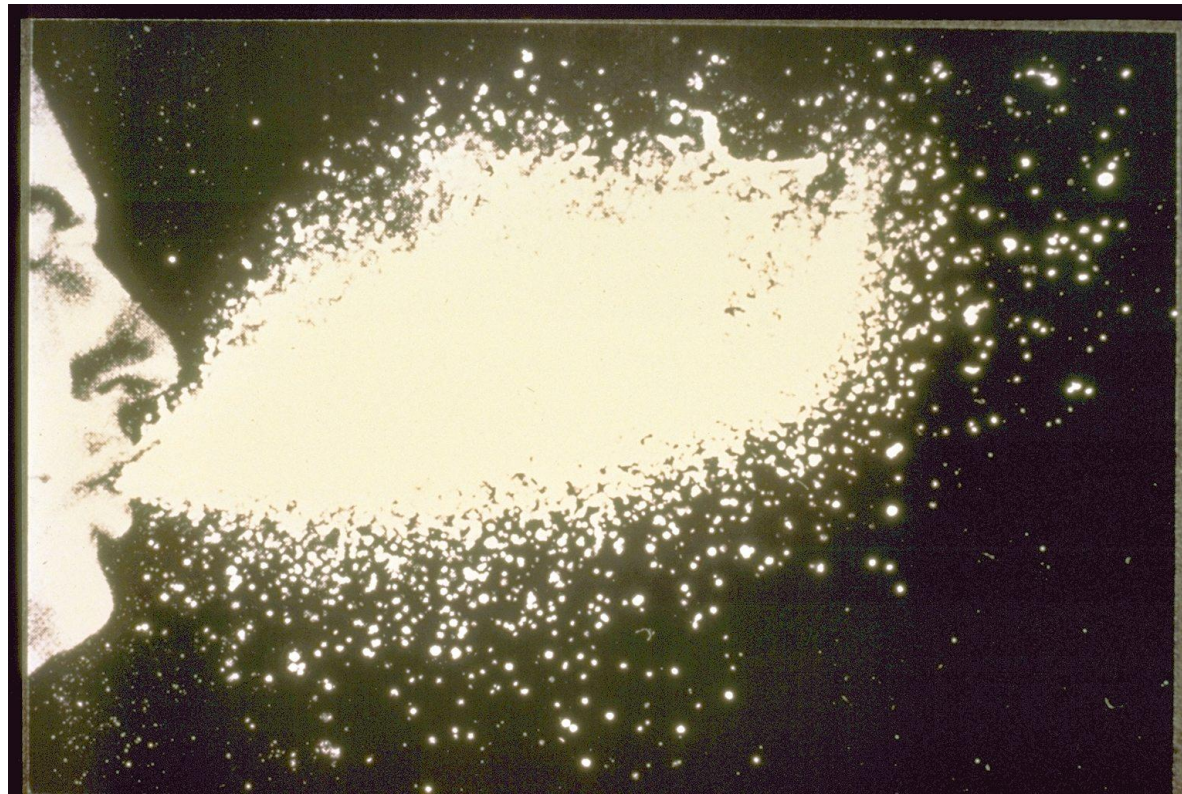


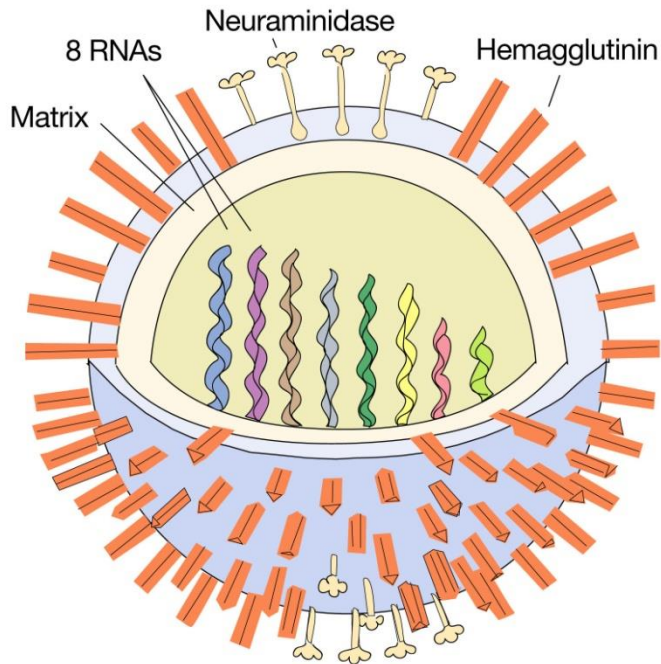
The evolution of infectious disease



Influenza

- **We generally think of the flu as nothing more than a minor annoyance**
- **In an average year, however, the flu kills 20,000 Americans**
- **Occasional global pandemics are much more severe**
 - **The ‘Spanish flu’ of 1918 sickened 20% of the world’s population and killed between 20-100 million people**
 - **Other highly lethal pandemics occurred in 1957 and 1968**

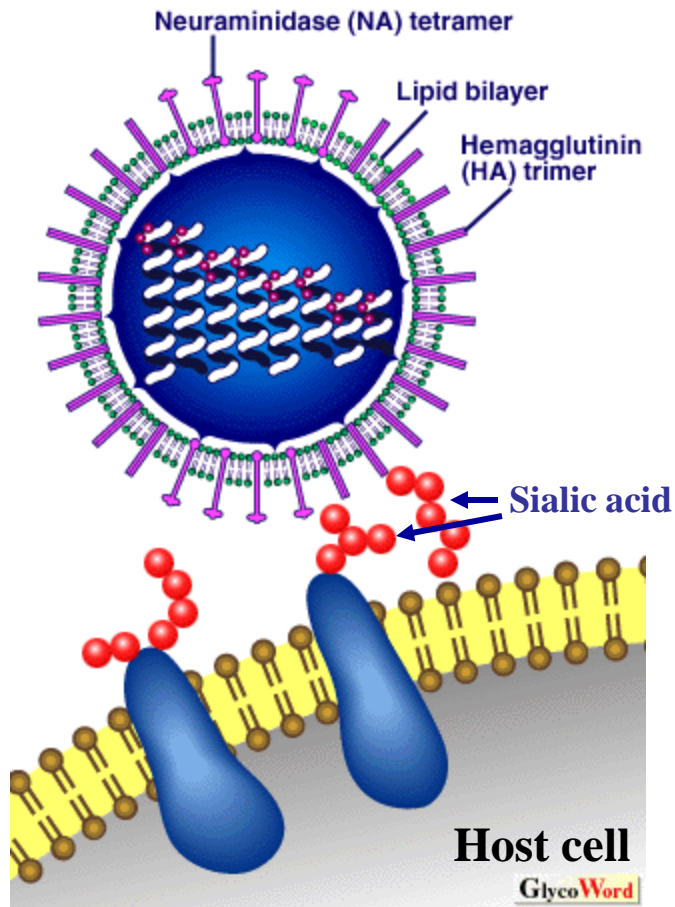
What is Influenza?



Copyright © 2004 Pearson Prentice Hall, Inc.

