

Evolution

Lecture 21

Biology W3310/4310

Virology

Spring 2016

*Anything produced by evolution is bound to
be a bit of a mess.*

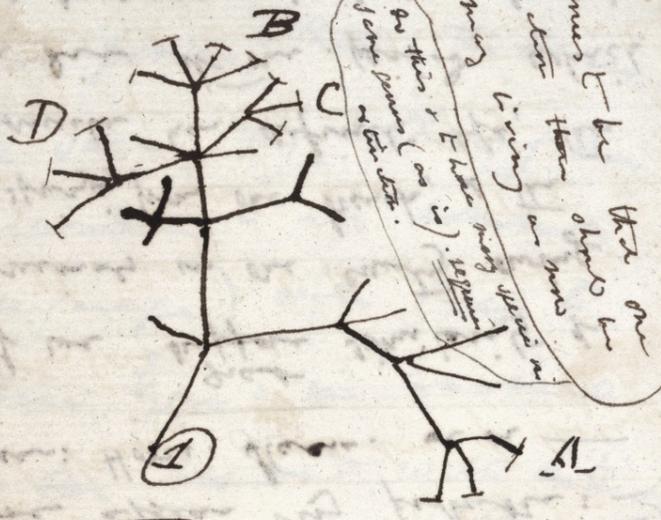
SYDNEY BRENNER

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

LEWIS CARROLL

Alice in Wonderland

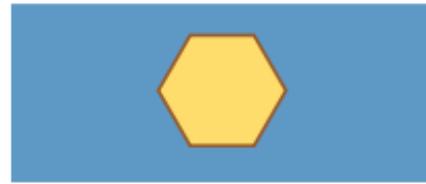
I think



Can't guess E. be that one
 because there should be
 some other than in mass
 living in same place as
 in mass
 Do you think E. later may have been
 a later species in the same line?

then between A & B. various
 sort of relation. C + B. The
 finest gradation, B & D
 rather greater distinction
 than genus would be
 formed. - bearing relation

Adaptation



Host protein adapted to environment

↓ Change in environment/
selective regime



Host protein maladapted to environment

↓ Adaptive evolution



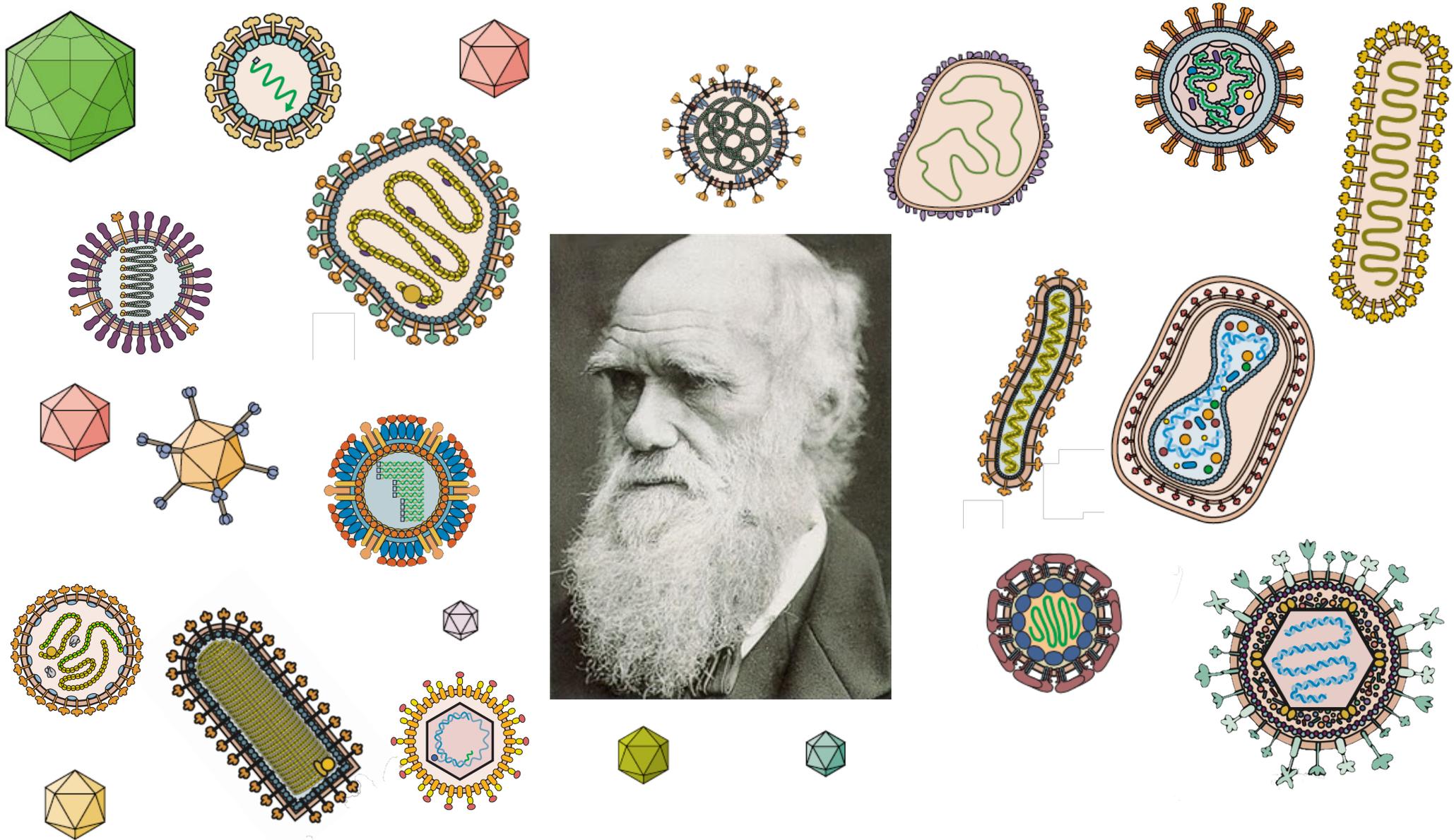
Host protein adapted to "new" environment

