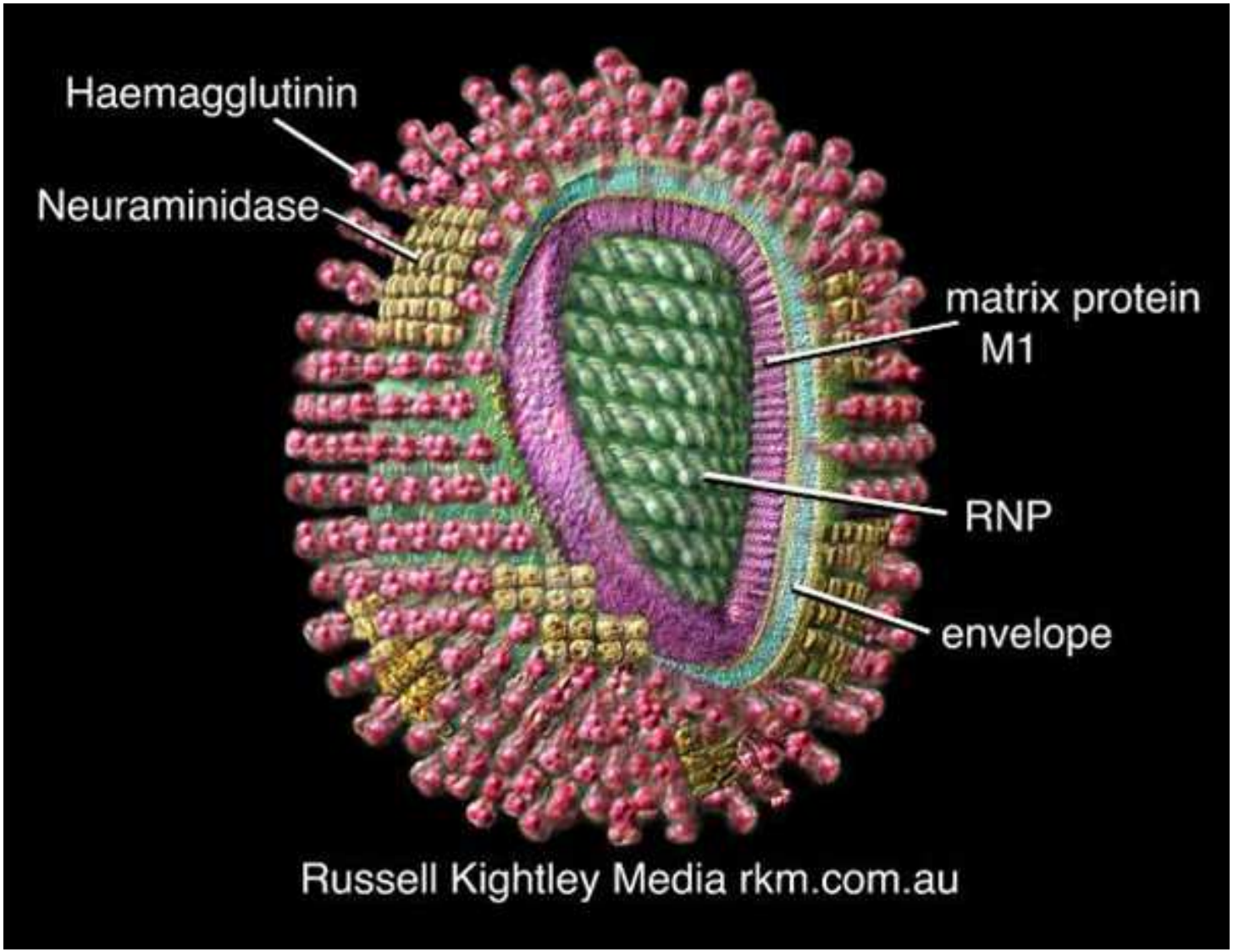

US Annual Mortality Rate
Deaths per 100,000 per year
2000
1800
1600
1400
1200
1000
800
600
400
200
0
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
All causes
Infectious Disease


Weekly influenza reports
Confirmed cases
120
100
80
60
40
20
0
1997
1998
1999
2000
2001
UK
AUS

Influenza A viruses

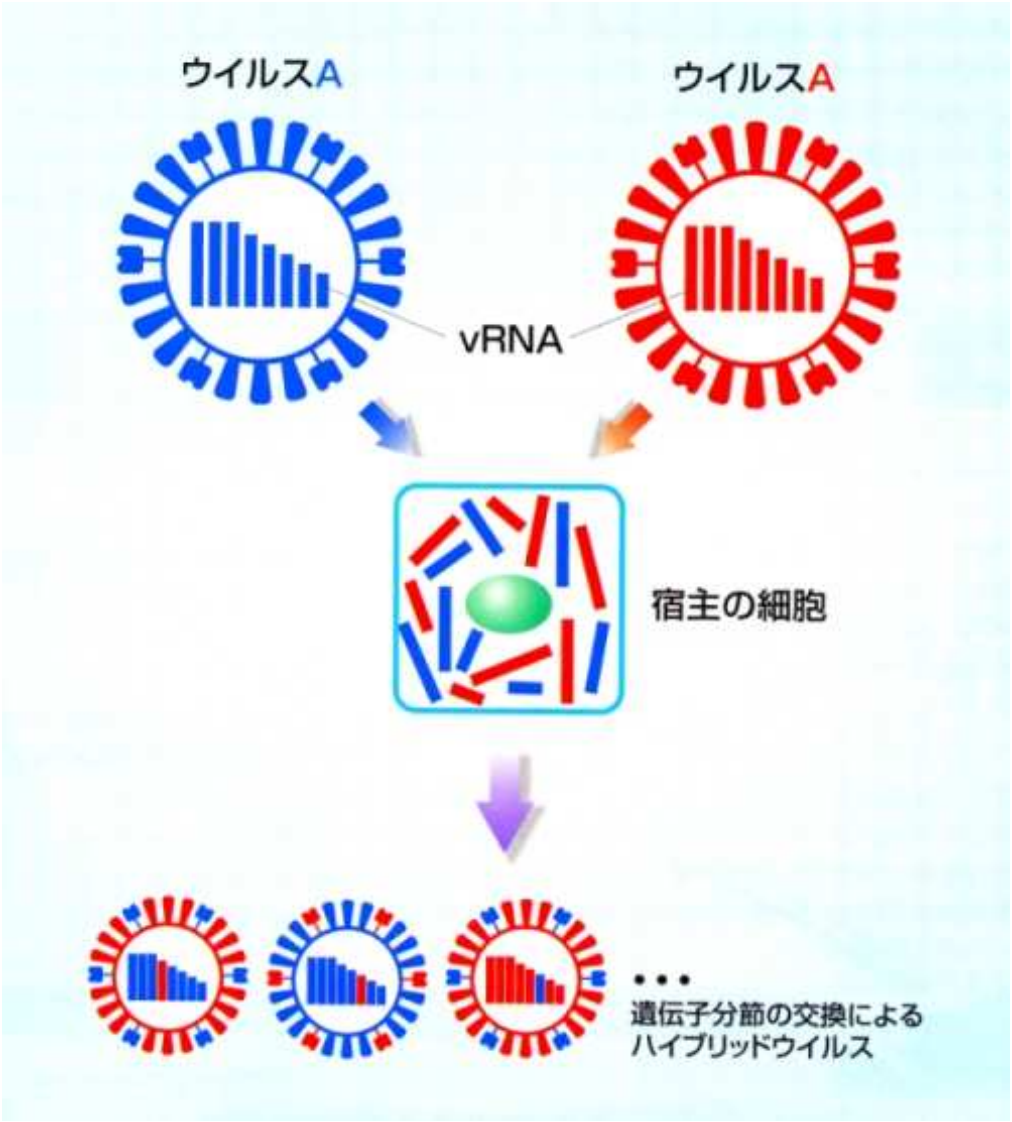
- An important cause of morbidity and mortality on an annual basis.
- Cause occasional pandemics, with extremely high infection rates, and sometimes extremely high mortality.
- Endemic in many mammal and bird populations, with tremendous, stable antigenic diversity in wild waterfowl populations.
- A remarkable capacity for antigenic evolution.
- Epidemiologically more significant than influenza B and C viruses, which circulate primarily in humans.

An influenza virion



A human host (juvenile)

http://homepage2.nifty.com/yamasaki-clinic/



Shift evolution

Major antigenic change caused by reassortment between human and avian virus segments.

1918 Spanish flu (H1N1) replaces earlier strain.

1957 H2N2 replaces H1N1.

1968 H3N2 replaces H2N2.

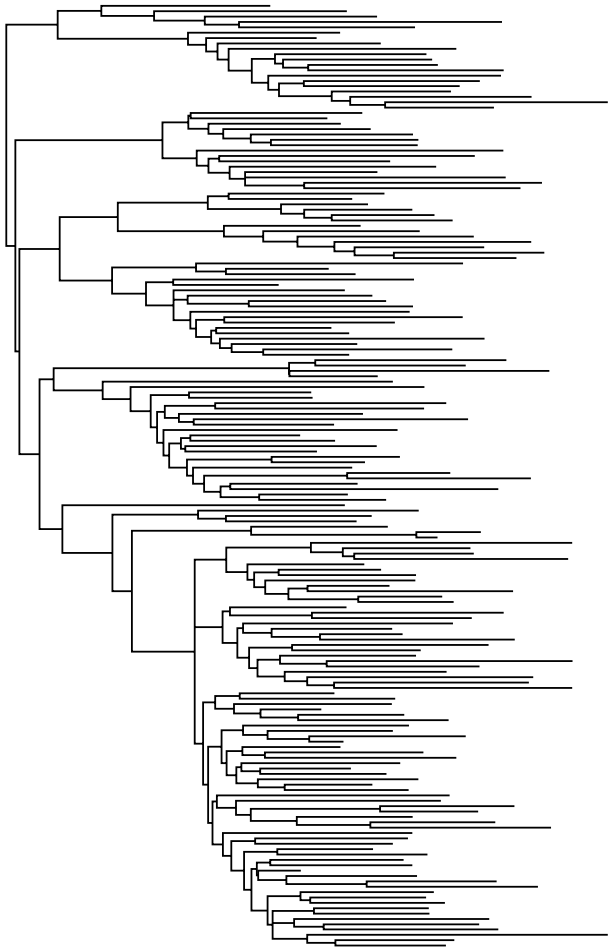
1977 H1N1 mysteriously reappears.

It is estimated that there have been roughly 10 influenza pandemics (presumably caused by shifts) in the last 250 years.

Drift Evolution

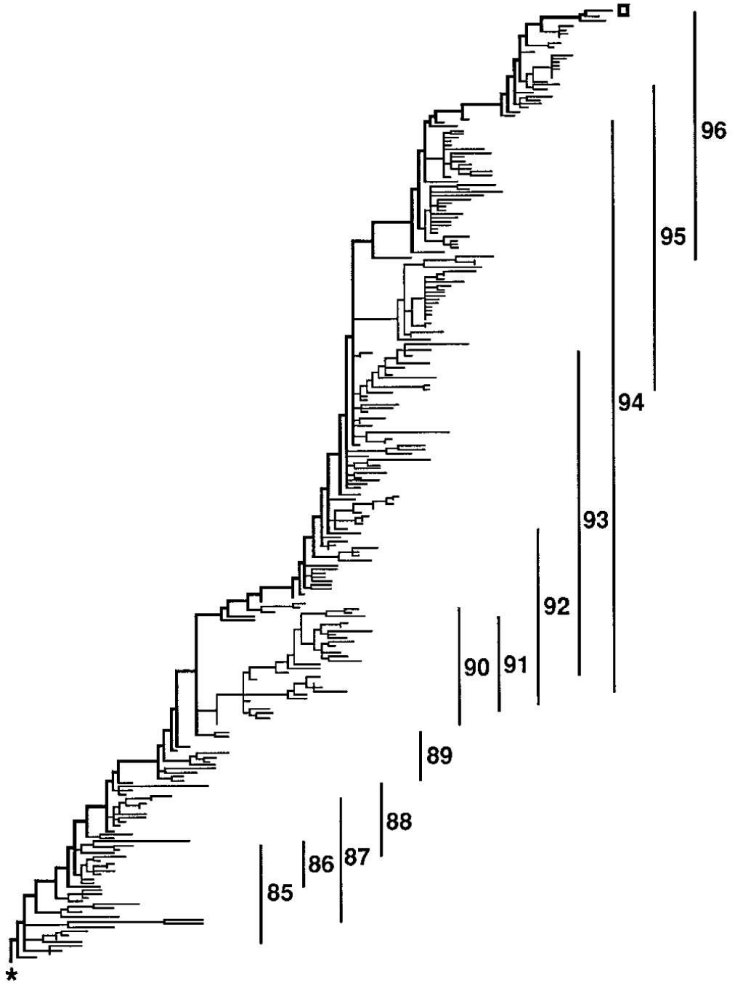
- Each influenza subtype undergoes gradual, antigenically significant mutations to HA.
- After a few years, descendants of an infecting 'strain' will have changed enough to re-infect most individuals.
- Unusual phylogenetic pattern generated: a great deal of diversity, but a dominant main trunk.
- Influenza B viruses show a similar, but less dramatic, pattern.

HIV-1



Rambaut, et al., 2001

Influenza A



Fitch, et al., 1997

Human influenza is an exciting system

- Important source of mortality and morbidity on an annual basis
- Unpredictable and sometimes devastating pandemics
- Excellent model system for evolutionary ecology.
  - Interactions across scales, from cellular to individual to group to globe.
  - Drift evolution is driven by, and shapes, landscape of human HA antibodies.

Influenza modeling framework

Vision A modeling framework that links scales from cellular, to individual, to population, to globe to predict influenza drift evolution, at the level of HA amino-acid sequences.

Reality This vision may or may not be achievable. Importantly, it provides a useful unifying framework for exciting research questions.

Links genomics, structural investigations of proteins, epidemiology and mathematical modeling.

