Around here, it takes all the running you can do just to stay in the same place.
LEWIS CARROLL
Alice in Wonderland
I think

Then there are A, B, C, D, E, F, G, H, I, J

http://commons.wikimedia.org/wiki/File:Darwin_Tree_1837.png
Phylogenetics

- Phylogenetic trees measure the genetic distance between organisms, and identify the nearest relatives.
- To construct a phylogenetic tree, compare differences in nucleotide sequences of many isolates of putatively related organisms.
- Each division in the tree is a ‘node’, the common ancestor of the organisms or the isolates identified to its right.
- After such branching, the organisms and their sequences evolve independently. The ‘root’ (at the extreme left) is the assumed common ancestor of all organisms in the tree.

https://evolution.berkeley.edu/evolibrary/article/evo_05
Adaptation

Host protein adapted to environment

Change in environment/selective regime

Host protein maladapted to environment

Adaptive evolution

Host protein adapted to "new" environment

Purifying selection

(removal of deleterious alleles)

Host protein maintained in "new" environment

(positive selection)
Darwin would have loved viruses!

The best exemplars of evolution by natural selection, and for RNA viruses, evolution is so rapid it can be followed in real time
Viral evolution: The constant change of a viral population in the face of selection pressures

- Where did viruses come from?
- Where are viruses going?
Modern virology has provided a window on the mechanisms of evolution

- As host populations grow and adapt, virus populations are selected that can infect them
  - *New viral populations emerge every day*
- It also works the other way
  - *Viral populations can be significant selective forces in the evolution of host populations*
- If a host population cannot adapt to a lethal virus infection, the population may be exterminated
The public is constantly confronted with the reality of viral evolution (even if they don’t believe in evolution)

- New viral diseases: AIDS, West Nile virus in the US, HCV, Ebolavirus, Zika virus
- Regular bouts every year with influenza and common cold viruses
- Drug resistant HIV

Simple fact: viruses evolve faster than many can comprehend
Four main drivers of virus evolution

- Large numbers of progeny
- Large numbers of mutants
- Quasi-species effects
- Selection
Virus-infected cells produce large numbers of progeny

<table>
<thead>
<tr>
<th>Virus in plasma</th>
<th>HBV</th>
<th>HIV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Half-life</td>
<td>24 h</td>
<td>6 h</td>
</tr>
<tr>
<td>Daily turnover</td>
<td>50%</td>
<td>90%</td>
</tr>
<tr>
<td>Total production in blood</td>
<td>$&gt;10^{11}$</td>
<td>$&gt;10^9$</td>
</tr>
</tbody>
</table>

The interface of host defense and virus replication is fertile ground for selection and evolution.
Replicating viruses produce large numbers of mutant genomes

- Evolution is possible only when mutations occur in a population
- Mutations are produced during copying of any nucleic acid molecule

Viral genomes are always mutating!
The Ebola Virus Is Mutating, Say Scientists

The outbreak has so far claimed 8,795 lives across the affected West African region.

Scientists at a French research institute say the Ebola virus has mutated and they are studying whether it may have become more contagious.

Researchers at the Institut Pasteur are analyzing hundreds of blood samples from Guinean Ebola patients in an effort to determine if the new variation poses a higher risk of transmission, according to the BBC.

A health care worker, right, takes the temperatures of school children for signs of the Ebola virus before they enter their school in the city of Conakry, Guinea, Monday, Jan. 19, 2015.
RNA viruses

- Lack of proofreading activity in RNA dependent RNA polymerase: high error frequencies (1 misincorporation / $10^3 - 10^4$ nt polymerized)
- Average error frequency: 1 in $10^4$ or $10^5$ nucleotides polymerized
- In a 10 kb RNA virus genome, a mutation frequency of 1 in $10^4$ results in about 1 mutation per genome
DNA viruses

- Genome replication not as error prone as RNA viruses
- Proofreading
- Most DNA viruses generate less diversity, evolve slower than RNA viruses
The graph shows the relationship between mutation rate ($\mu_{sn/c}$) and genome size (nt) across different types of viruses. The x-axis represents genome size in nucleotides (nt), while the y-axis represents the mutation rate in substitutions per nucleotide per generation ($\mu_{sn/c}$). Different types of viruses are represented by various symbols and colors.