↓ Purifying selection



Host protein maintained in "new" environment

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

| Virus in plasma | HBV | HIV |
|---------------------------|------------|---------|
| Half-life | 24 h | 6 h |
| Daily turnover | 50% | 90% |
| Total production in blood | $>10^{11}$ | $>10^9$ |

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

Kevin McSpadden @KevinMcspadden | Jan. 29, 2015



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.



Youssef Bah—AP

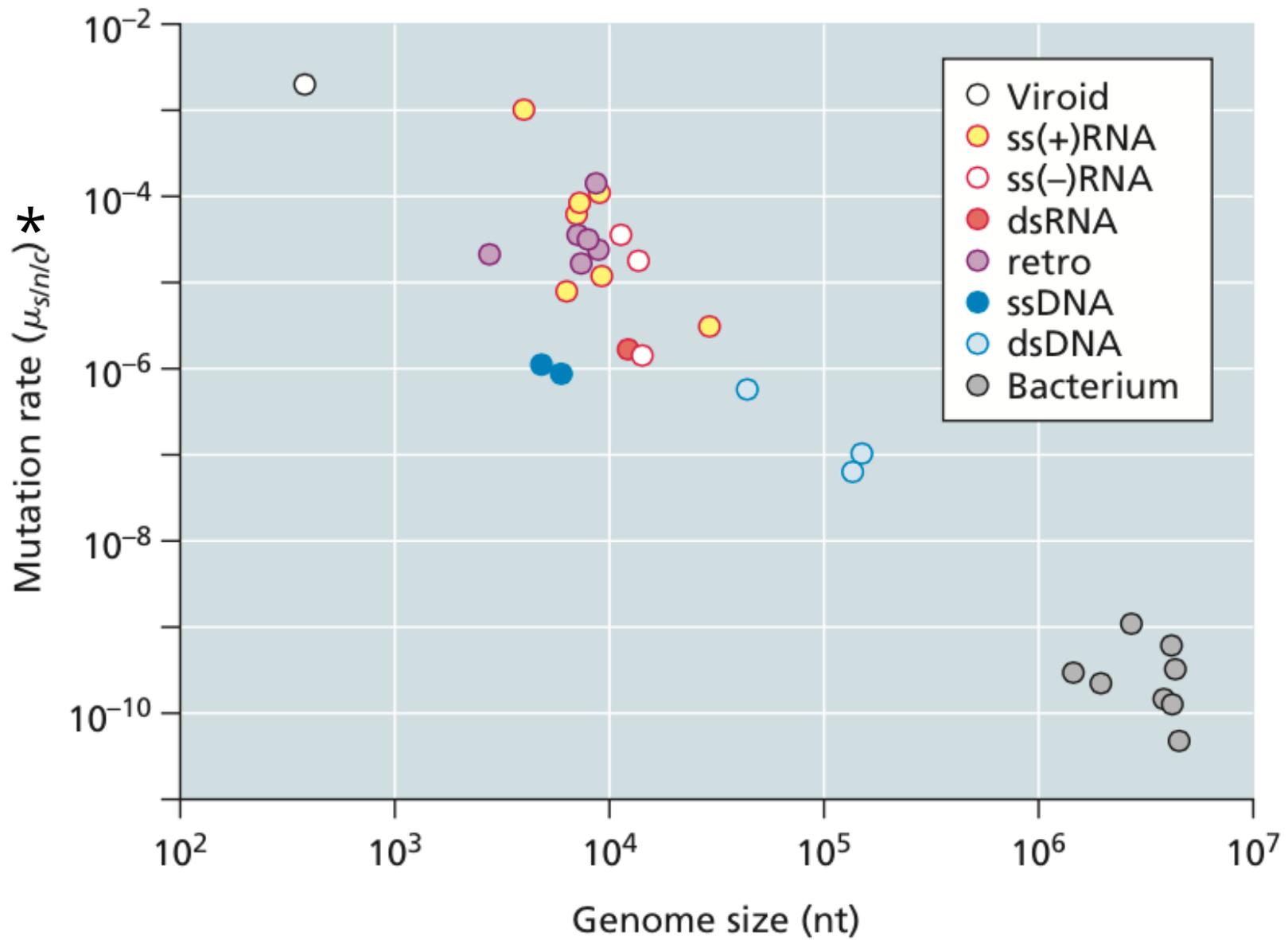
A health care worker, right, takes the temperatures of school children for signs of the Ebola virus before they enter their school in the city of Conakry, Guinea, Monday, Jan. 19, 2015