Overview

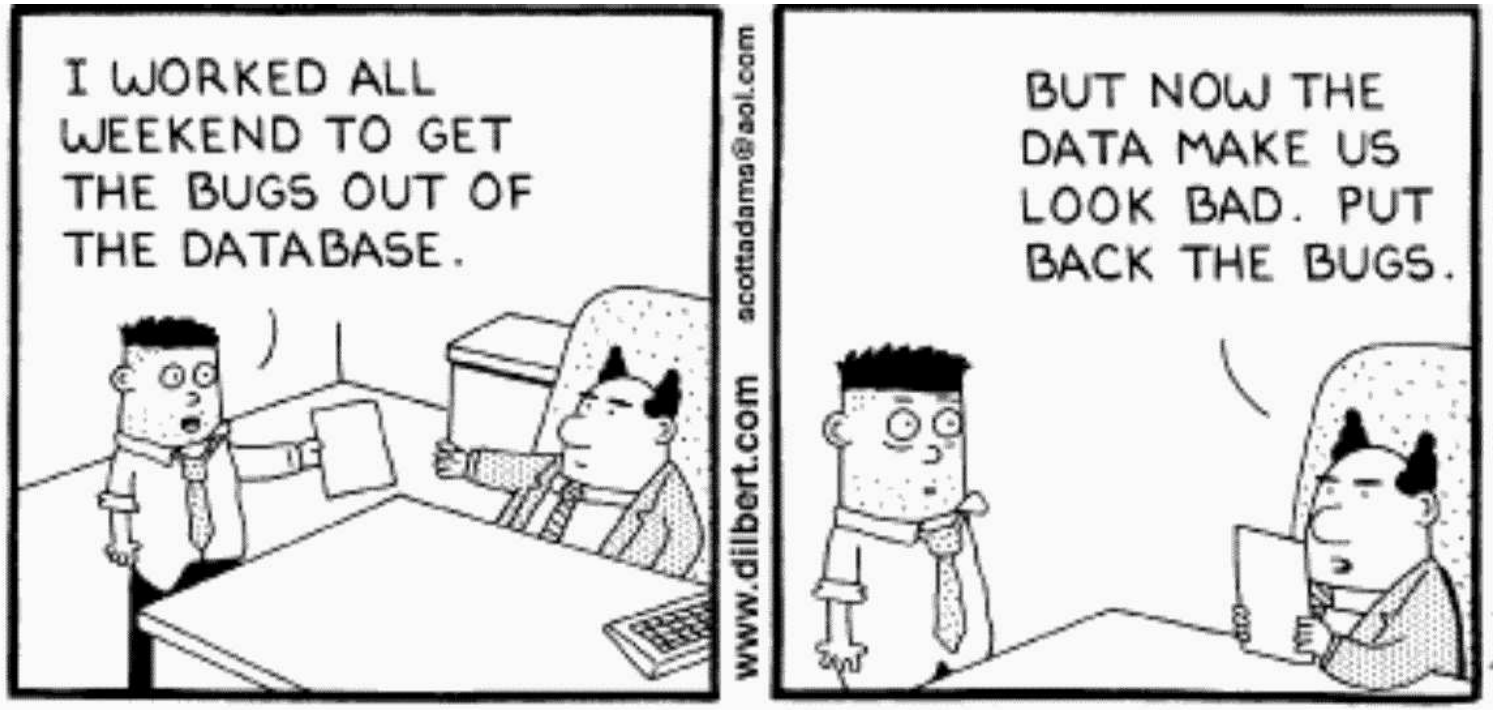
- Mortality burden of influenza
- Evolution and quasispecies structure
- Simple, individual-based models

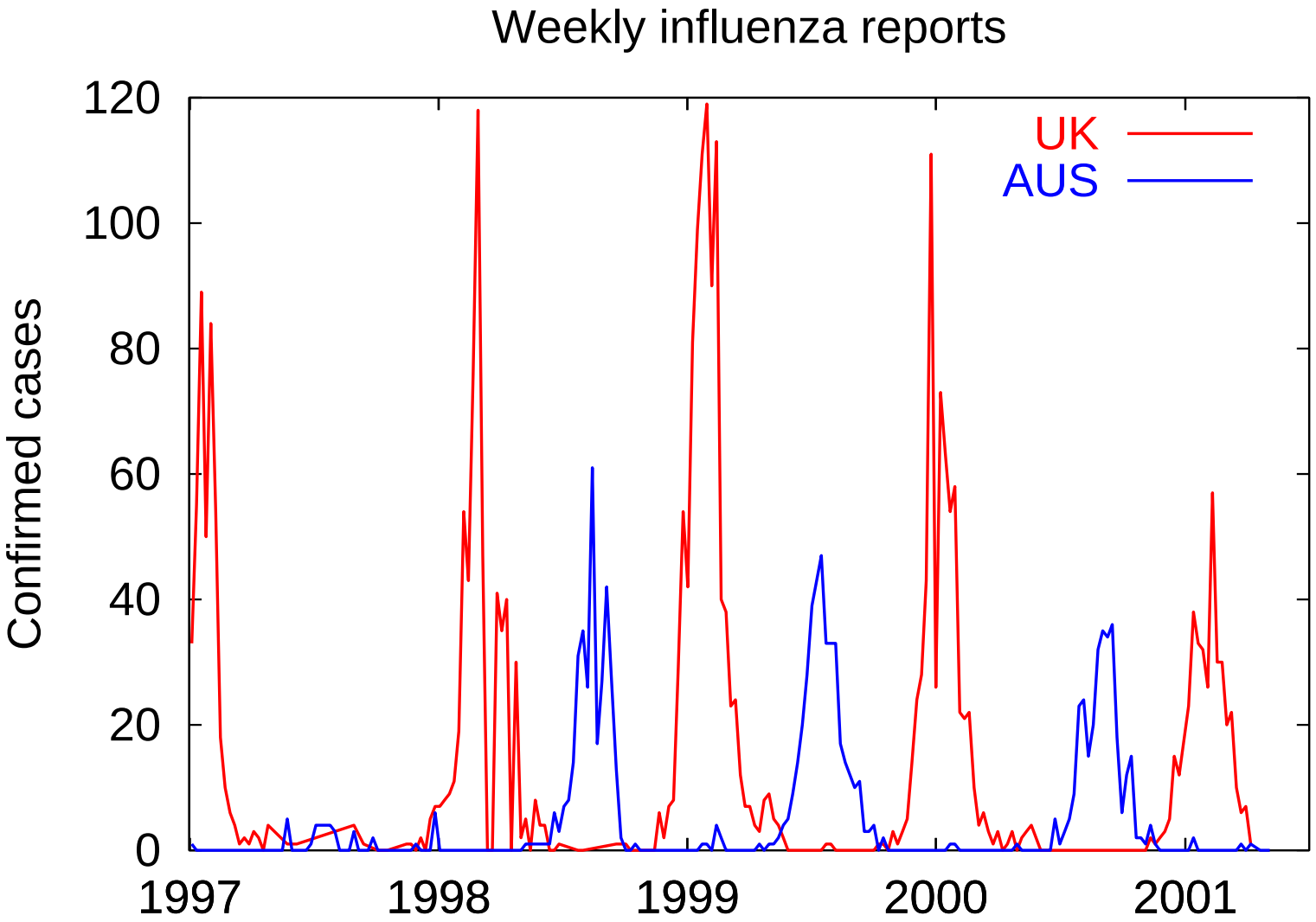
Overview

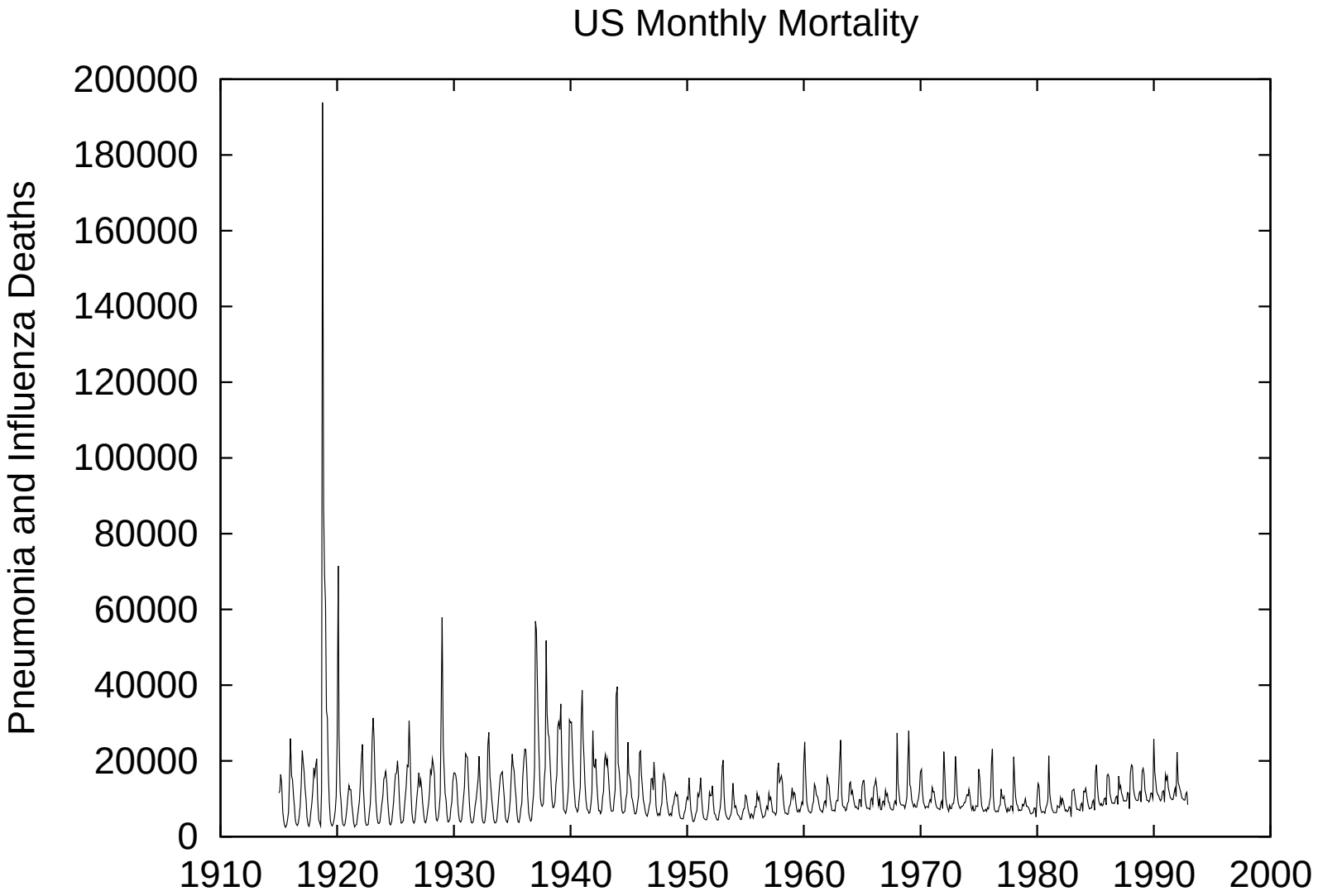
- **Mortality burden of influenza**
- Evolution and quasispecies structure
- Simple, individual-based models

The mortality burden of influenza

Dushoff, Plotkin, Viboud, Earn and Simonsen (submitted)




Weekly influenza reports
Confirmed cases
120
100
80
60
40
20
0
1997
1998
1999
2000
2001
UK
AUS


US Monthly Mortality
Pneumonia and Influenza Deaths
200000
180000
160000
140000
120000
100000
80000
60000
40000
20000
0
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000

How many people die of influenza in temperate countries?

Source	Method	Estimate (ann. deaths per 100K)
U.S. Public-use data	influenza deaths	1
Keatinge & Donaldson	daily regression	3
Simonsen et al.	excess mortality	12
Reichert et al.	excess mortality	20
Thompson et al.	weekly regression	20
U.S. Public-use data	pneumonia deaths	80**
Nichol et al.	cohort	$> 100^{*}$

Pitfalls of weekly (or monthly) regression

- Do you have the right shape of the seasonal trend (sinusoid or exponentiated sinusoid are what are used)?
- How to fit the model: do you remove seasonal components first, or fit the whole model at once?
- What about deaths that occur weeks after infection?
- At what time scale are you measuring deaths?

Annualized approach

Weekly approach powerful but problematic, because so many things follow a seasonal pattern: school, weather, other viruses.

Annual approach less statistically powerful, but more robust.

Compare:

- Timing of different death series
- Annual deaths vs. viral prevalence

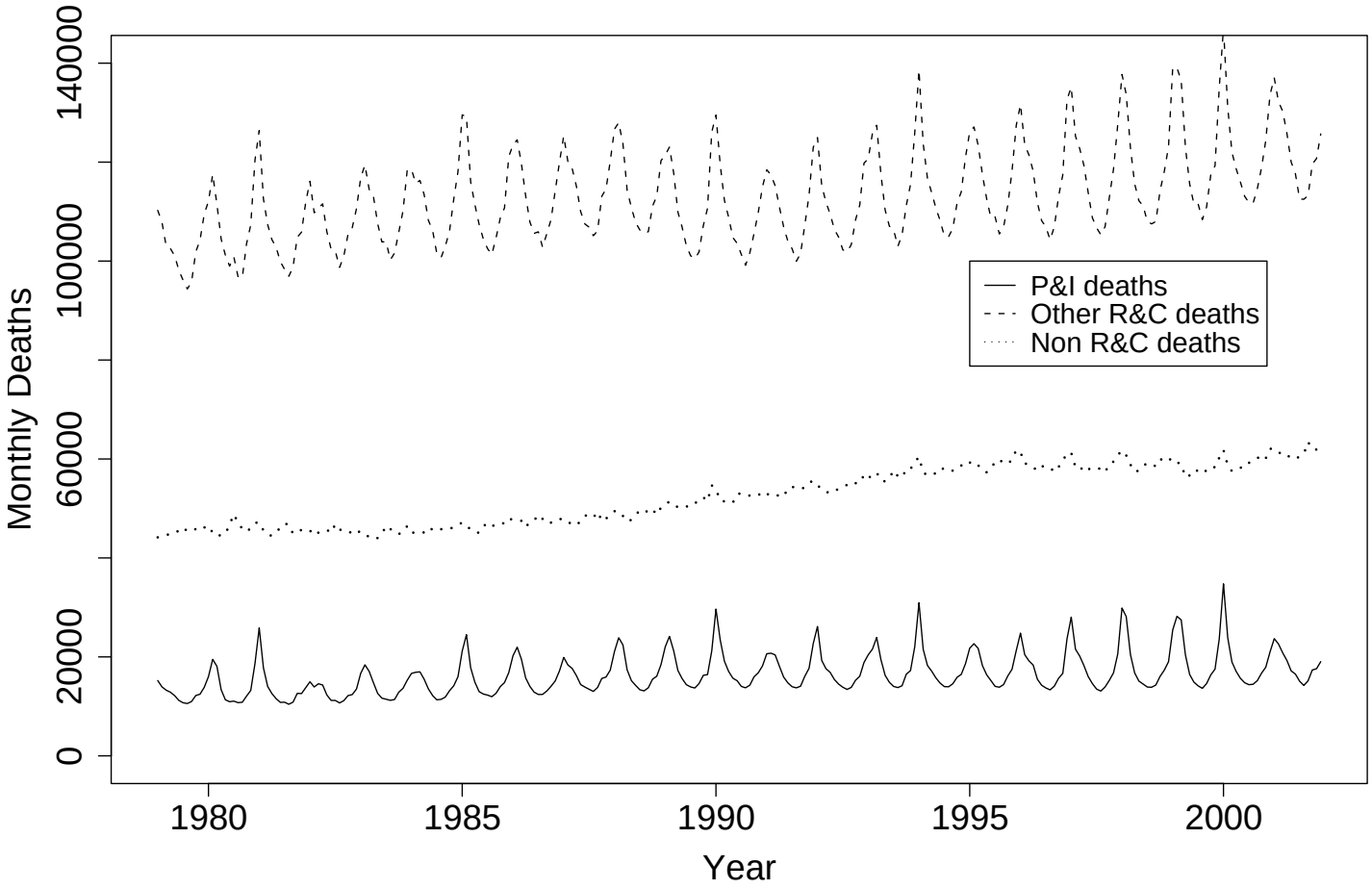
using years as data points.