- **RNA virus**
- **Genome composed of 8 RNA strands**
- **These encode 10 proteins**
- **The predominant coat protein is called hemagglutinin**
- **Hemagglutinin is responsible for binding to host cells**

Influenza infection



- Hemagglutinin binds to sialic acid on the host cell
- This allows the virus to enter the host cell
- Hemagglutinin is also the primary protein recognized by the host immune system
- Hosts that have been previously infected **cannot be reinfected by the same strain**

How, then, can we be infected multiple times?

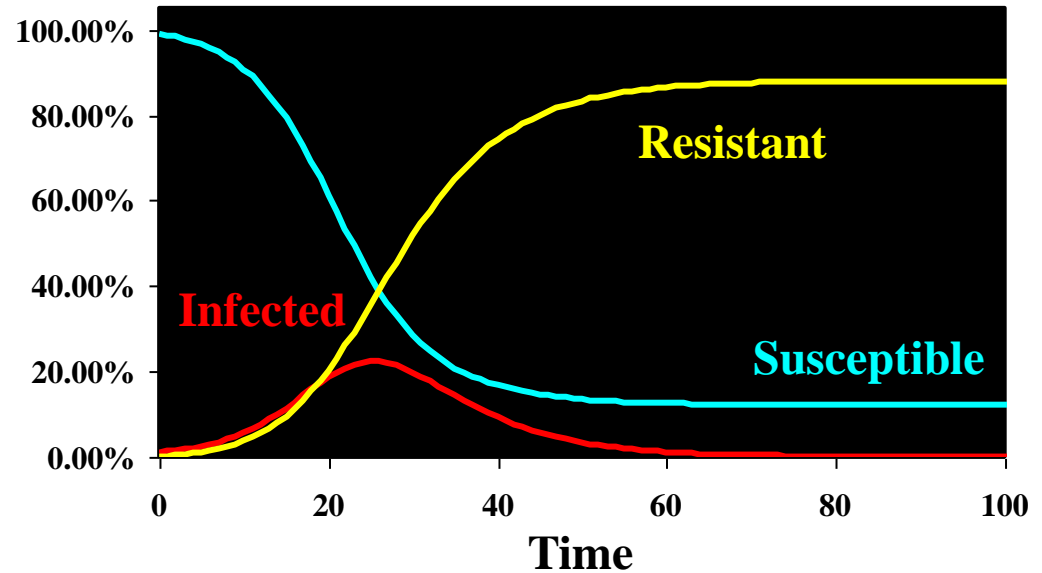
As a strain of influenza spreads, it uses up its finite supply of susceptible hosts

$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

% of host
population
in the three
classes



Where do new susceptible hosts come from?

An evolutionary hypothesis:

- New mutations arise in the Influenza hemagglutinin gene
- Some of these mutations generate hemagglutinin molecules not recognized by the host immune system
- These mutant strains enjoy a renewed supply of susceptible hosts

Testing the evolutionary hypothesis

Fitch et. al. 1991:

- Collected DNA from human Influenza samples that had been stored in freezers
- This collection spanned the years 1968-1987
- Sequenced the **Antigenic sites*** of the hemagglutinin gene
- Measured the **genetic distance** between samples

* Antigenic sites are the specific parts of the protein recognized and remembered by the immune system

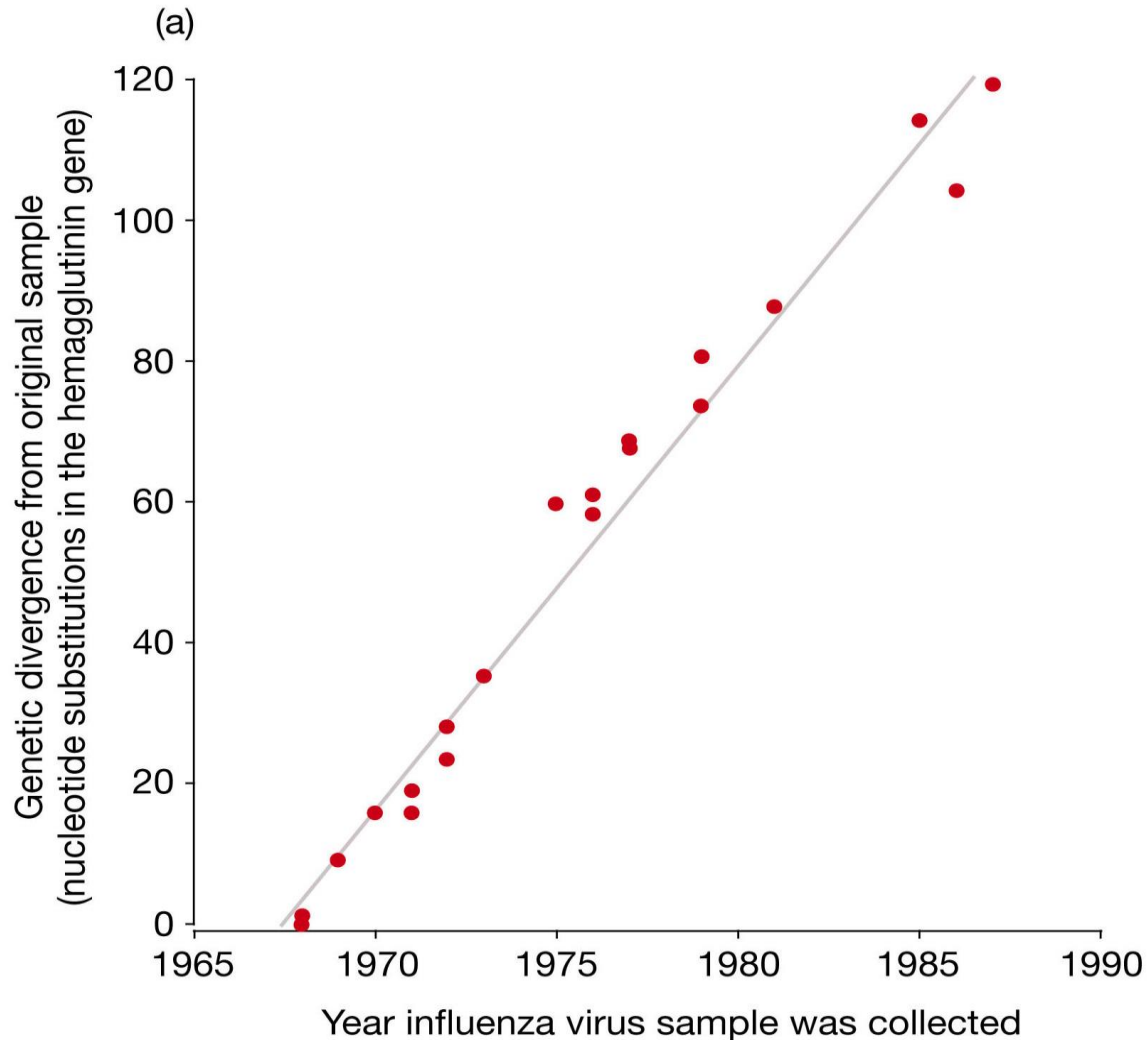
A caricature of measuring genetic distance...