- Viroid
- ss(+)RNA
- ss(-)RNA
- dsRNA
- retro
- ssDNA
- dsDNA
- Bacterium

*substitutions/nucleotide/generation
The quasispecies concept

- Analysis of an RNA bacteriophage population (Qβ):

  “A Qβ phage population is in a dynamic equilibrium with viral mutants arising at a high rate on the one hand, and being strongly selected against on the other. The genome of Qβ cannot be described as a defined unique structure, but rather as a weighted average of a large number of different individual sequences.” E. Domingo, D. Sabo, T. Taniguchi, C. Weissmann. 1978. Nucleotide sequence heterogenity of an RNA phage population. Cell 13:735-744.

- This discovery was far ahead of its time (40 years ago this month!), not appreciated by most virologists

- Virus populations exist as dynamic distributions of nonidentical but related replicons, called quasispecies
Viral quasispecies

this

not this
Viral infections are initiated by a population of particles, not a single virus particle.

The large number of progeny produced are products of selective forces inside the host.

The survivors that can re-infect a new host reflect the selective forces outside the host.

Quasispecies effects
The myth of consensus genome sequences

- For a given RNA virus population, the genome sequences cluster around a consensus or average sequence, but virtually every genome can be different from every other.
- A genome with the consensus sequence may not exist in the population.
Quasispecies

Variation further generated by recombination and reassortment
Selection

- *Survival of the fittest*: A rare genome with a particular mutation may survive a selection event, and this mutation will be found in all progeny genomes.

- *Survival of the survivors*: However, the linked, but unselected mutations, get a free ride.

- Consequently, the product of selection after replication is a new, diverse population that shares only the selected mutations.
Diversity is selected

- Mutations in viral polymerases that reduce the frequency of incorporation errors
  - Do not have a selective advantage when wild type and anti-mutators are propagated together
  - Lower rates are neither advantageous nor selected in nature
  - Mutants are often less pathogenic
- High mutation rates are selected during virus evolution: mutation is good for viral populations

http://www.virology.ws/2009/05/15/increased-fidelity-reduces-viral-fitness/
Error threshold

- Mutation is a powerful advantage, but selection and survival balances genetic fidelity and mutation rate
- This limit is called the error threshold
  - Exceed it: loss of infectivity
  - Below it: cannot produce enough mutations to survive selection
- RNA viruses: evolve close to their error threshold
- DNA viruses: evolve far below their error threshold
Error threshold

- Expose a cell culture infected with a DNA virus to a base analog such as 5-azacytidine
- 5-azacytidine is incorporated as a C, but templates as a T (G to A transitions)
- Mutation rate among viral progeny increases several orders of magnitude
- When a similar experiment is done with an RNA virus, the error frequency per genome increases only two- to threefold at best - cannot make any more mutations
Error threshold

Antiviral ribavirin and poliovirus

RNA genome specific infectivity (% wt)

mutations/RNA genome

LI_{50}
Importance of Quasispecies: Genetic bottlenecks

- Extreme selective pressures on small populations that result in loss of diversity, accumulation of non-selected mutations, or both
- A single RNA virus plaque is picked and expanded
- Next, a single plaque is picked from the expanded stock
- The process is repeated over and over
Genetic bottlenecks

- After about 20-30 cycles of single-plaque amplification, many virus populations are barely able to grow.
- They are markedly less fit than the original population.
- The environment is constant, and the only apparent selection is that imposed by the ability of the population of viruses from a single plaque to replicate.
- Why does fitness plummet?
Genetic bottlenecks

- The bottleneck arises by restricting further viral replication to the progeny found in a single plaque
  - A few thousand progeny viruses derived from a single founder virus
Genetic bottlenecks

- Another way to look at this problem: Muller’s ratchet: Small, asexual populations accumulate deleterious mutations
- Replicating RNA viruses are close to error threshold
- By restricting population growth to serial single founders (the bottleneck) under otherwise nonselective conditions, so many mutations accumulate (exceed the threshold) that fitness decreases
The ratchet metaphor: each of the new mutations works like a ratchet, allowing the gear to move forward, but not backward.

Each round of error-prone replication works like a ratchet, “clicking” relentlessly as mutations accumulate at every replication cycle.
Fitness decline compared to initial virus clone after passage through a bottleneck

<table>
<thead>
<tr>
<th>Virus</th>
<th># of bottleneck passages</th>
<th>% Decrease in fitness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacteriophage φ6</td>
<td>40</td>
<td>22</td>
</tr>
<tr>
<td>Vesicular stomatitis virus</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>Foot-and-mouth disease virus</td>
<td>30</td>
<td>60</td>
</tr>
<tr>
<td>HIV</td>
<td>15</td>
<td>94</td>
</tr>
<tr>
<td>Bacteriophage MS2</td>
<td>20</td>
<td>17</td>
</tr>
</tbody>
</table>
Bottlenecks in the real world?