Time series used

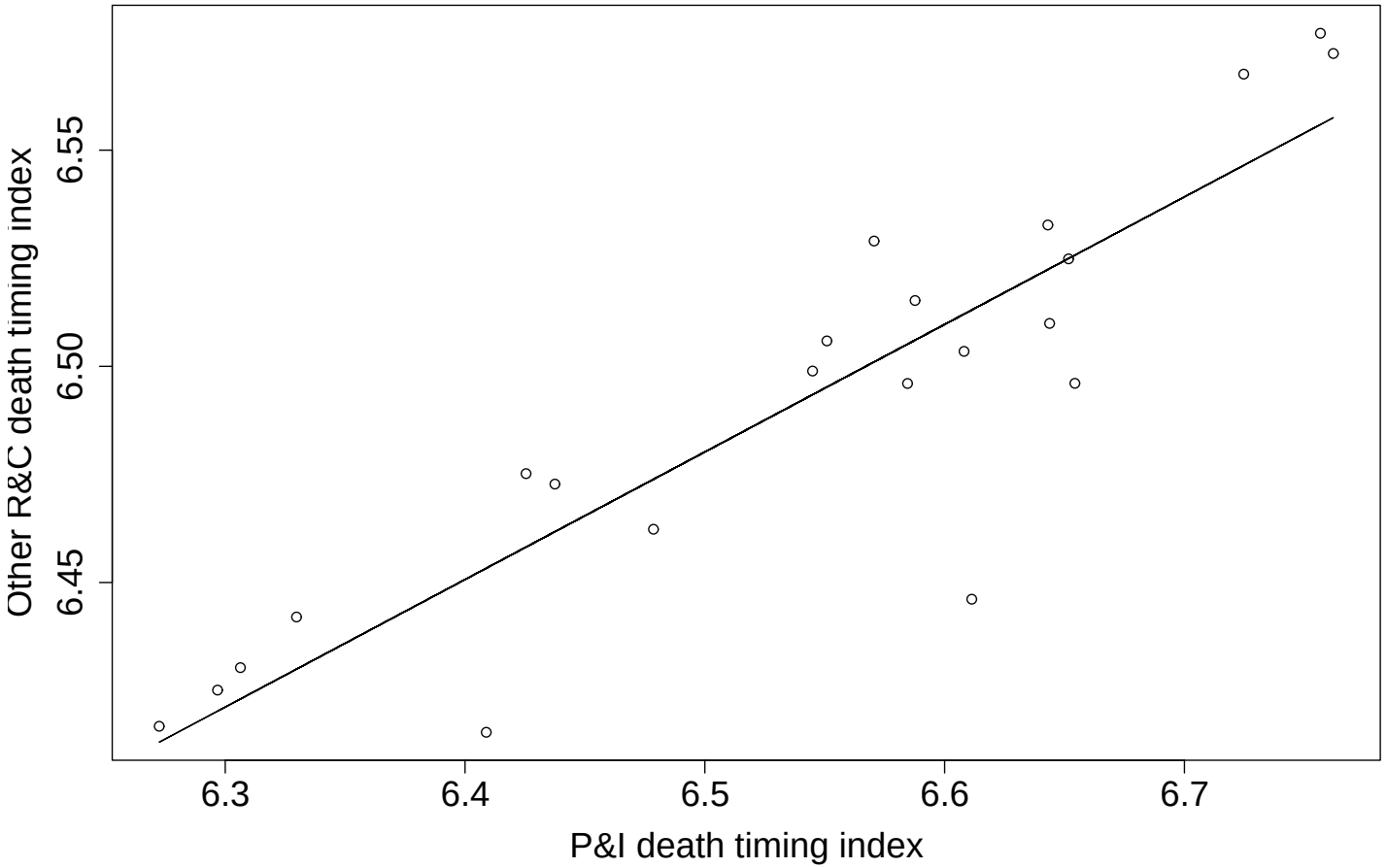
Unlike earlier research, take advantage of 'multiple-cause' data to classify deaths by causes appearing anywhere on the death certificate.

Name	Definition	Deaths/month
P&I	deaths with any mention of pneumonia or influenza	16,405
R&C	deaths with any mention of respiratory and circulatory conditions	129,279
non-R&C	deaths with no mention of respiratory and circulatory conditions	52,447
All deaths		181,726
Other R&C	R&C − P&I	112,874

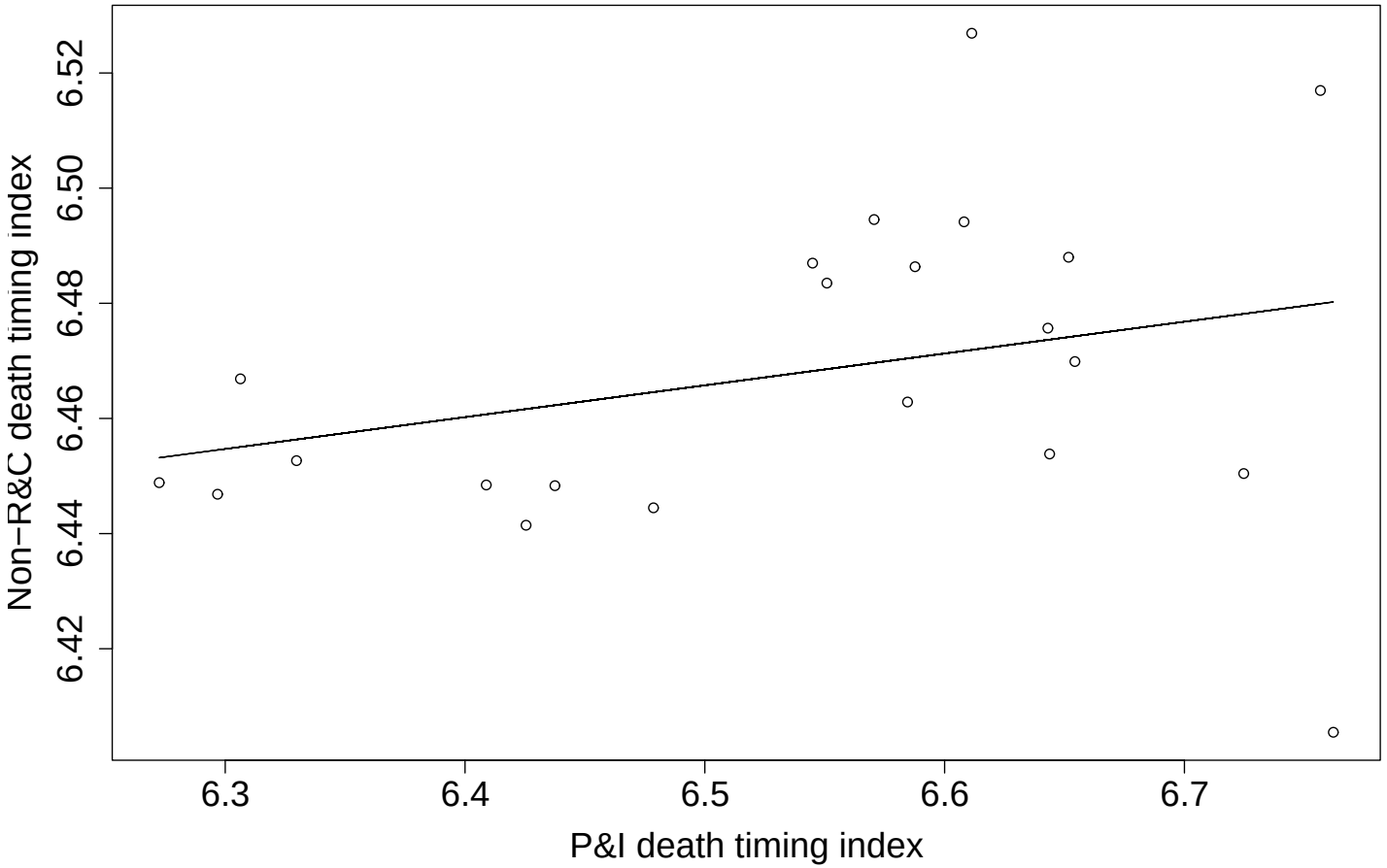
Mortality time series



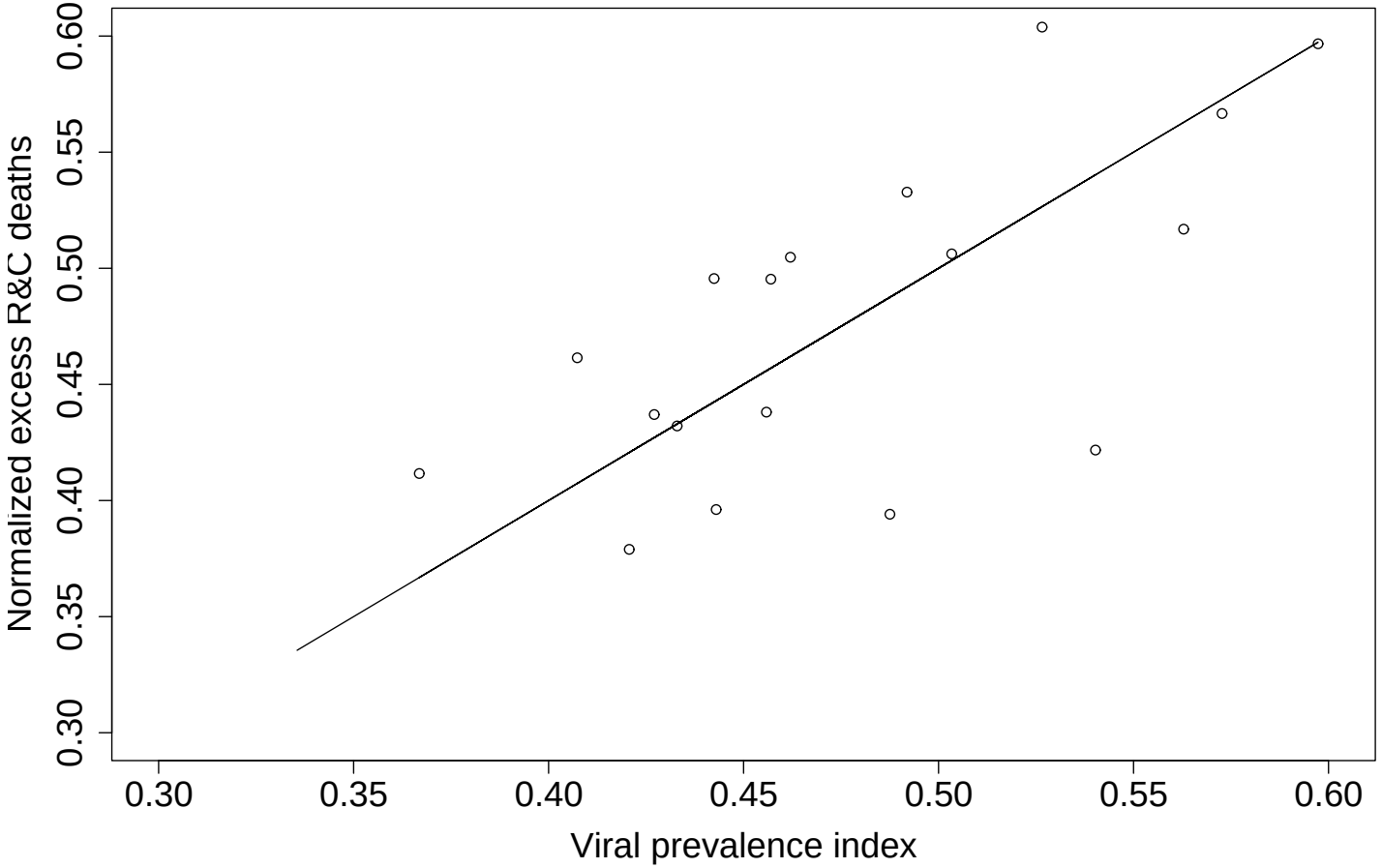
Timing of winter peak: Other R&C vs. P&I ($P < 0.001$)