1968 hemagglutinin gene sequence: ...GCAUGCAAAUGGC...
1969 hemagglutinin gene sequence: ...GCAUCCAAAUGGC...
1970 hemagglutinin gene sequence: ...GCAUCCAAUUGGC...
1971 hemagglutinin gene sequence: ...GCAUCCAAUUGAA...

D=1
D=2
D=4

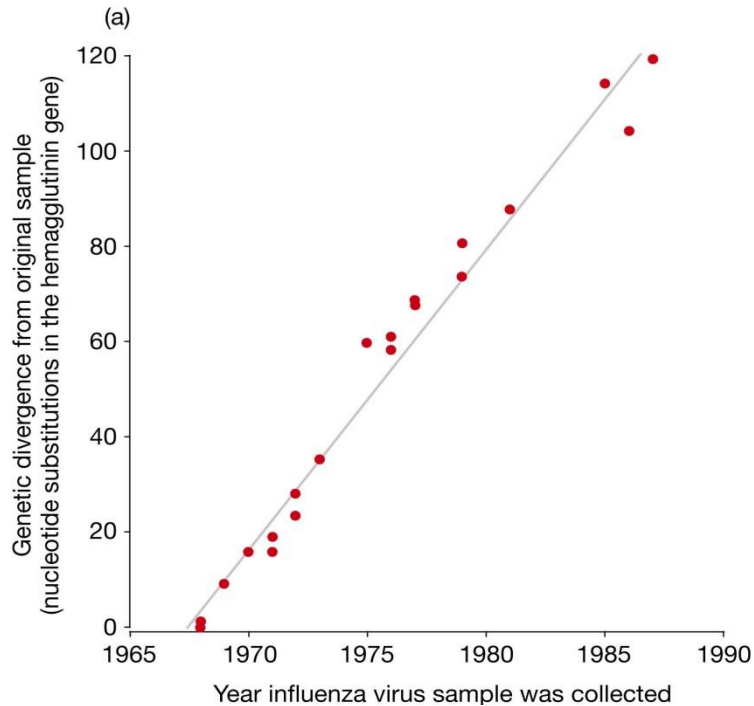
This hypothetical scenario is consistent with the hypothesis that mutant strains of influenza are at a selective advantage

So what did Fitch et. al. (1991) find?



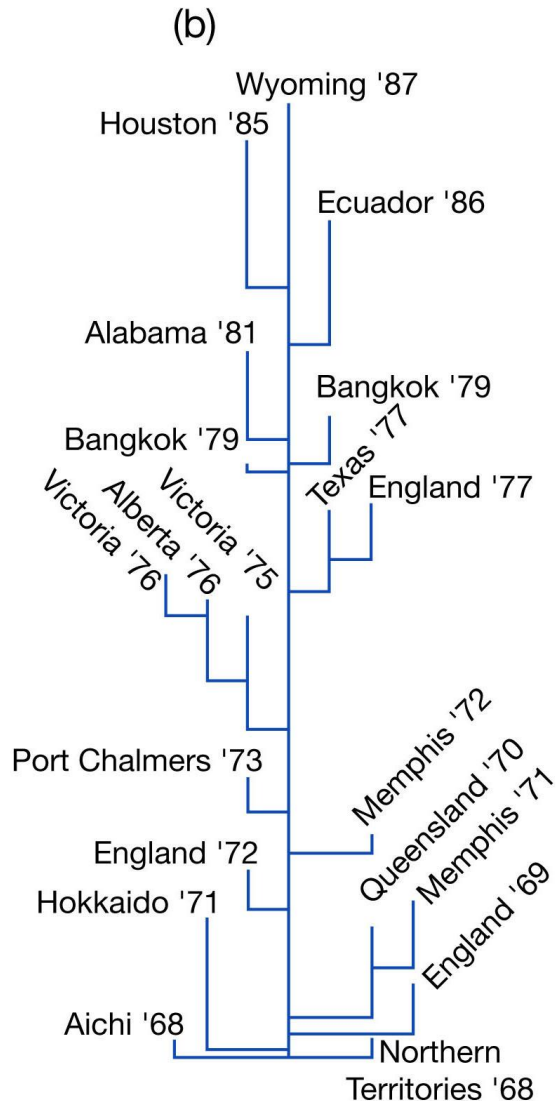
What does this pattern suggest?

Summary of Fitch et. al. (1991)



- DNA sequence of the hemagglutinin gene has diverged over time
- This divergence has occurred at a **rate which is too rapid to be explained by genetic drift** (Bush et. al. 1999)
- Suggests that **genetic divergence of the hemagglutinin gene has been driven by natural selection for immune evasion**

Summary of Fitch et. al. (1991)



- A novel strain of influenza emerges
- Hosts develop immunity, reducing the number of susceptible hosts
- This creates strong selection for new mutant strains not recognized by the host immune system
- Ultimately, this generates a historical pattern of sustained molecular divergence

Applications to vaccine development

- **Flu vaccine works by exposing potential hosts to dead virus**
- **Even though the virus is dead, its protein coat elicits an immune response**
- **This creates a ‘memory’ of the protein coat that allows the host to mount a rapid immune response when exposed to live virus**
- **This immune response is specific to the strain used as the vaccine**
- **Production of large quantities of vaccine takes months**

→ Forecasting the next years strain is critical

Can we use the results from the evolutionary studies to forecast?

Bush et. al. (1999):

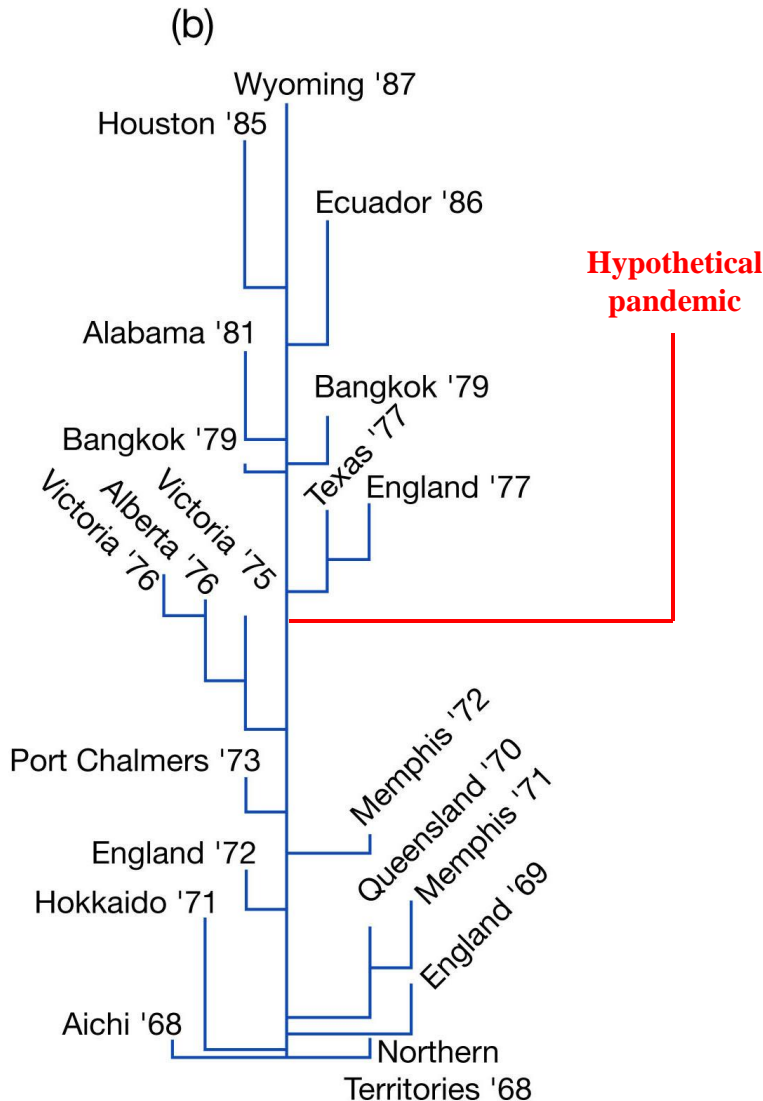
- Reasoned that the surviving strain will be the one with the most mutations in the Hemagglutinin gene (the gene under selection by the human immune system)
- Tested this idea for the previous 11 flu seasons
- Accurately predicted the surviving strain 9/11 times!
- Allows an appropriate vaccine to be developed in time for the next years flu season