- Infection by a limited virus population and subsequent amplification are often found in nature
  - Small droplets of suspended virus during aerosol transmission
  - Activation of a latent virus from a limited population of cells
  - Small volume of inoculum introduced in infection by insect bites

- How do infections that spread by these routes escape Muller’s ratchet?
Avoiding the ‘ratchet’

- Subject a more diverse viral population to serial passage
  - *Don’t pick a single plaque, pool several plaques*

- More diversity in the replicating population facilitates construction of a mutation-free genome by recombination or reassortment, removing or compensating for mutations that affect growth adversely
By exchange of genetic information

Reassortment
Avoiding the ‘ratchet’

- The message is simple: Diversity of a viral population is important for the survival of individual members
  - Remove diversity, and the population suffers
An example of selection: genetic shift & drift

- Selection of viral mutants resistant to elimination by antibodies or cytotoxic T cells inevitable when sufficient virus replication occurs in an immunocompetent individual
- Drift - diversity arising from copying errors and immune selection - may occur each time a genome replicates
- Shift - diversity arising after recombination or reassortment
Influenza A viruses are classified by antigenic composition, by serologic testing of HA and NA.

- Combinations of H and N are called HxNy
- x = 1-18; y = 1-11
- H1-17 can infect birds; H1, H2, H3 can infect and transmit between humans
Antigenic drift: Influenza virus
Host-virus arms race

Around here, it takes all the running you can do just to stay in the same place.

LEWIS CARROLL
Alice in Wonderland
Evolution-guided functional analysis of host-virus arms races

Phase 1: Analyze sequences for positive selection

Primate phylogeny

Ortholog sequences

Concept: synonymous and non synonymous changes
Phase 2: Determine phenotypic consequences of ortholog variation
Phase 3: Map positively selected changes onto host-virus interaction

Virus winning

Host winning

GLELHPDYKTWSPEQVCSFLRRGGF
GPELHPDHKTWGPPEQVCSFLRRGGF
GLELHPDYKTWGPEQVCSFLRRGGF

Mutate $S \rightarrow G$

Virus winning

Host winning
Virus-host conflicts have driven evolution of the immune system

Host defense protein

Positively selected amino acids

Viral counter-defense proteins

TRIM5α
- RBCC domain
- B30.2 domain
  - HIV capsid

Protein kinase R
- RNA-binding domain
- Kinase domain
  - Herpesvirus US11
  - Influenza NS1A
  - Poxvirus E3L
  - Hepatitis C NS5A
  - Poxvirus K3L and E3L
  - Ranavirus RelF2H
  - HIV Tat

Referenced:

TRF1 evolution in rodents shaped by two virus-host races

- Arenavirus Binding Site (gray)
- MMTV Binding Site (blue)

Human TfR1 Dimer

Transferrin binding site

Cell surface
Despite this genome diversity...

- There are only 3 serotypes of poliovirus (>150 rhinoviruses)
- One measles serotype, continuous influenza variation
- Why?
Selection: Is virulence a positive or negative trait?

- **Positive trait**: increased virulence is a consequence of high viral loads, facilitating transmission
- **Negative trait**: increased virulence reduces transmissibility because hosts die faster, reducing exposure to uninfected hosts
- Probably both: there are many virulent and virulent viruses
An experiment in the evolution of virulence

- In 1859, the European rabbit was introduced to Australia for hunting purposes
- Lacking natural predators, it reproduced to plague proportions in a short time
Evolution of virulence

- Myxoma virus released in Australia in the 1950s in an attempt to rid the continent of the rabbits
- Natural host of myxoma virus is the cottontail rabbit
- Virus spread by mosquitoes; infected rabbits develop superficial warts on their ears
- European rabbits are a different species, infection is 90-99% fatal
Evolution of virulence

- In the first year, the released virus was efficient in killing rabbits with a 99.8% mortality rate.
- After the second year the mortality dropped to 30%.
- Rate of killing was lower than the reproductive rate of the rabbits, and hope for 100% eradication was dashed.
Evolution of virulence

- Both rabbits and viruses produce large numbers of offspring
- Virus evolved to kill fewer rabbits and to extend the life of lethally infected rabbits so that the virus could overwinter and spread in spring mosquitoes
- The rabbits evolved to become more resistant or tolerant of the virus
- As predicted for an evolving host coming to an equilibrium with the pathogen

**Doubts raised over Australia’s plan to release herpes to wipe out carp**

*Warm water may render the virus ineffective against the invasive fish, say researchers.*