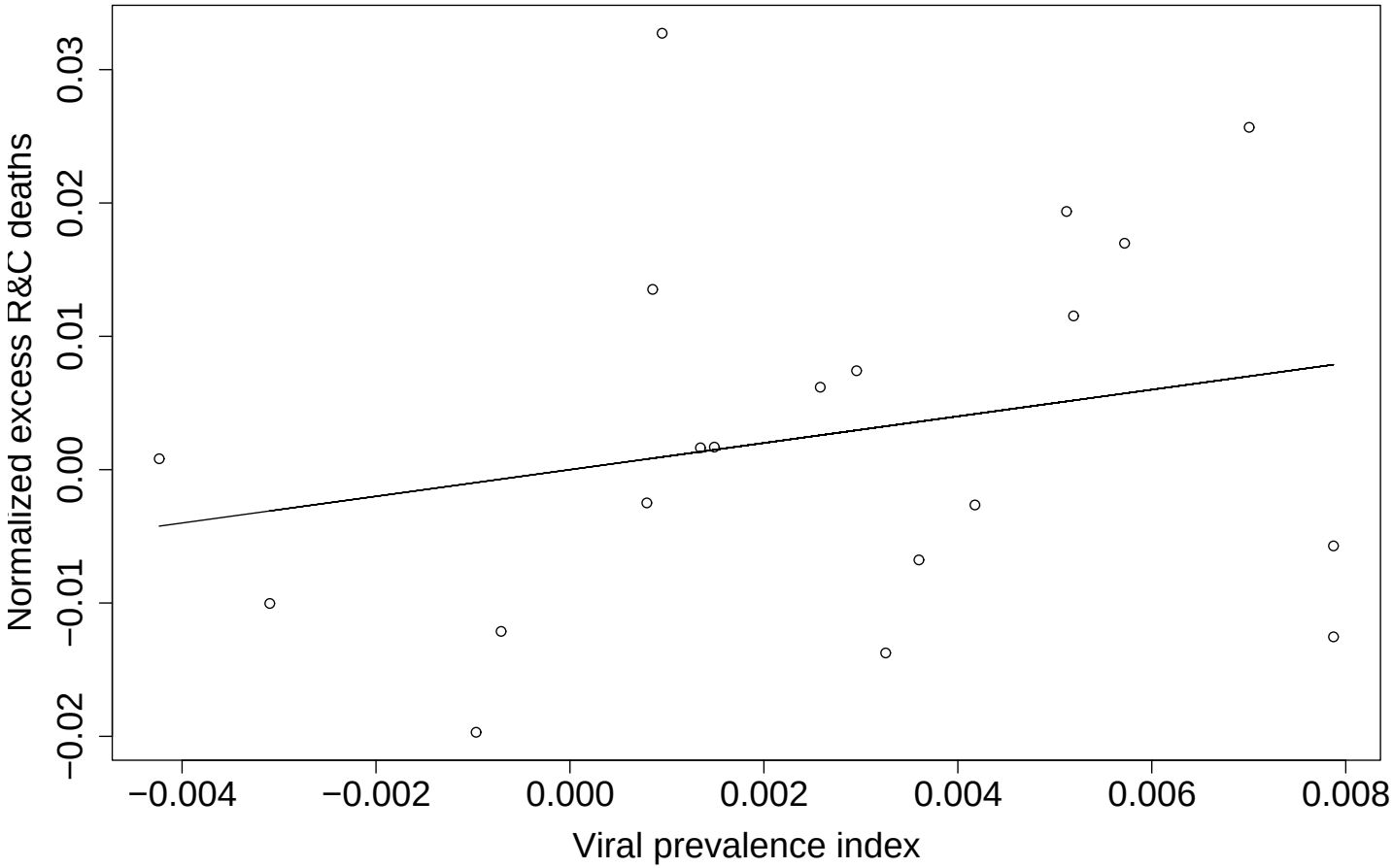
Timing: Non-R&C vs. P&I ($P > 0.1$)



Total R&C mortality vs. virus prevalence ($P < 0.001$)



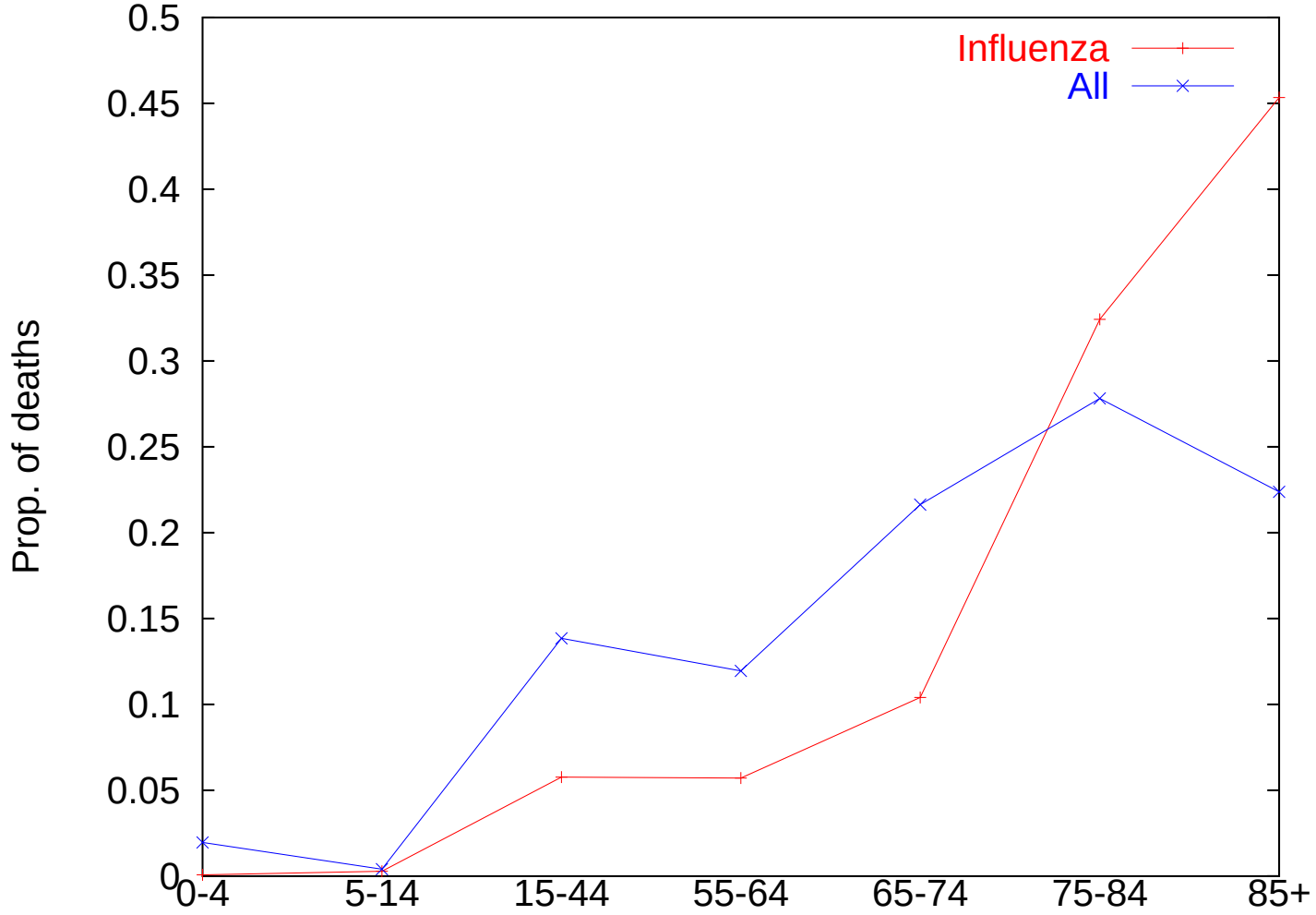
Other mortality vs. virus prevalence ($P > 0.1$)



Annual mortality

Cause	P&I	Other R&C	Non R&C
H3N2	11861***	12317*	1124
H1N1	1531*	1673	226
B	4018***	4068	775
Total	17411***	18058*	2126
Adj. R^2	0.8405	0.2785	-0.1189

Age distribution



Burden summary

- Multiple-cause mortality data is a useful tool
- R&C deaths with no mention of P&I are correlated with P&I deaths:
- Annualized approach supports less conservative excess-mortality or weekly regression approaches.
- Cold temperature does not affect mortality *on an annual time scale.*
- Influenza really is killing people indirectly, on an annual time scale. Years of life lost are considerable, but less per life than for other causes.

Ongoing work

Inferring pandemic risks by comparing current, historical mortality data.

Spatiotemporal patterns for mortality data base. Linking to viral strain surveillance where possible.

For example: when antigenically similar influenza strains circulate in consecutive years, can we detect a different spatial pattern of deaths?

Overview

- Mortality burden of influenza
- **Evolution and quasispecies structure**
- Simple, individual-based models

Quasispecies structure and the antigenic evolution of Influenza A

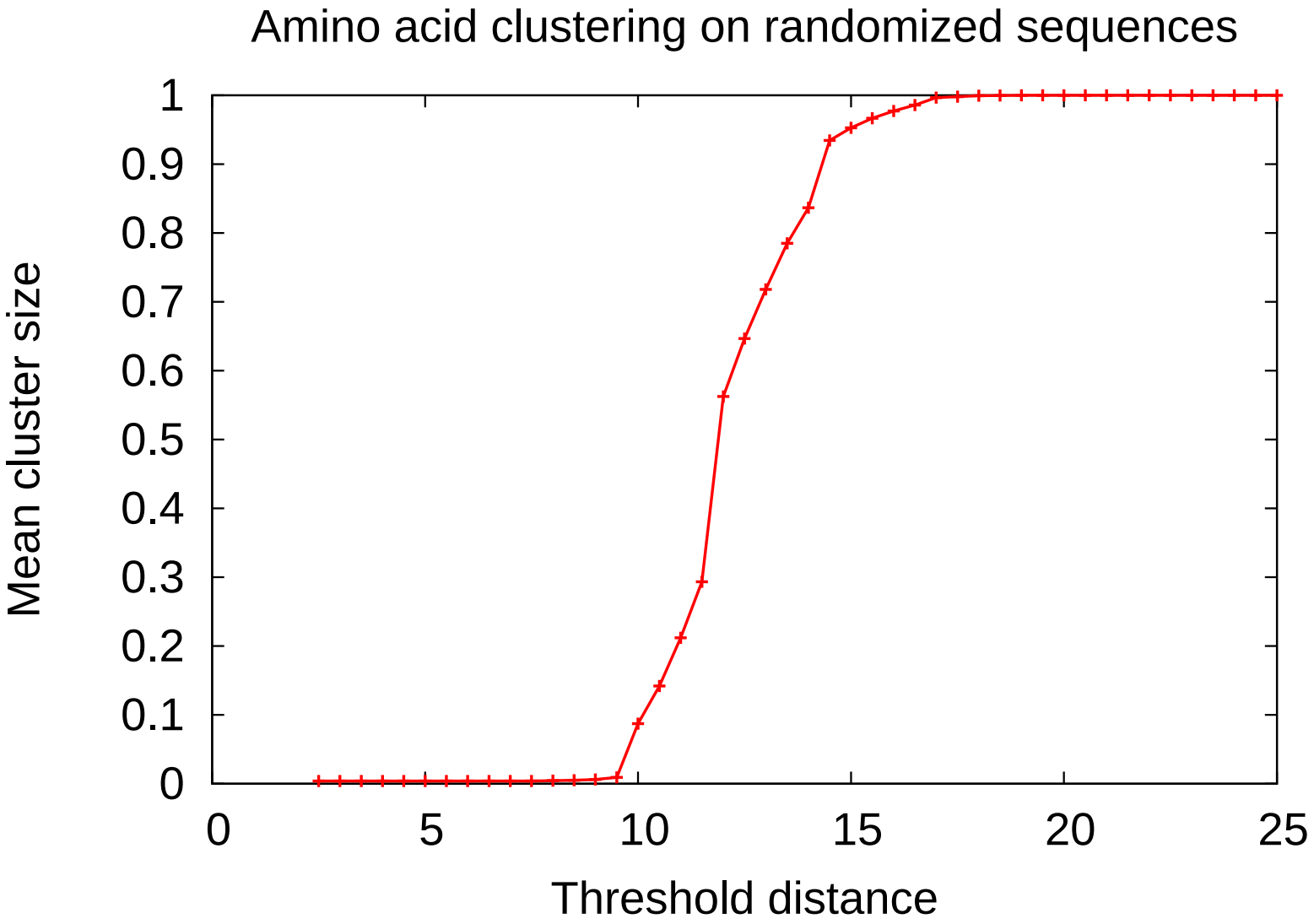
Plotkin, Dushoff and Levin; PNAS 99:6263, 2002