What causes pandemics?

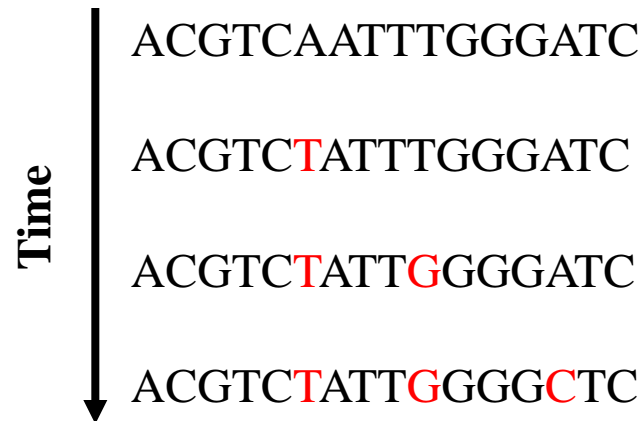
Three major pandemics have been recorded in recent history:

- 1918 ‘Spanish flu’ 500,000 deaths in the United States
- 1957 ‘Asian flu’ 70,000 deaths in the United States
- 1968 ‘Hong Kong flu’ 34,000 deaths in the United States
- 2009 ‘Swine Origin Influenza Virus’ > 18,000 deaths worldwide in first year and had infected more than 43,000,000 people within the U.S. alone by April 2010

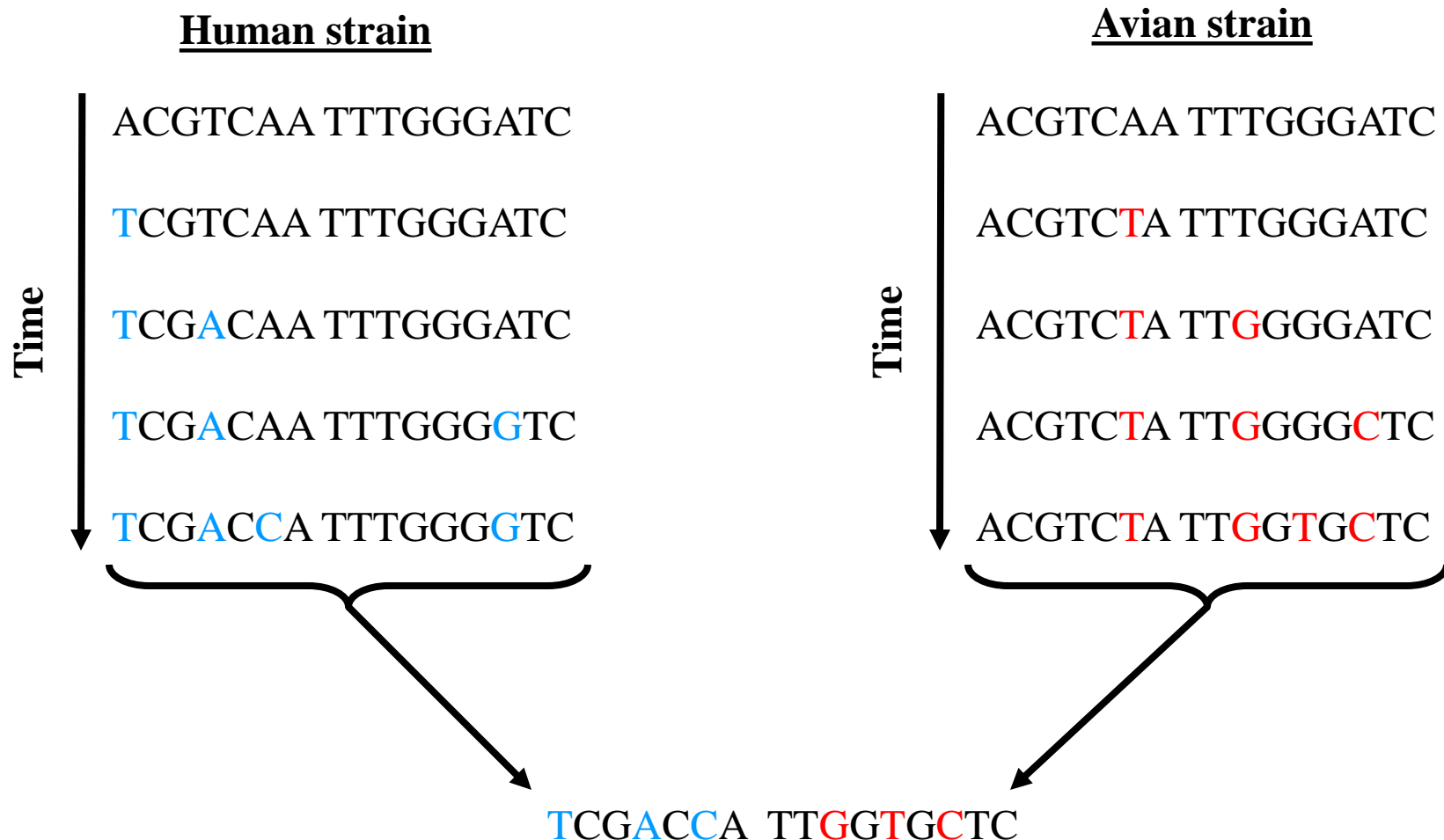
New mutations?



- New mutations generally have small effects
- New mutant strains will not be sufficiently different to evade immune recognition altogether
- Suggests that pandemics may arise from a process other than mutation



Reassortment?



This new recombinant differs from any other human strain at 3 positions all in one step!

Reassortment?

- **Reassortment can create new virus genotypes that differ radically**
- **These novel strains (genotypes) may be so divergent that they are not readily recognized by the host immune system**
- **As a result, these novel strains may lead to pandemics**

How could this occur?