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



*substitutions/nucleotide/generation

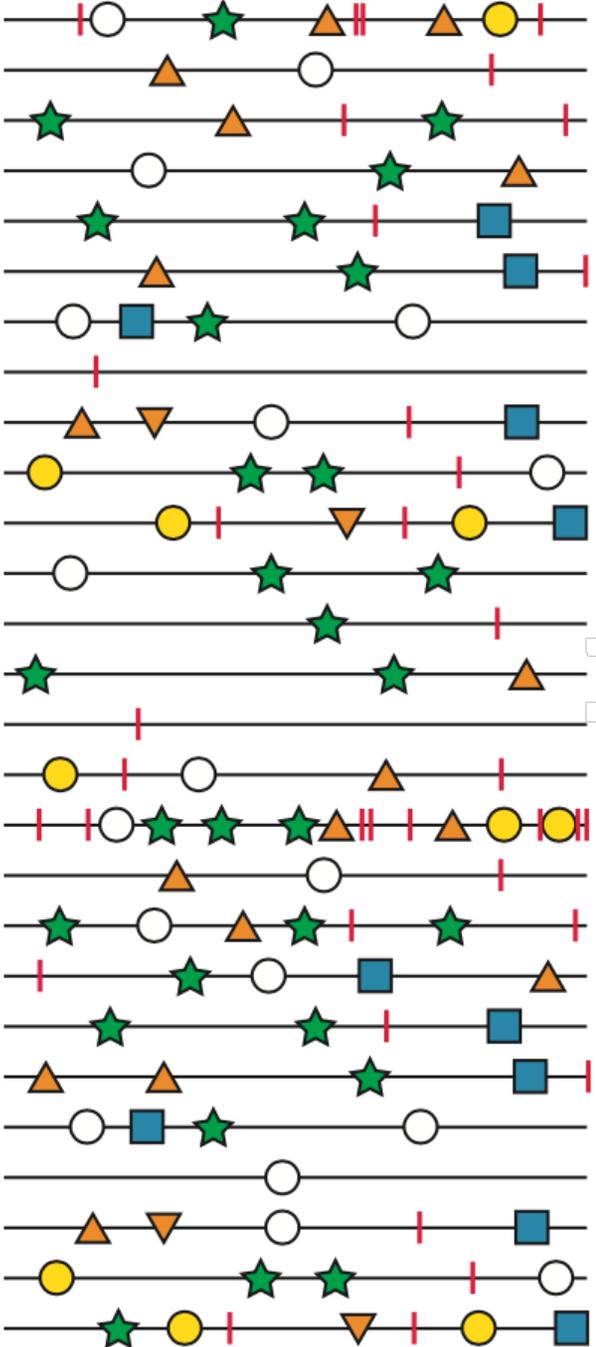
The quasispecies concept

- Analysis of an RNA bacteriophage population (Q β):

“A Q β 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 Q β 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 heterogeneity of an RNA phage population. Cell 13:735-744.

- This discovery was far ahead of its time, not appreciated by most virologists
- Virus populations exist as dynamic distributions of nonidentical but related replicons, called *quasispecies*