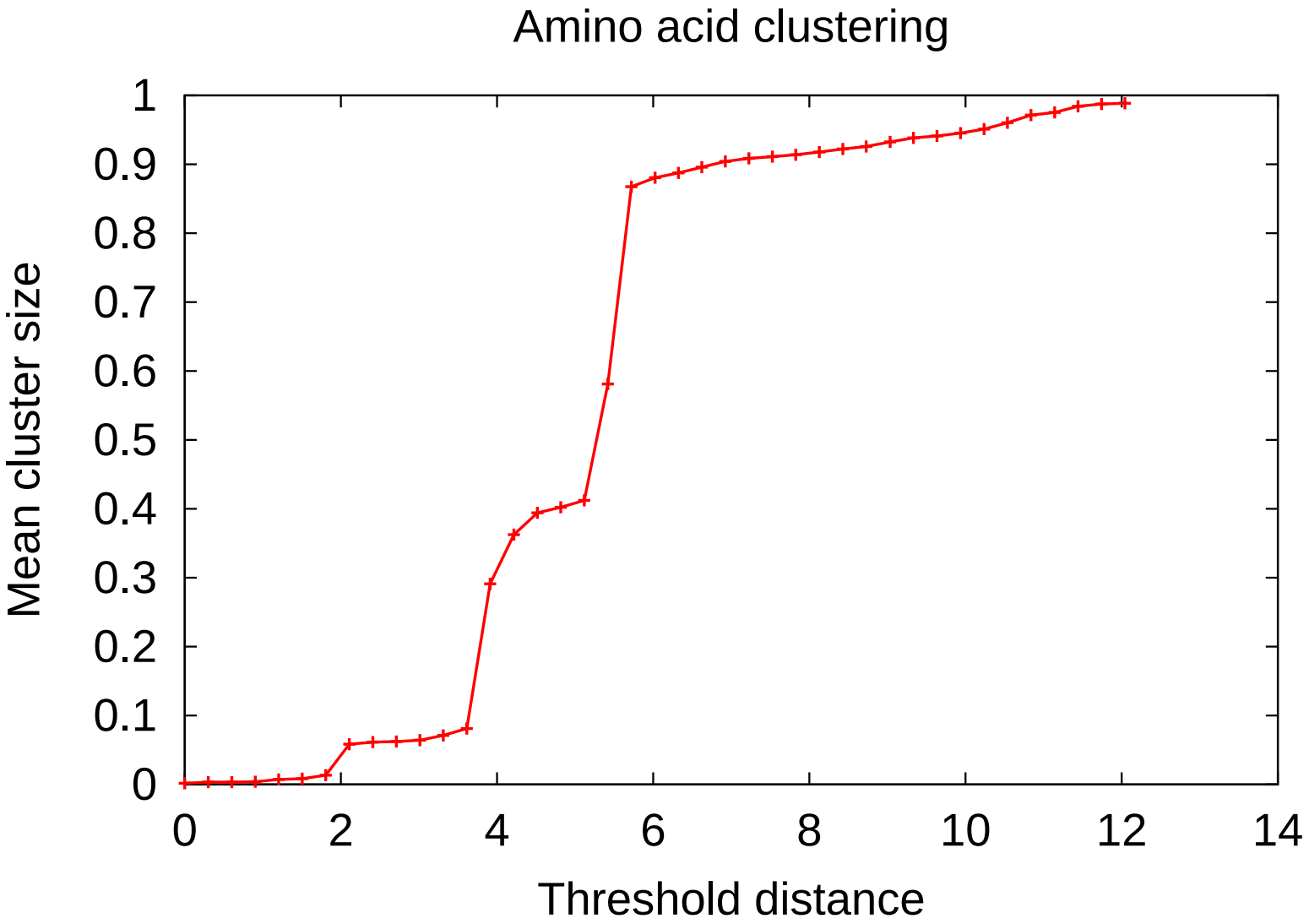
Questions

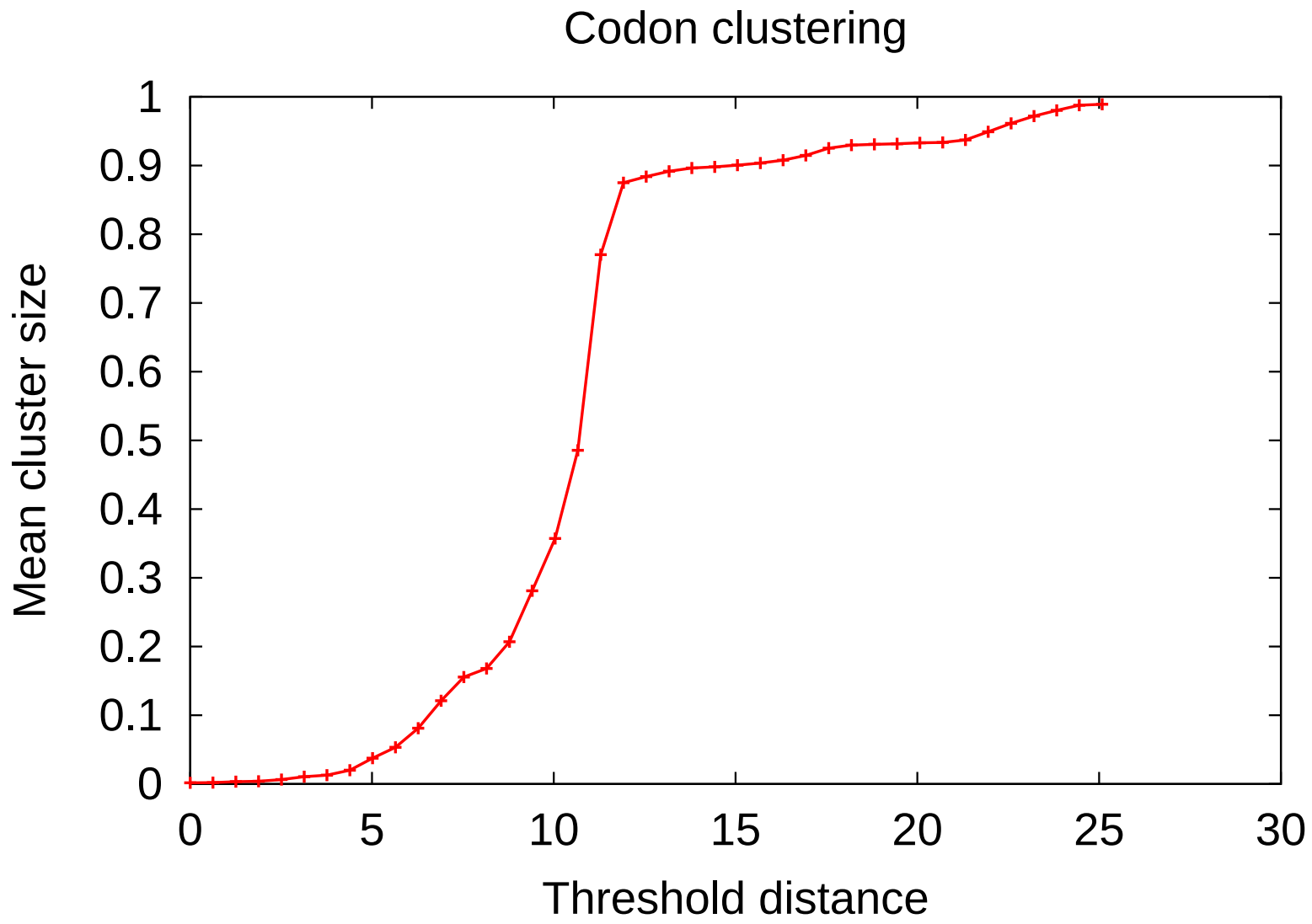
- What do modelers mean by a 'strain'?
- What does strain space look like?
- Do influenza viruses cluster into 'quasispecies'?

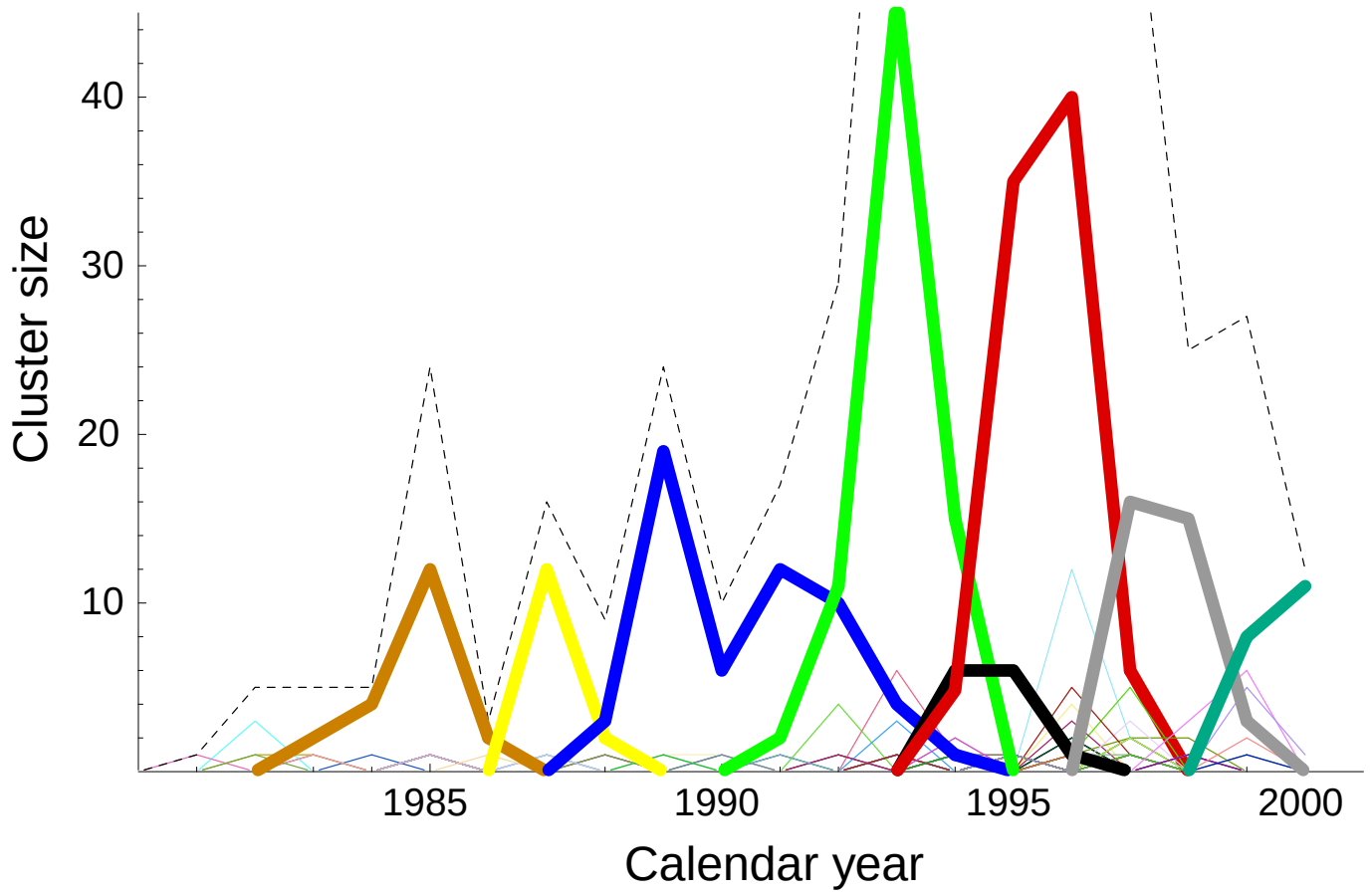
Random clustering technique

- Look at sequences in amino-acid space
- Examine random clusters at different length scales
- Look for scales at which clusterings are stable; these are natural clusterings


Amino acid clustering on randomized sequences
Mean cluster size
Threshold distance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
5
10
15
20
25

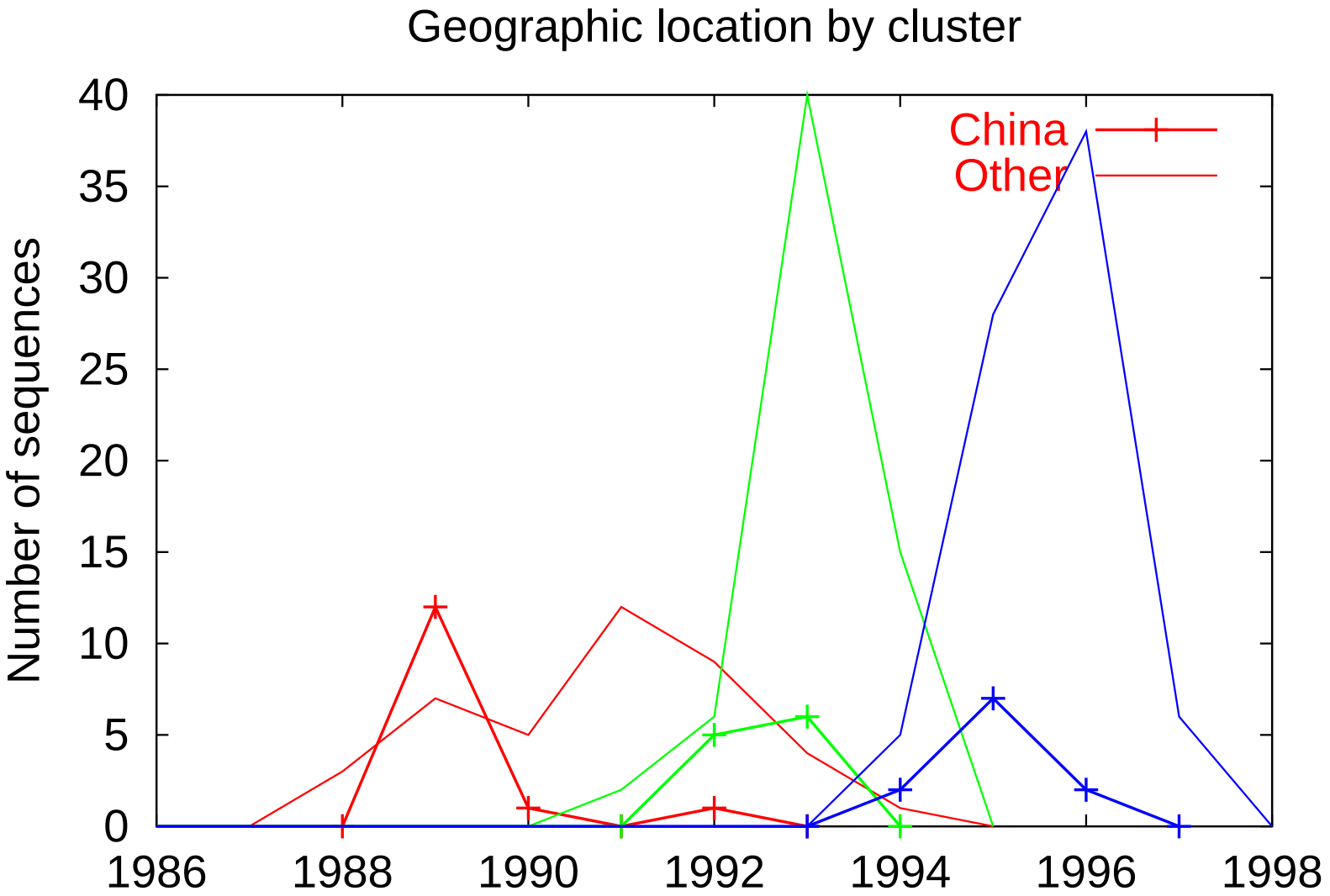

Amino acid clustering
Mean cluster size
Threshold distance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
2
4
6
8
10
12
14

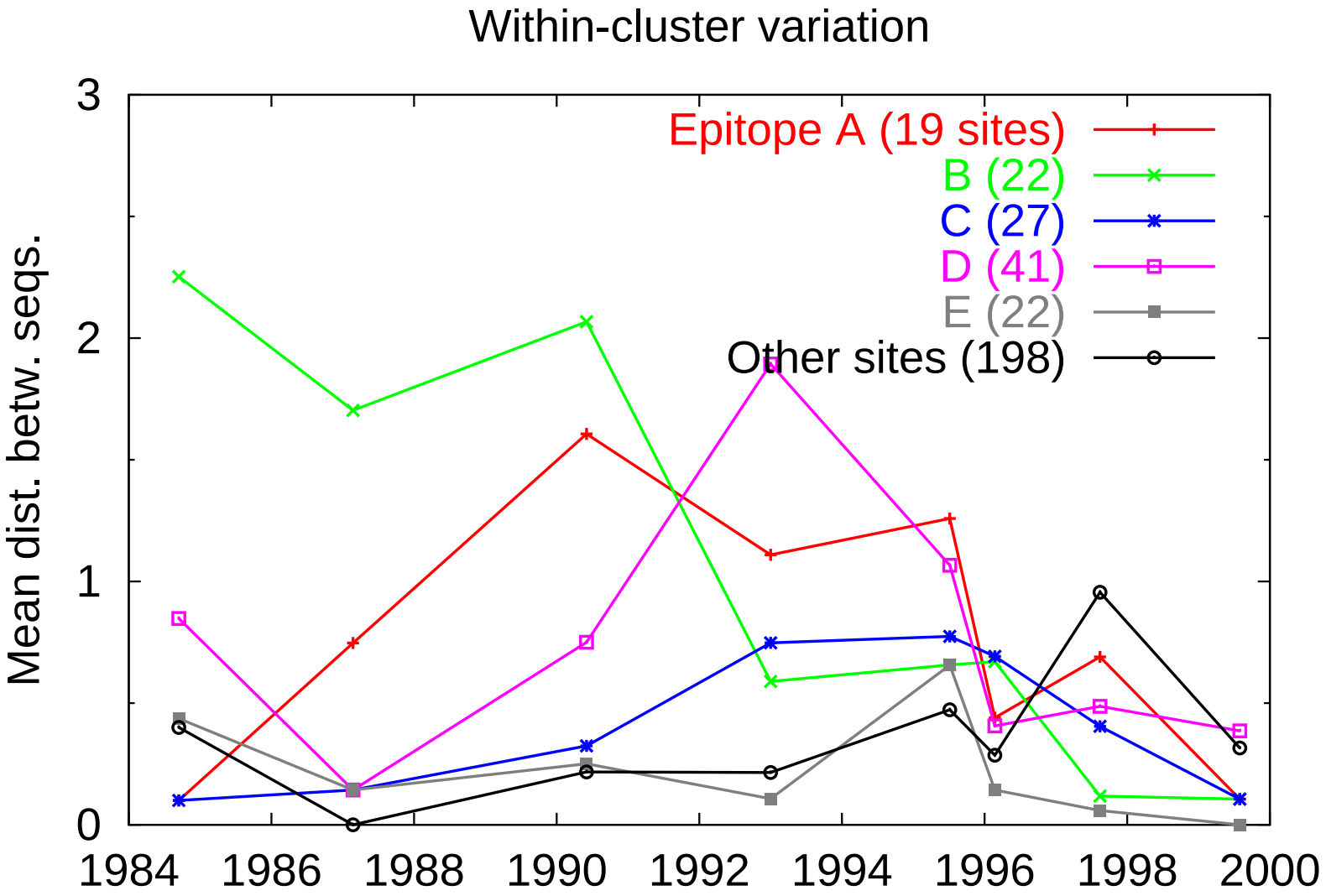

Codon clustering
Mean cluster size
Threshold distance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
5
10
15
20
25
30