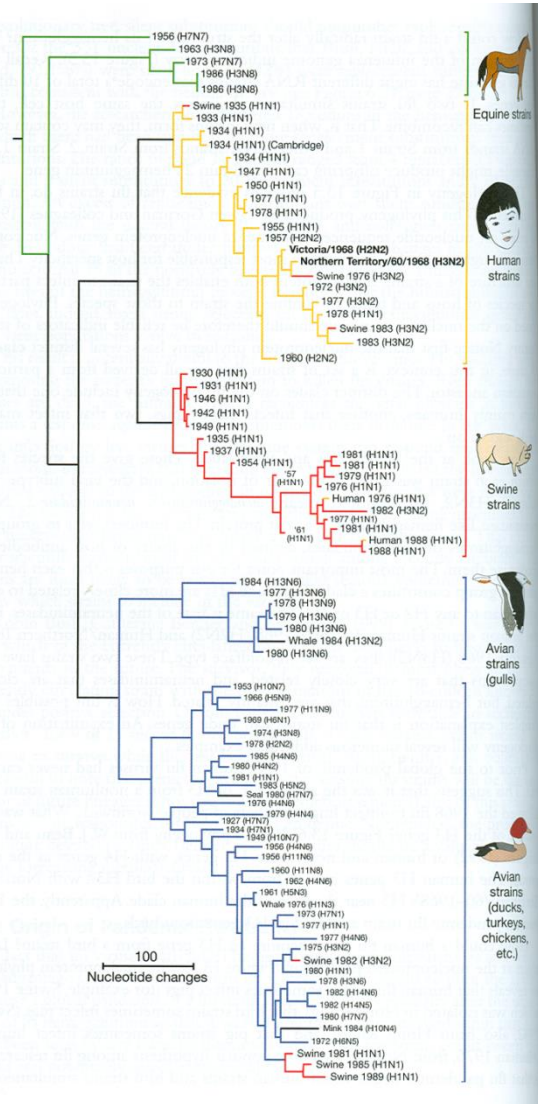


- **Human Influenza requires receptors found only in human respiratory cells**
- **Avian Influenza requires receptors found only in intestinal cells**
- **Pigs have both types of receptors in their trachea**
- **Pigs serve as ‘mixing vessels’ where coinfection of individual host cells can occur**
- **Thus, pigs enable reassortment of avian and human strains**

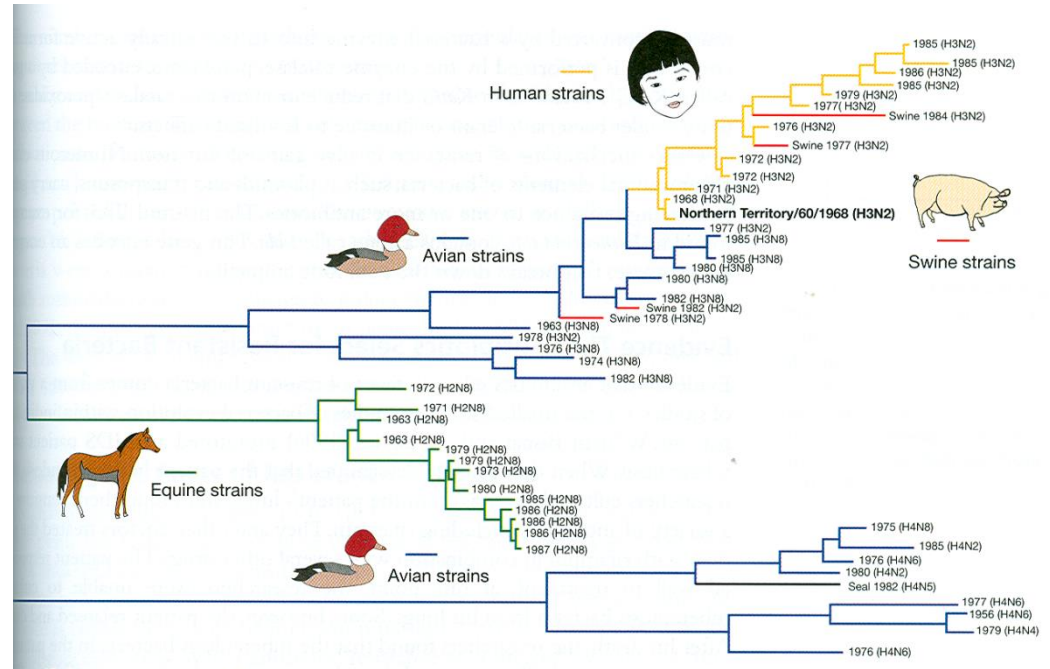
Is there evidence that this has occurred?

Gorman et. al. (1992) & Bean et. al. (1992)

Based on nucleoprotein sequence



Based on hemagglutinin sequence



- Nucleoprotein sequences cluster together by host group
- Hemagglutinin sequences do not
- The gene phylogenies are not congruent!

Summary of Influenza evolution

- **Influenza evolution seems to be driven by natural selection for host immune system evasion**
- **Most of the time, this occurs through the origin of novel mutant strains which increase in frequency and do not lead to pandemics**
- **Occasional pandemics appear to be caused by reassortment of different virus strains**

Practice Problem: Applying the SIR model

Imagine that an emerging infectious disease has been identified in the human population of the United States. Scientists from the CDC have studied this viral disease intensively during its first several weeks and determined that $\beta = .24$ and $\gamma = .12$. They have also determined that the entire human population is likely to be initially susceptible to this disease. Use this information to answer the following questions:

A. Derive a general mathematical expression for the minimum population size that will lead to an epidemic, starting from the standard SIR model which assumes that the rate of

change in the density of infected individuals per unit time is: $\frac{dI}{dt} = \beta SI - \gamma I$.

B. What would the minimum human population density have to be for this emerging infectious disease to lead to an epidemic?

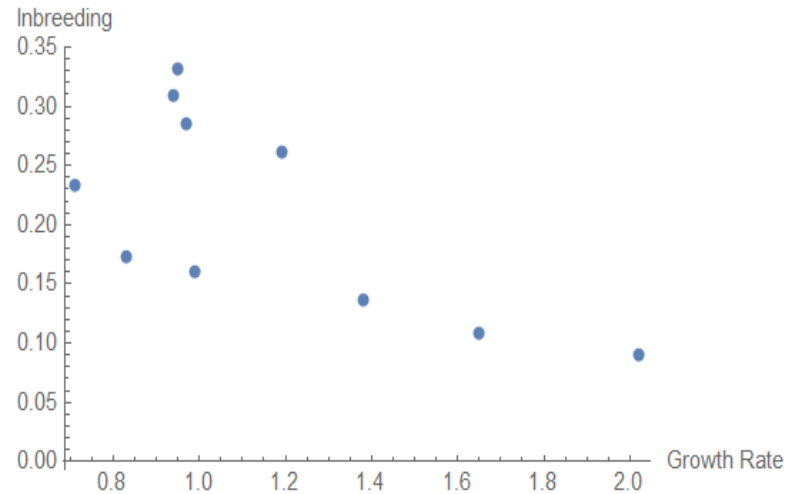
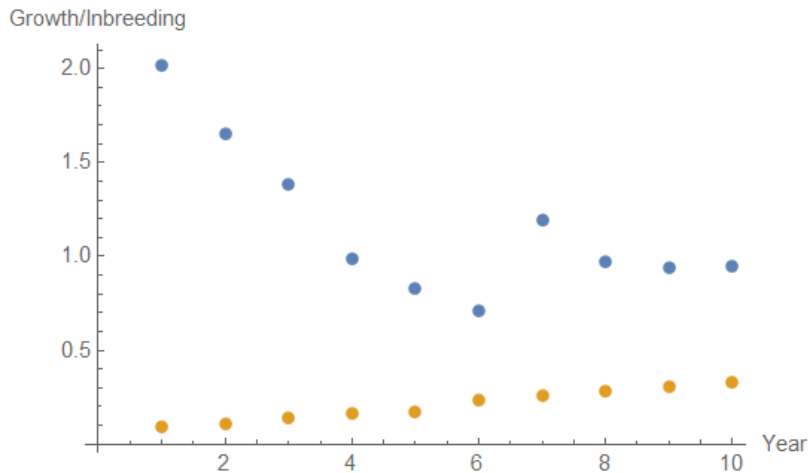
C. Now assume that the human population density is actually .98. Would this disease now lead to an epidemic? Why?