**NEWS** • 22 FEBRUARY 2018
Evolution of viral virulence in humans?

- Experience with Lassa virus, Ebolavirus, HIV: animal-human virus transfers tend to be virulent
- But viruses from older jumps (measles, poliovirus) are less virulent
- What happened in the meantime?
Nevertheless, we are obsessed with increased viral virulence

- Ebolavirus is mutating, will go airborne (Osterholm, NY Times, 11 Sept 2014)
- Ebolavirus, Lisa Henley: “Is it getting better at replicating as it goes from person to person?” \textit{(Ebola Wars, Richard Preston, New Yorker)}
- Peter Hotez, \textit{NYT} Op-Ed 8 April 2016: “There are many theories for Zika’s rapid rise, but the most plausible is that the virus mutated from an African to a pandemic strain a decade or more ago and then spread east across the Pacific from Micronesia and French Polynesia, until it struck Brazil.”
- It’s easy to blame mutation - but usually there are other explanations (e.g. poliovirus, see ‘Emerging Viruses’)
We have no data on the effect of evolution on viral virulence in humans
The origin of viruses

Oldest viral stocks:
1918 influenza virus
*Pithovirus sibericum*
(30,000 y)
Origins of DNA viruses

Molecular clocks: By relating timescale of herpesviral genome evolution with that of hosts, believe that three major groups of herpesviruses (alpha, beta, gamma) arose ~180-220 million years ago.
Endogenous viruses - retrovirus and otherwise

Phylogenomics
History of ssDNA virus integrations

- Parovirus-like
- Dependovirus-like
- Circovirus-like

90Mya 60Mya 30Mya

- Humans
- Baboon
- Tarsier
- Mouse
- Rat
- Squirrel
- Kangaroo rat
- Guinea pig
- Pika
- Rabbit
- Alpaca
- Cow
- Dolphin
- Pig
- Cat
- Panda
- Dog
- Horse
- Megabat
- Microbat
- Armadillo
- Sloth
- Tenrec
- Hyrax
- Elephant
- Opossum
- Wallaby
- Platypus
- Finch
- Frog
- Tetraodon
- Fugu
How old are viruses?

- Estimates of molecular evolution suggest marine origin of some retroviruses >450 Ma, Ordovician period
- Likely originated billions of years ago - before cells?

Nobu Tamura (http://spinops.blogspot.com)

Orthoceras, a nautiloid cephalopod, 488 Ma

©Principles of Virology, ASM Press
Origins of viruses

Start small, acquire genes
Start large, lose genes
All known types of viruses likely evolved long before humans appeared on Earth.

All human viruses have therefore evolved from animal viruses.
Origins of smallpox virus

- Phylogenetic analysis of related viruses and their hosts suggests emergence 3000-4000 years ago in East Africa, transmitted from camels
- Camels were introduced to Africa 3500-4500 year ago
- Camels likely infected with a smallpox virus ancestor from gerbils
Origins of measles virus

- Measles virus is closely related to rinderpest virus, a bovine pathogen.
- Probably evolved from an ancestral rinderpest virus when humans first began to domesticate cattle (11th-12th centuries, 1000-1200).
- Established in the Middle East when human populations began to congregate in cities (MV maintenance requires populations of 250,000-500,000).
- Spread around the world by colonization and migration, reaching Americas in 16th century and destroying native Americans.
Evolution of new viruses

• Assumption: new viruses can only arise from viruses that are now in existence, not de novo

• What is the number of all possible mutations of a viral genome?

• Sequence comparisons of several RNA virus genomes have demonstrated that well over half of all nucleotides can accommodate mutations
Evolution of new viruses

- For a 10 kb viral genome, $4^{5000}$ sequences
- Deletions, recombination, and reassortment increase the numbers
- $\sim 4^{135}$ atoms in the visible universe
The fundamental properties of viruses constrain and drive evolution

- Despite many rounds of replication, mutation, selection, we can recognize a herpesvirus or influenza virus genome by sequence analysis.
- Viral populations often maintain master or consensus sequences, despite opportunities for extreme variation.
- How is stability maintained?
Constraining viral evolution

- Extreme alterations in viral consensus genome do not survive selection
- The viral genome is one constraint
  - DNA cannot become RNA, or vice versa
  - Replication - interaction with host proteins, replication signals
  - mRNA synthesis signals, poly(A) addition, processing
  - RNA structure, codon usage,
- Physical nature of capsid
  - Icosahedral capsids: defined internal space, fixes genome size
- Selection during host infection
  - A mutant too efficient in bypassing host defenses will kill host, have the same fate as one that does not sufficiently replicate
Imagine what a virus can do with 8 million years
Next time: Emerging viruses