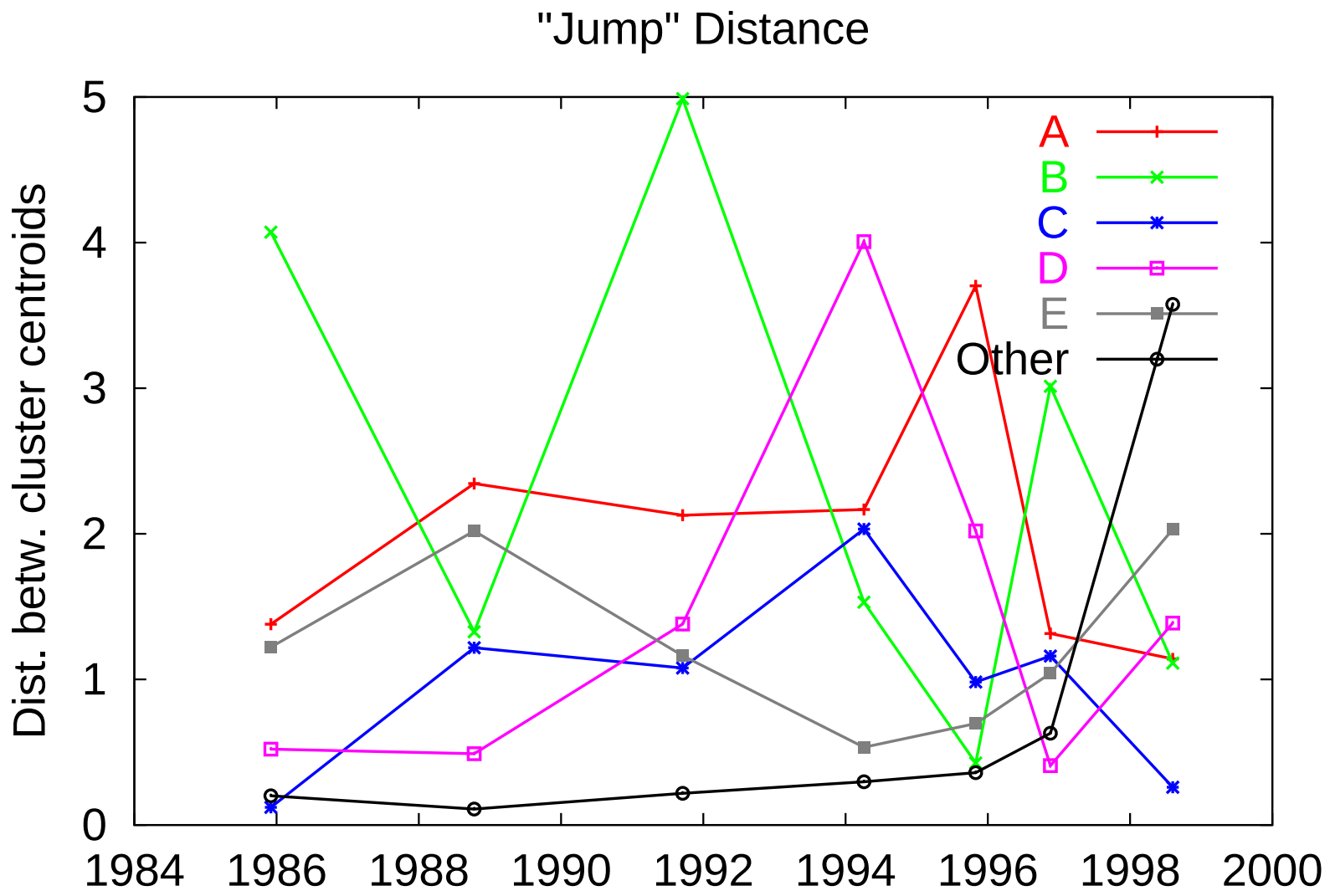

Cluster size
40
30
20
10
1985
1990
1995
2000
Calendar year

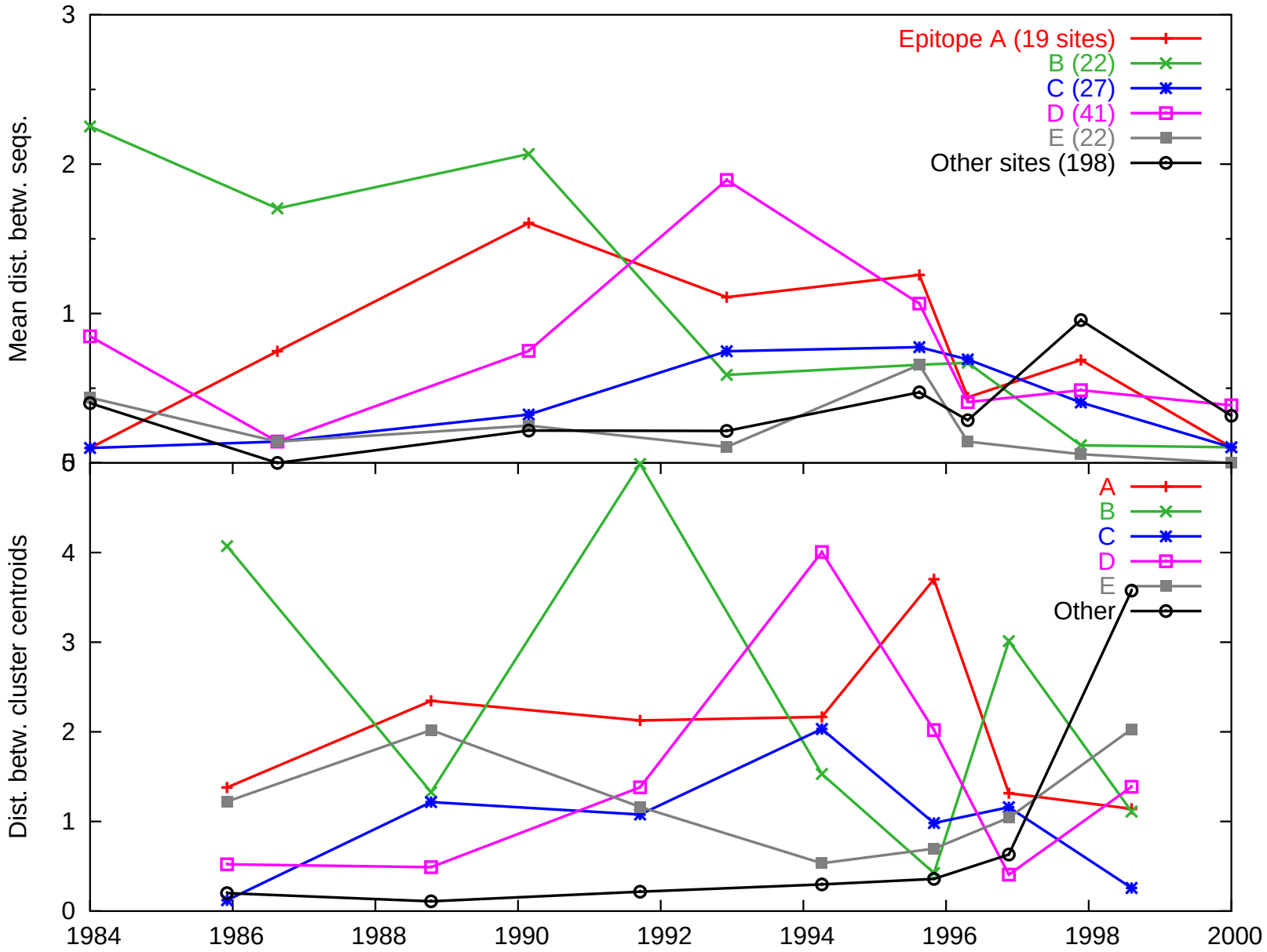
Clusters through time

- Quasispecies have limited temporal range
- Dominant quasispecies replace each other on a time scale of 2–5 years
- Evolution is linear over this time span in amino-acid space


Geographic location by cluster
Number of sequences
40
35
30
25
20
15
10
5
0
1986
1988
1990
1992
1994
1996
1998
China
Other


Within-cluster variation
Mean dist. betw. seqs.
0
1
2
3
1984
1986
1988
1990
1992
1994
1996
1998
2000
Epitope A (19 sites)
B (22)
C (27)
D (41)
E (22)
Other sites (198)


"Jump" Distance
Dist. betw. cluster centroids
0
1
2
3
4
5
1984
1986
1988
1990
1992
1994
1996
1998
2000
A
B
C
D
E
Other


Mean dist. betw. seqs.
Epitope A (19 sites)
B (22)
C (27)
D (41)
E (22)
Other sites (198)
Dist. betw. cluster centroids
A
B
C
D
E
Other
0
1
2
3
4
1984
1986
1988
1990
1992
1994
1996
1998
2000

Clustering summary

- Sequences are clustered in amino-acid space, forming natural ‘quasispecies’.
- Clusters replace each other on a time scale of 2–5 years.
- Clusters display interesting interactions with antibody-combining regions (epitopes).
- Formal clustering methods have potential for predicting the direction of influenza evolution.

Ongoing work

Structural clustering

Use stereochemical properties, location of side chains.

Work on homology modeling.

Investigate more advanced clustering techniques.

- Evaluating random cluster-size curves, and constructing corresponding clusterings.
- Paramagnetic clustering.

Methods of characterizing whole molecules.

Overview

- Mortality burden of influenza
- Evolution and quasispecies structure
- **Simple, individual-based models**