D. Assume again that the density of the human population is .98. What proportion of the population would need to be vaccinated to prevent an epidemic?

Practice Problem

| Year | Wolf Growth Rate | Wolf Inbreeding Coefficient |
|------|------------------|-----------------------------|
| 2002 | 2.02 | 0.09 |
| 2003 | 1.65 | 0.109 |
| 2004 | 1.38 | 0.136 |
| 2005 | 0.99 | 0.161 |
| 2006 | 0.83 | 0.173 |
| 2007 | 0.71 | 0.233 |
| 2008 | 1.19 | 0.261 |
| 2009 | 0.97 | 0.285 |
| 2010 | 0.94 | 0.309 |
| 2011 | 0.95 | 0.332 |

Data from the Isle Royale Wolf Project: <http://www.isleroyalewolf.org/>



- Is there a significant correlation?

$$\text{Covariance}[\text{Growth Rate}, \text{Inbreeding}] = -0.0218867$$

$$\text{SD}[\text{Growth Rate}] = 0.38702$$

$$\text{SD}[\text{Inbreeding}] = 0.08202$$

$$\text{Correlation}[\text{Growth Rate}, \text{Inbreeding}] = \frac{-0.0218867}{0.38702 * 0.08202} = -0.6895$$

$$t = -0.6895 * \sqrt{\frac{8}{1 - (-0.6895)^2}} = -2.69$$

$$t_{crit} = 2.306$$

We can reject the null hypothesis that inbreeding and growth rate are independent

The evolution of disease virulence*



European rabbit



Introduced to Australia in 1859
as game for hunting

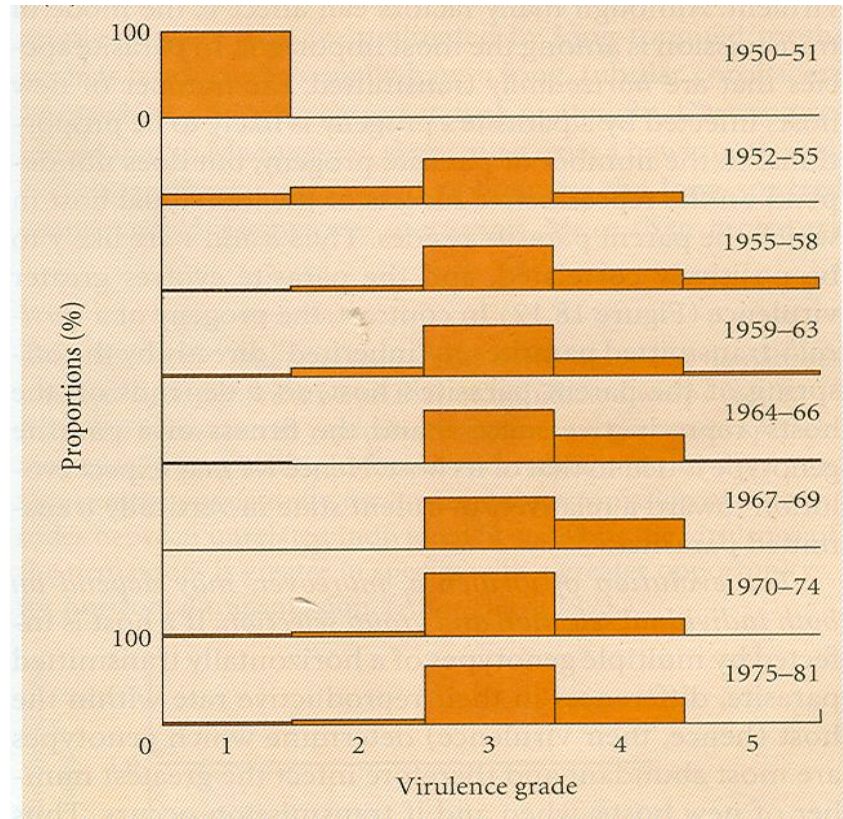
*Virulence is the extent to which a disease lowers host fitness

The evolution of disease virulence



- Rabbits became rangeland pests
- In an effort to control rabbit density, myxoma virus was introduced into the rabbit population in 1950
- Initially this virus was highly virulent, killing almost 99.9% of infected rabbits within 6-10 days

Over time, decreased virulence evolved



**High
virulence**

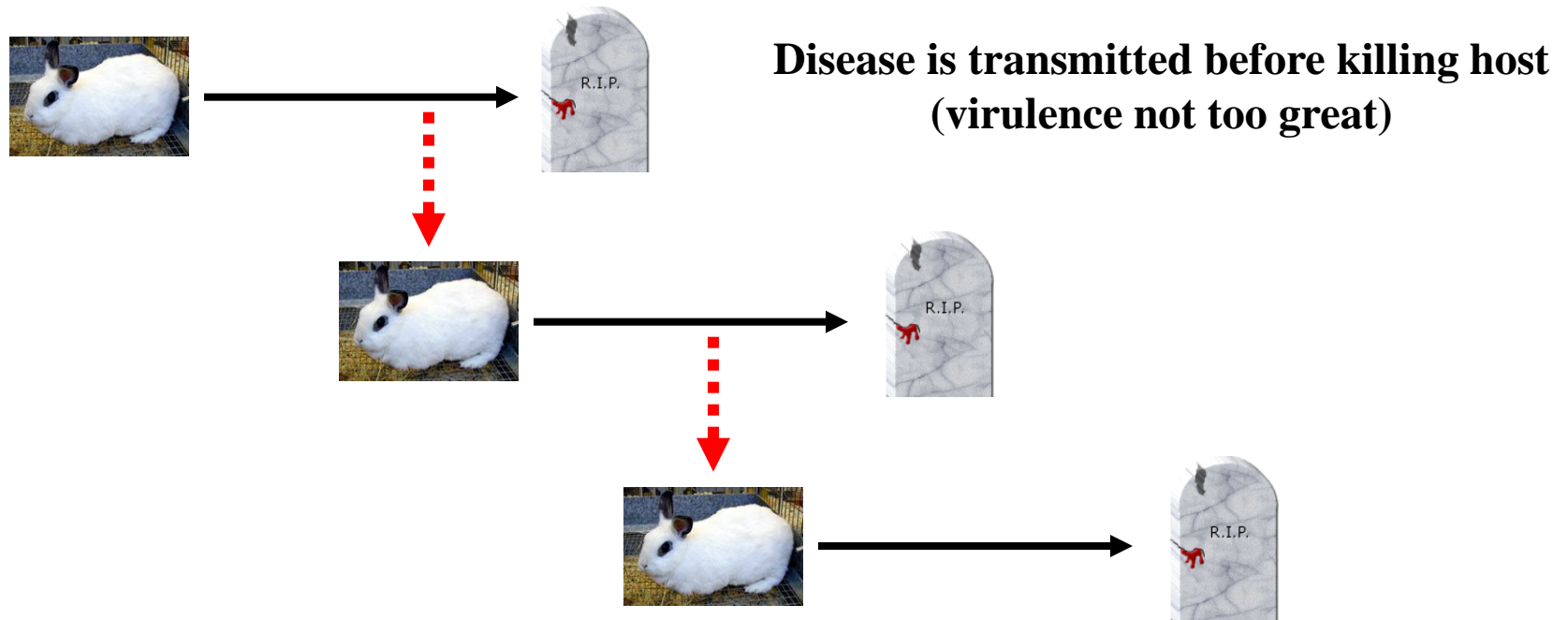
**Low
virulence**

Why?

The trade-off hypothesis

- A pathogen cannot reproduce inside the host without doing some harm
 - At the very least, host resources must be ‘stolen’
- All else being equal, those **pathogens with a higher rate of within host reproduction should be more likely to be transmitted** to a new host
- Because reproducing faster within the host causes greater harm, however, too great a rate of pathogen reproduction may debilitate the host sufficiently for transmission to a new host to be curtailed

The trade-off hypothesis



Virulence is largely determined by how frequently opportunities for transmission occur

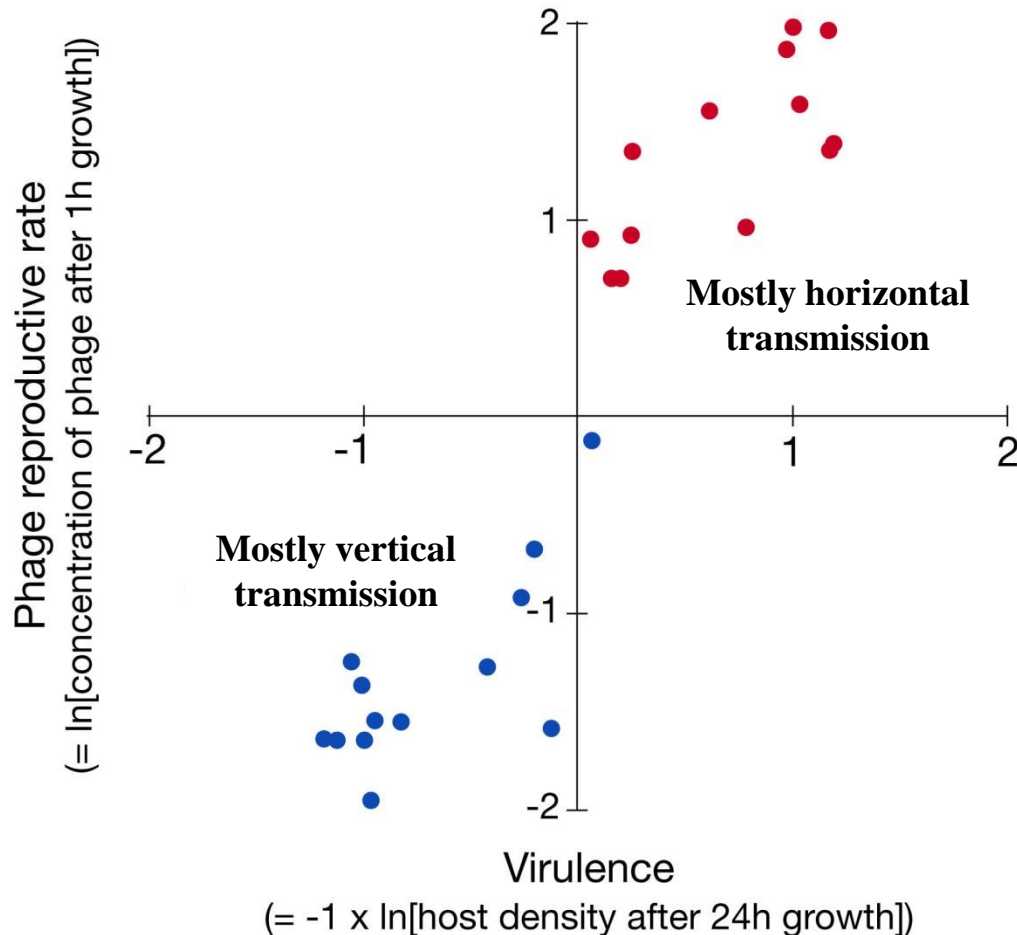
The trade-off hypothesis: an experimental test

(Messenger et. al. 1999)

- Bacteriophage f1 infects bacterial cells
- f1 can be transmitted *horizontally* or *vertically*
- Predicted that horizontally transmitted phage would be more virulent than vertically transmitted phage
- Predicted that those phage with the greatest rate of within host reproduction would be the most virulent

The trade-off hypothesis: an experimental test

(Messenger et. al. 1999)

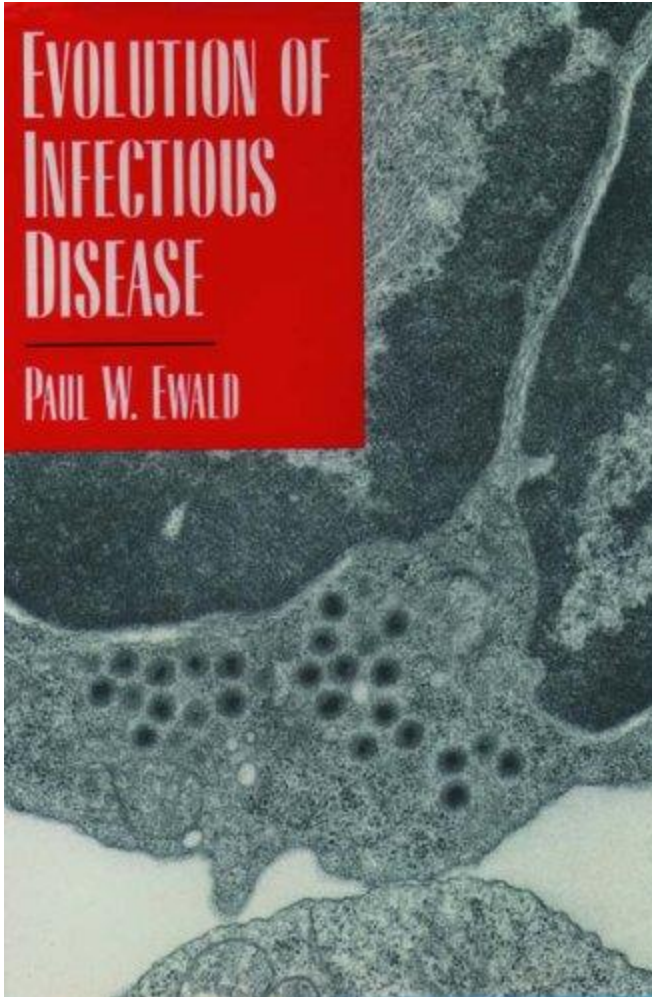


- Horizontal transmission led to increased virulence

- Phage with the greatest rate of within host reproduction were the most virulent