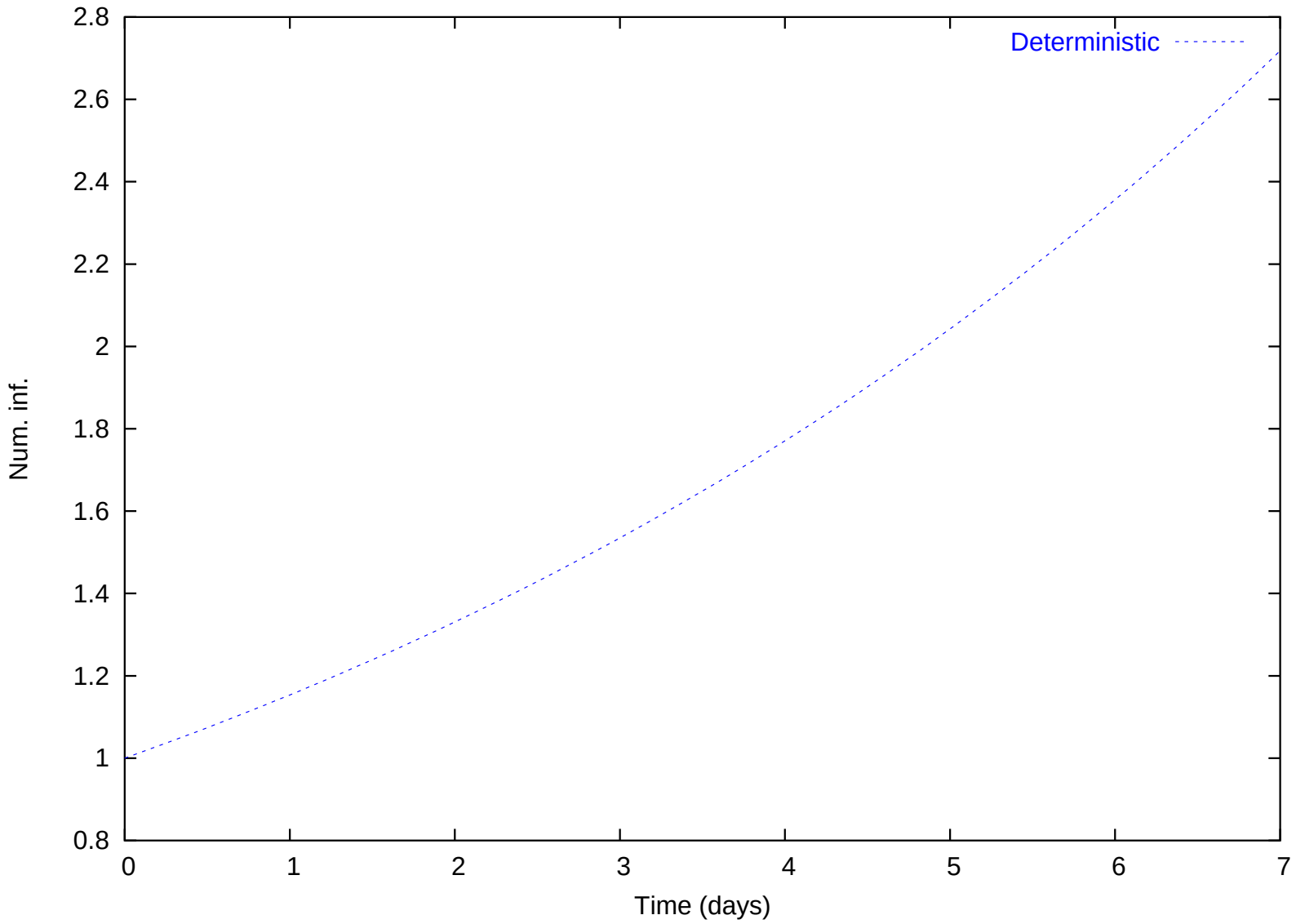
Influenza modelling

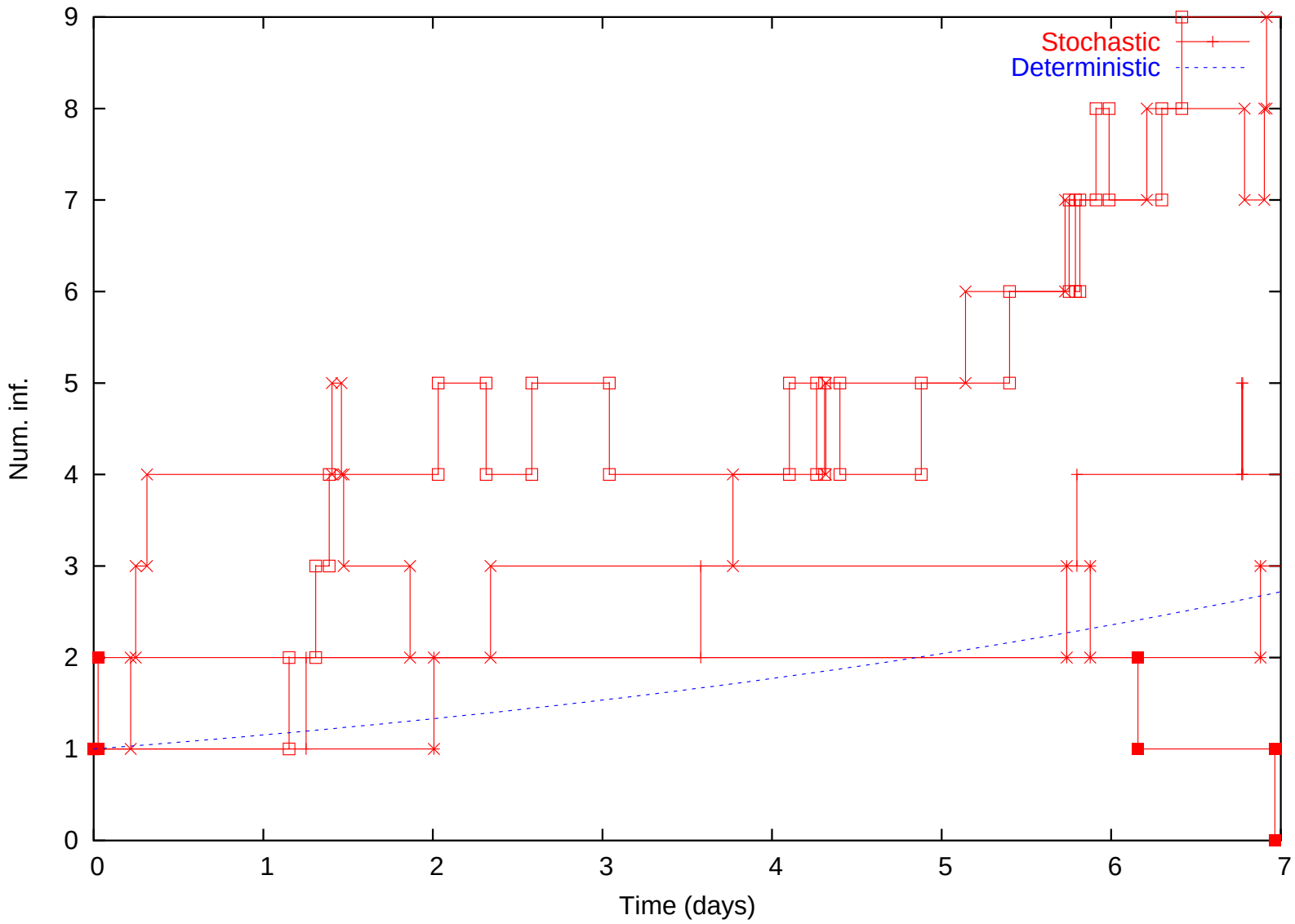
- How do strains interact, and what does this tell us about drift and shift?
- Why does influenza incidence show such marked seasonal oscillations?
- Understanding simple stochastic models as a step towards navigating more complex, realistic models.

Event-based modelling

Levin, Dushoff and Plotkin (Math. Biosci. 188:17, 2004); Dushoff, Plotkin, Levin and Earn (PNAS 101:16915, 2004)

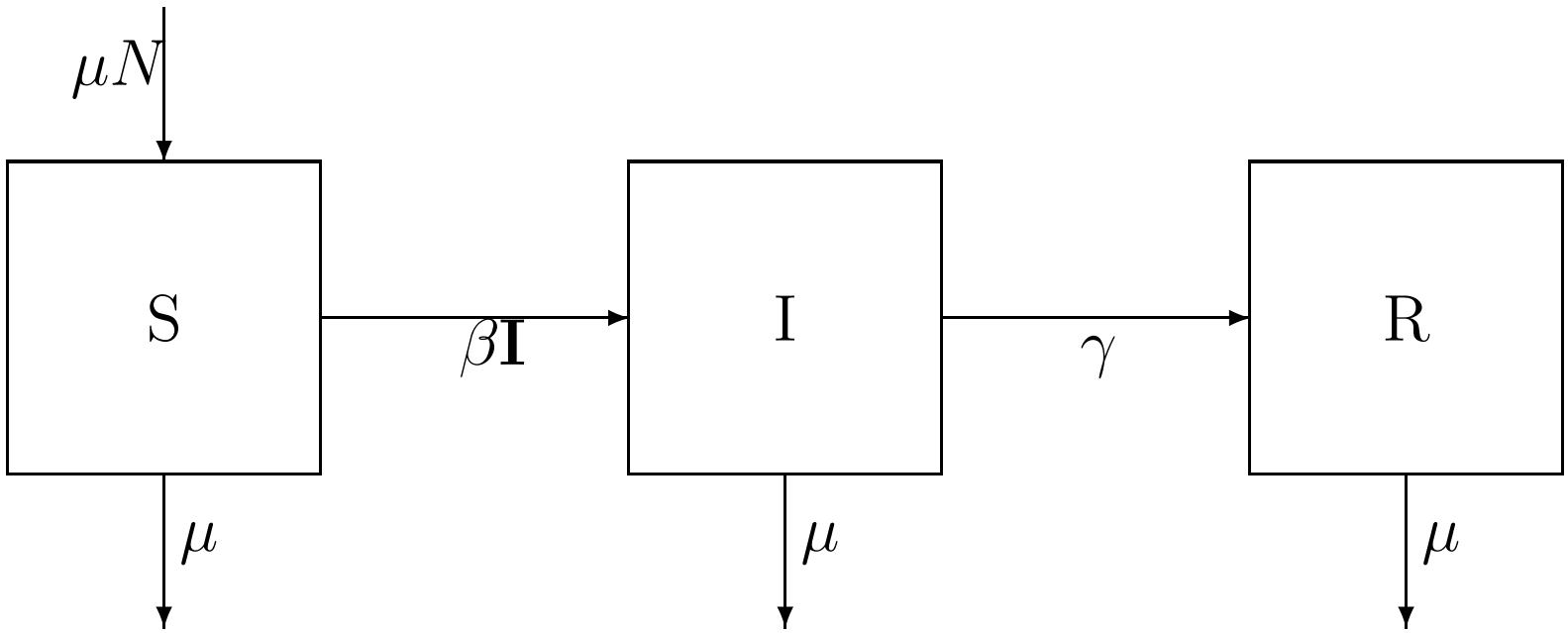
- Instead of modelling continuous changes through time, model discrete events.
- Fundamentally stochastic: when does a particular person become infected? What particular mutations occur?
- Stochasticity due solely to treating individuals as individuals is called *demographic stochasticity*


Deterministic
Num. inf.
Time (days)
2.8
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0
1
2
3
4
5
6
7


Stochastic
Deterministic
Num. inf.
Time (days)
0
1
2
3
4
5
6
7
8
9

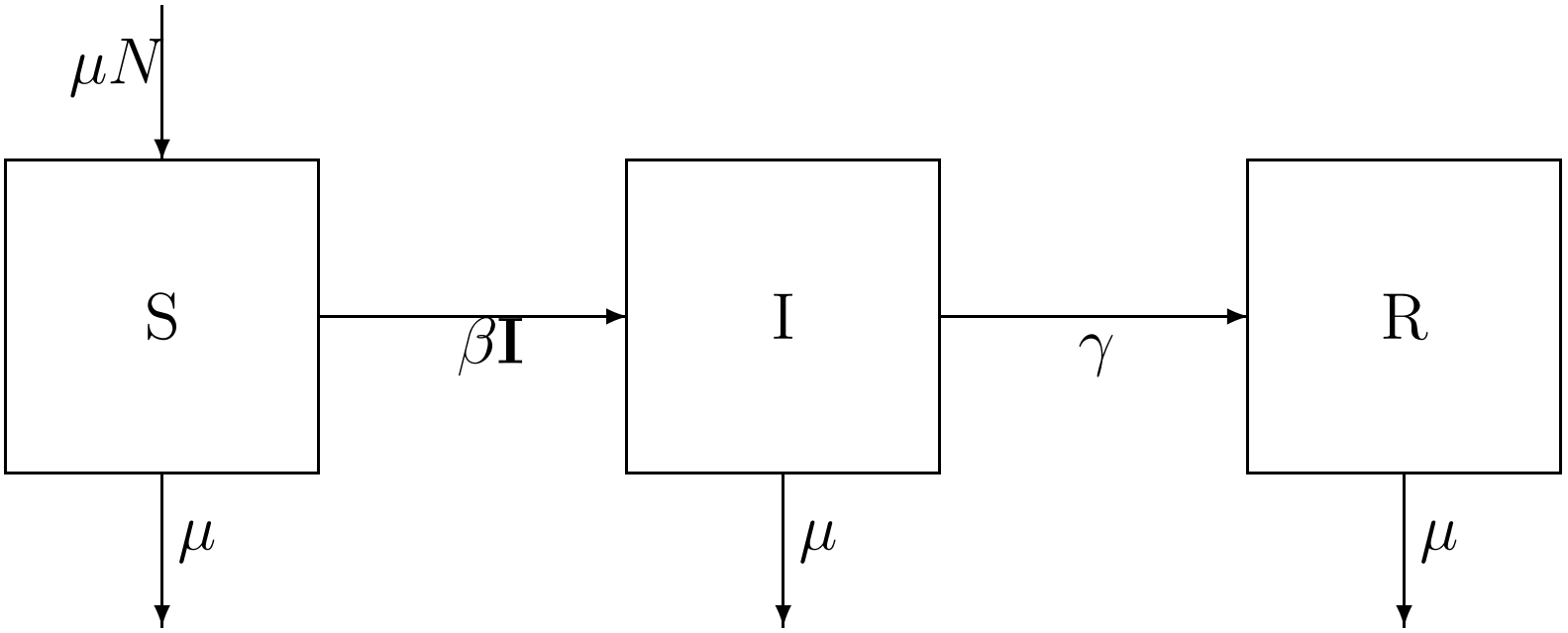
A stochastic potential-transmission event