The trade-off hypothesis and human health

Ewald (1993, 1994)



Two predictions:

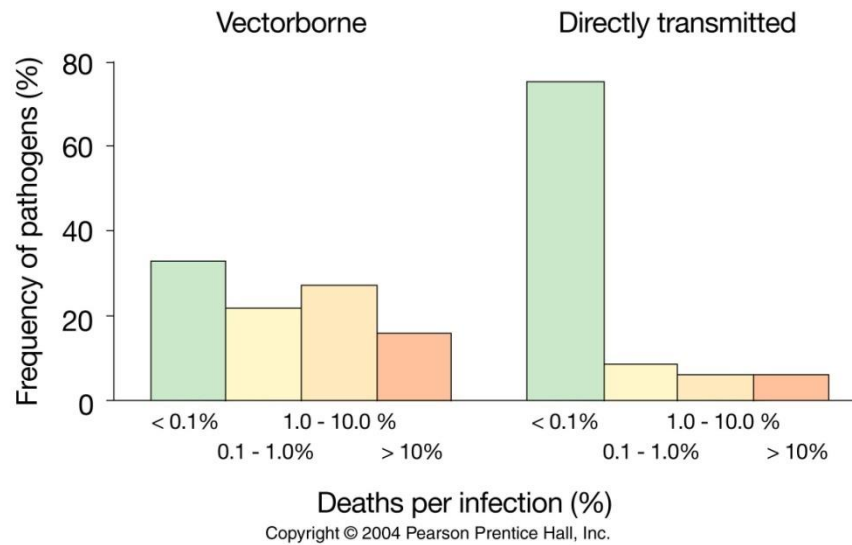
1. Vector borne diseases will be more virulent than directly transmitted diseases

- directly transmitted diseases require a mobile host able to encounter other potential hosts
- Vector born diseases may be (more) efficiently transmitted by immobile very ill hosts

2. Diseases that cause diarrhea will be more virulent

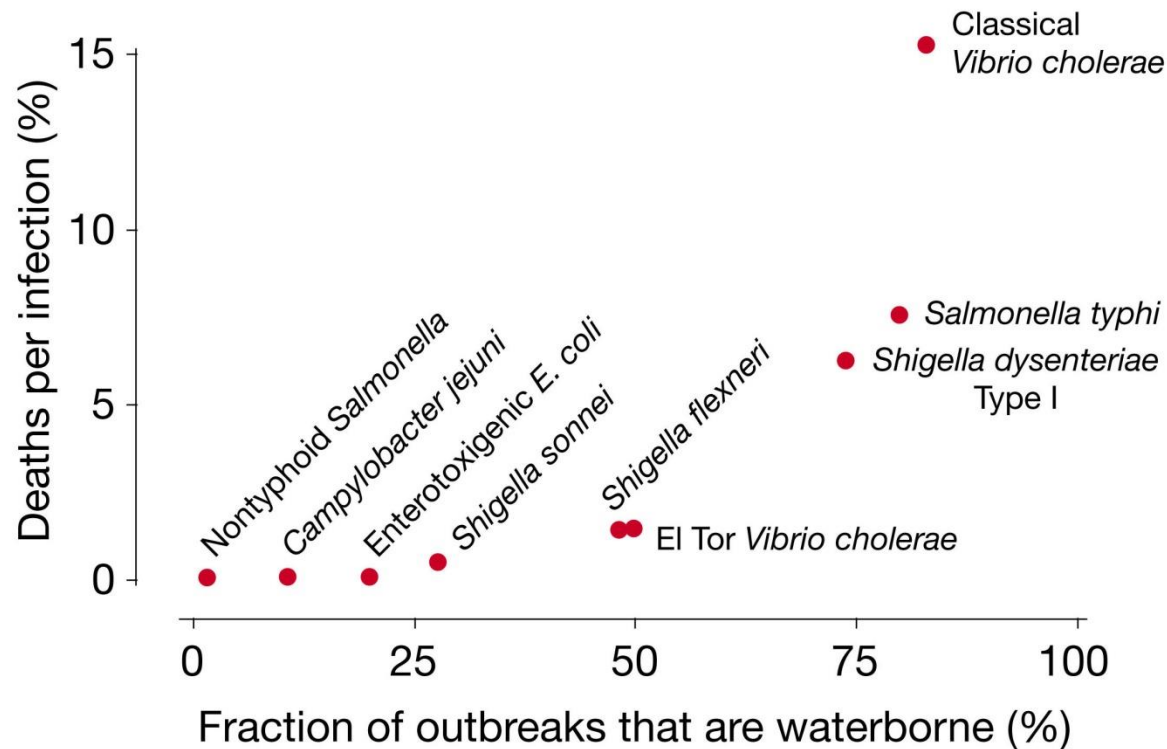
- Diarrheal diseases can be transmitted (potentially better) even by very sick hosts

Are vector borne diseases more virulent?



Vector borne pathogens are more likely to be virulent!

Are diarrheal diseases more virulent?



Copyright © 2004 Pearson Prentice Hall, Inc.

Pathogens transmitted through diarrhea are more likely to be virulent

The evolution of infectious disease: summary

1. Pathogen populations are under constant pressure from the immune system
2. Mutant pathogen strains that can elude the immune system are selectively favored
3. Recombinant pathogens can elude the host immune system entirely, potentially leading to pandemics
4. Levels of pathogen virulence evolve in response to opportunities for transmission
5. Indirectly transmitted pathogens should be more virulent than directly transmitted pathogens

Practice Problem

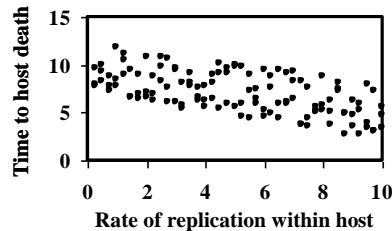
A team of scientists from the CDC has compiled data on a recently discovered infectious disease. This disease has been observed in 5 different geographic regions and has probably been present, although previously unreported, for the last 100 years. Some of the data collected by the CDC scientists is presented below:

| Geographical region | % of outbreaks that are vector transmitted |
|---------------------|--|
| 1 | 33.6 |
| 2 | 1.2 |
| 3 | 38.2 |
| 4 | 35.5 |
| 5 | 42.6 |

A (10pts). In which geographic region do you expect the disease to be the least virulent? Why?

B (10pts). What is the trade-off theory for the evolution of virulence?

C (20pts). To test whether the trade-off theory for the evolution of virulence applied to this infectious disease, the CDC scientists infected rabbits with 20 different genetic strains of the disease (each strain was tested in 5 different rabbits). They then measured the rate of disease replication within the rabbits, and the time it took for the rabbit to succumb to the disease. A plot of their data is shown below:



Does this data support the trade-off hypothesis? Justify your response.