SIR model with birth and death

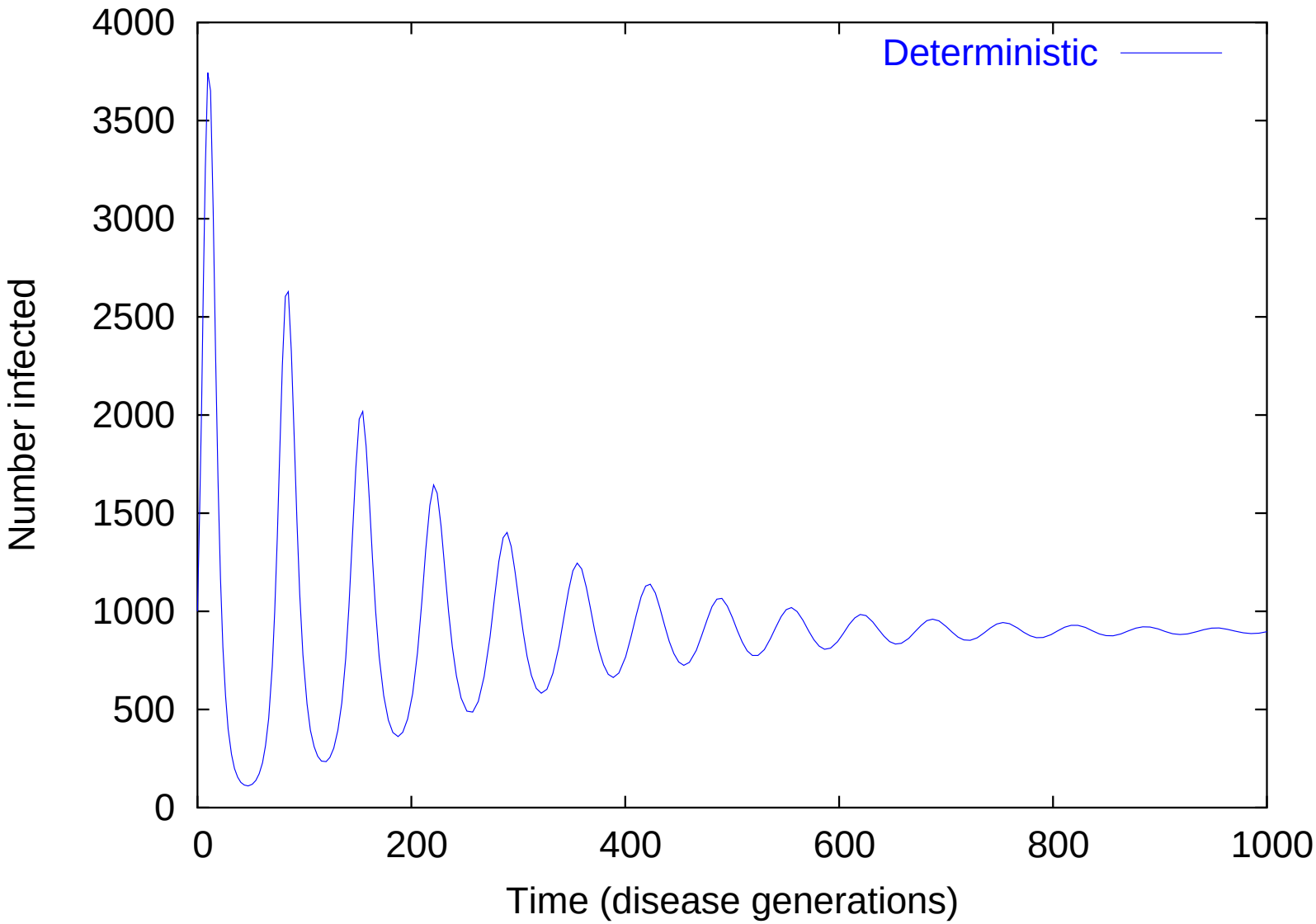


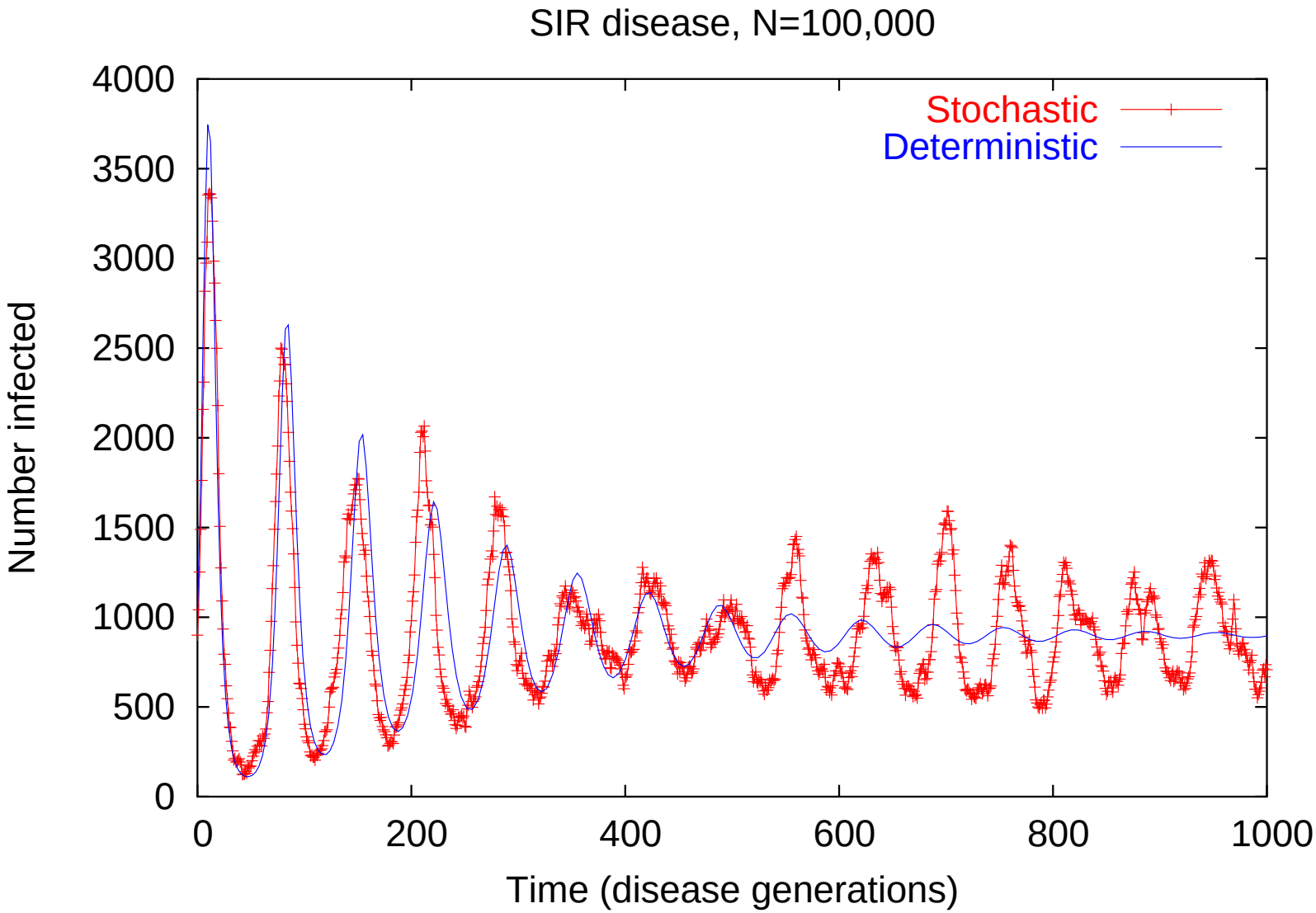
$$
\begin{aligned}
dx/dt &= \rho(1-x) - R_0 xy \\
dy/dt &= R_0 xy - y
\end{aligned}
$$

SIR model with birth and death



Event	**transition**	**rate**
Rebirth	$I \to S,\ R \to S$	ρ
Infection	$S \to I$	$R_0 I/N$
Recovery	$I \to R$	$1-\rho$


Deterministic
Number infected
Time (disease generations)
0
500
1000
1500
2000
2500
3000
3500
4000
0
200
400
600
800
1000


SIR disease, N=100,000
Stochastic
Deterministic
Number infected
Time (disease generations)
0
500
1000
1500
2000
2500
3000
3500
4000
0
200
400
600
800
1000

Demographic stochasticity

Persistent oscillations even in large populations, when there is long-term immunity.

Frequency well predicted by the 'natural frequency' of the deterministic system.

Size of (small) oscillations well predicted by diffusion approximations.

No good theory for time to extinction, even in simplest case.

Seasonality

Influenza (and other diseases) are strongly seasonal, but no obvious strong underlying mechanism.

Possible mechanisms include:

- cold stress on immune system
- crowding
- virus persistence in environment
- school terms

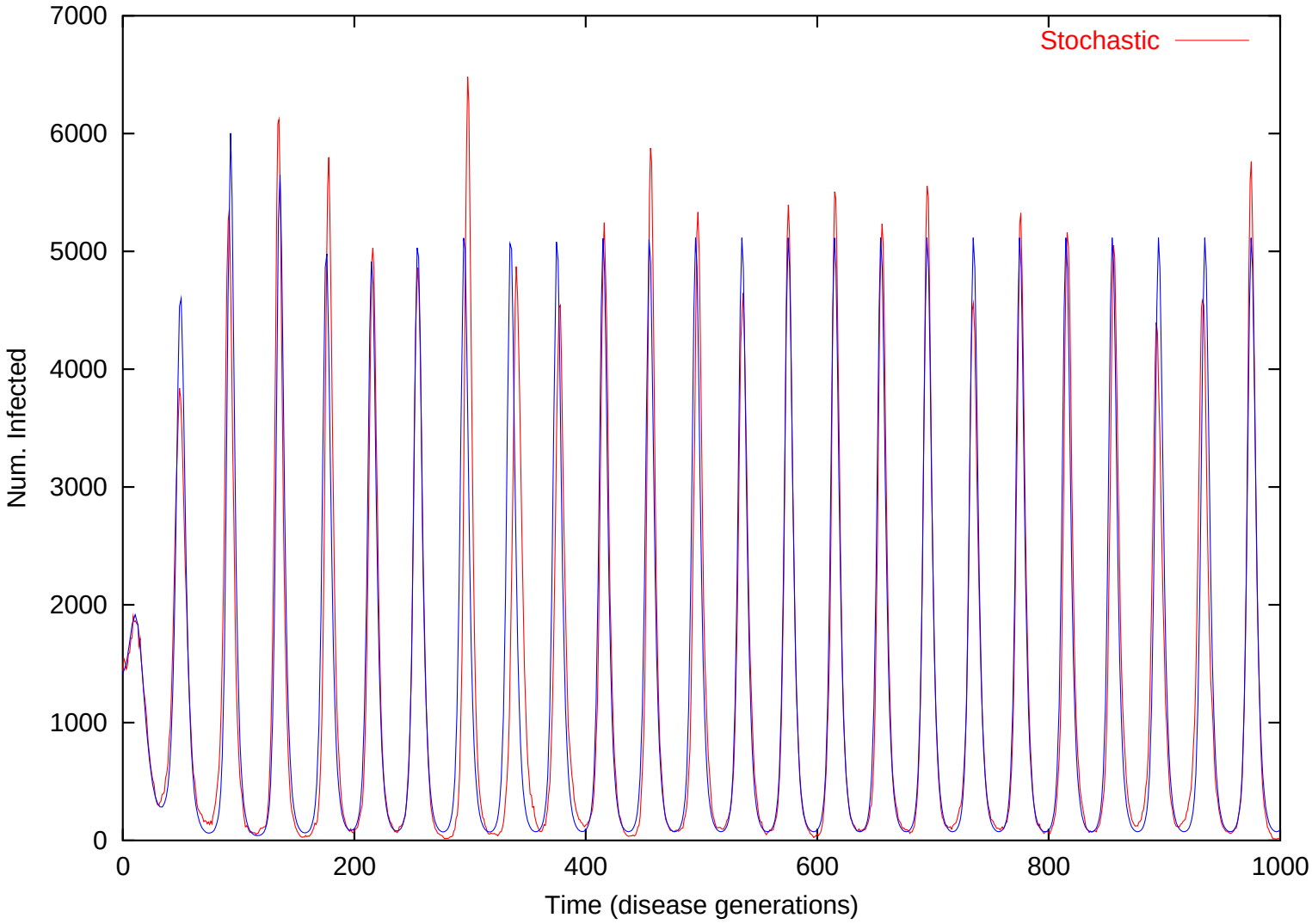
But why is cause not evident?

Effects of intrinsic dynamics on seasonality of disease

Can a small underlying seasonal variation in transmission lead to large seasonal variation in disease incidence?

- Demographic stochasticity
- 'Natural' period of disease dynamics

Moderate forcing ($\pm 10\%$), strong resonance (11 mo.)

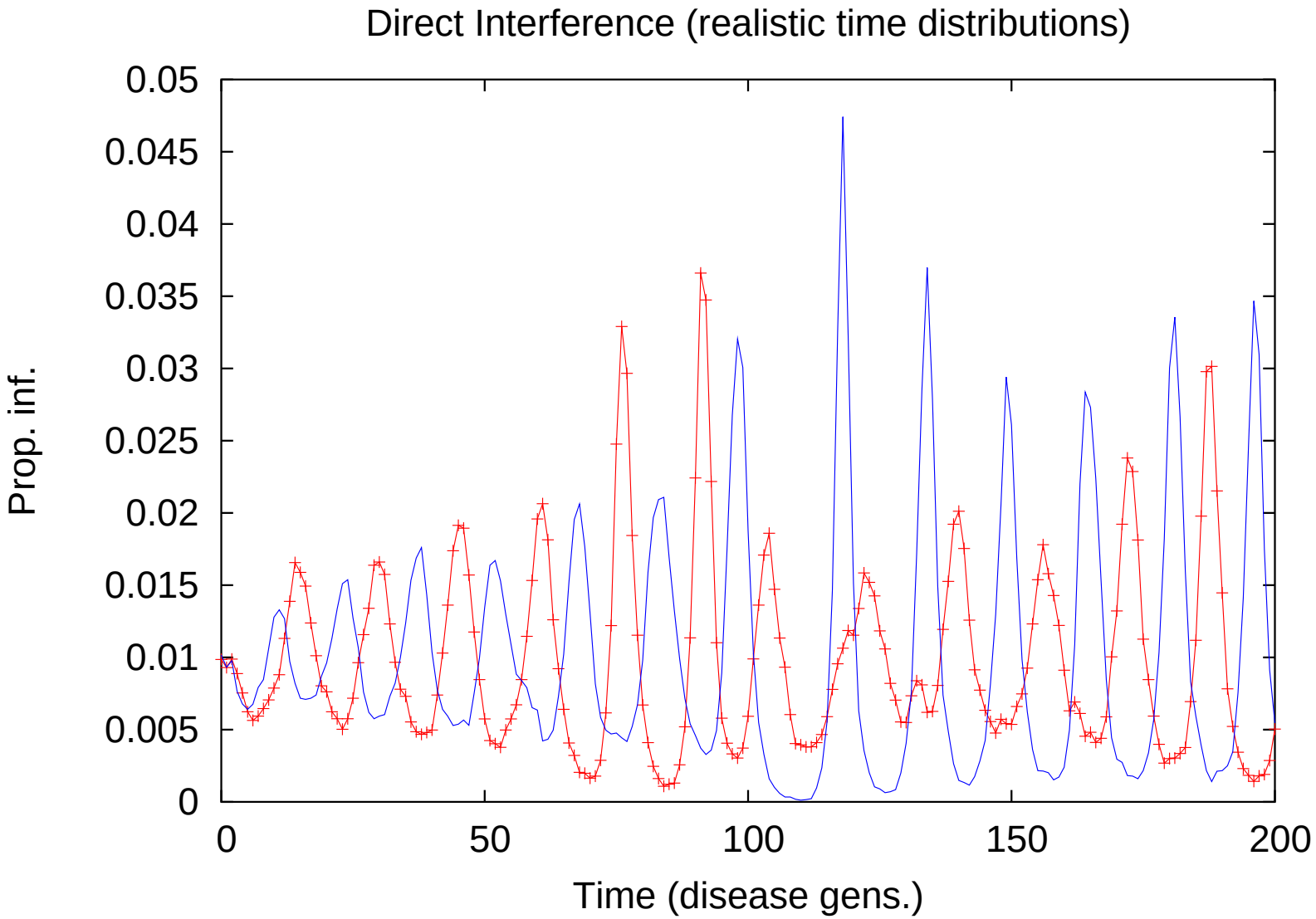


Demographic stochasticity and strain competition

Does demographic stochasticity change predictions about strain interactions?

Are dynamics sensitive to standard mathematical assumptions?

What does the exponential distribution really mean?


Direct Interference (realistic time distributions)
Prop. inf.
0.05
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
0
50
100
150
200
Time (disease gens.)

Modeling summary

- Stochastic models are necessary for understanding important biological phenomena.
- Better understanding of *simple* stochastic models can help us to understand the behavior of more complex ones.
- Modelling must be guided by data analysis, and vice versa

Ongoing work

Basic scaling laws for disease systems with demographic stochasticity

Invasion, persistence and competition in realistic models.

- Understanding time delays
- Population and spatial heterogeneities

Conclusion

Tying things together

Can scaling be extended to more realistic time distributions, to allow more efficient simulations?

Can we mathematically summarize community and regional interactions that will enable us to scale from small populations to the globe?

Can we understand what makes particular HA structures functional and effective at escaping existing antibodies?

Can we expand our understand of immunodominance, and the importance of T-cell and innate immunity?

Thanks

Co-authors (Earn, Fraser, Levin, Plotkin, Simonsen, Viboud)

Sergey Kryazhimskiy, Catherine Macken, Ben McMahon, Walt Mankowski, Ellis McKenzie, Tom Reichert, Peter Palese, David Smith.

National Institutes of Health (NIGMS, Fogarty Int'l Center)

Academy of Motion Picture Arts and Sciences

This audience