Viral quasispecies



← this

not this →

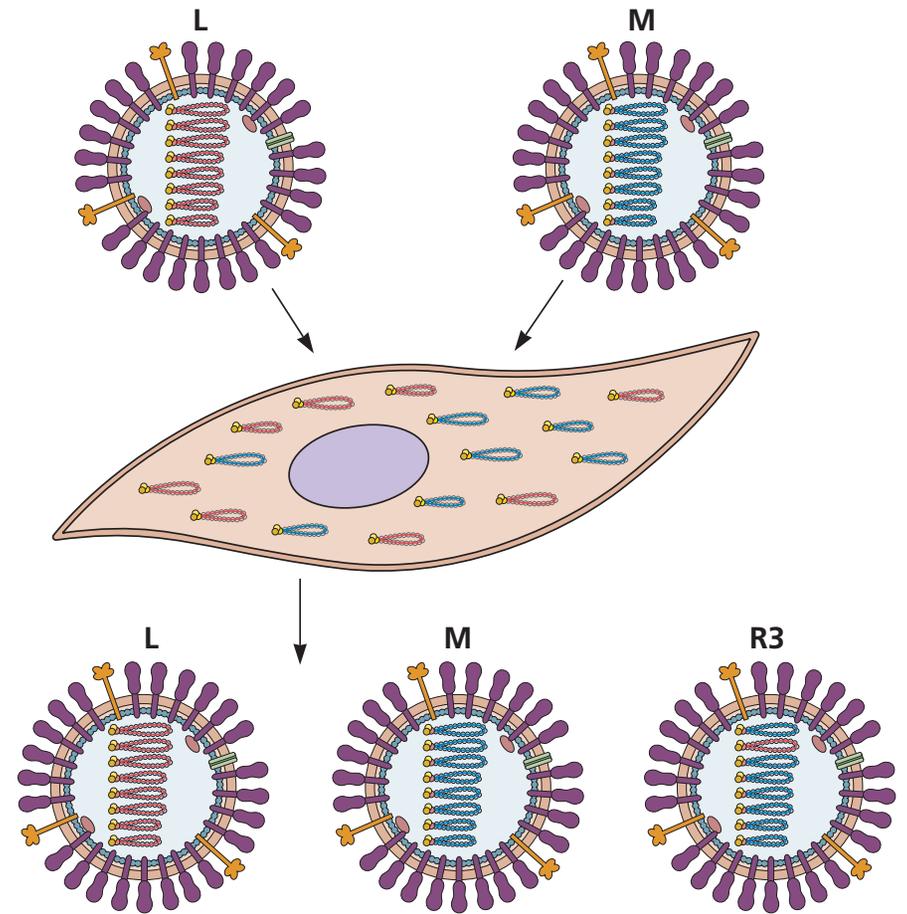
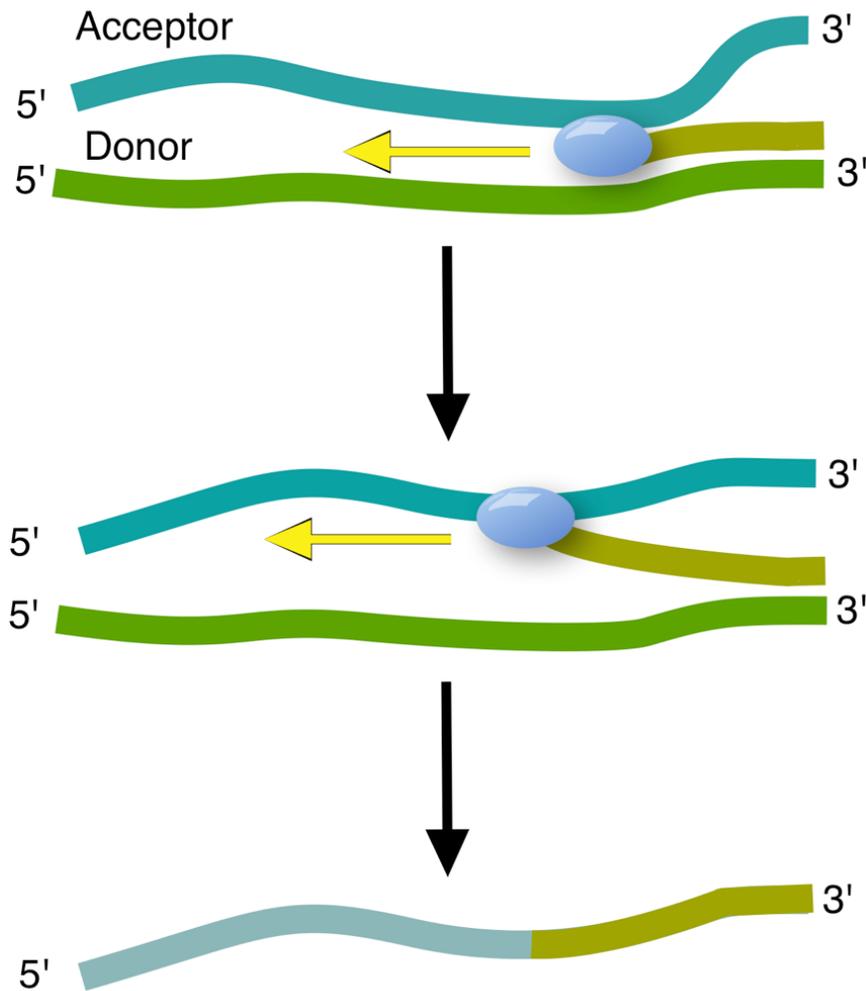
```
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180
190 200 210 220 230 240 250 260 270 280 290 300 310 320 330 340 350
360 370 380 390 400 410 420 430 440 450 460 470 480 490 500
510 520 530 540 550 560 570 580 590 600 610 620 630 640 650
660 670 680 690 700 710 720 730 740 750 760 770 780 790 800
810 820 830 840 850 860 870 880 890 900 910 920 930 940 950
960 970 980 990 1000 1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200 1210 1220 1230 1240 1250
1260 1270 1280 1290 1300 1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500 1510 1520 1530 1540 1550
1560 1570 1580 1590 1600 1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800 1810 1820 1830 1840 1850
1860 1870 1880 1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
2010 2020 2030 2040 2050 2060 2070 2080 2090 2100 2110 2120 2130 2140 2150
2160 2170 2180 2190 2200 2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
2310 2320 2330 2340 2350 2360 2370 2380 2390 2400 2410 2420 2430 2440 2450
2460 2470 2480 2490 2500 2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
2610 2620 2630 2640 2650 2660 2670 2680 2690 2700 2710 2720 2730 2740 2750
2760 2770 2780 2790 2800 2810 2820 2830 2840 2850 2860 2870 2880 2890 2900
2910 2920 2930 2940 2950 2960 2970 2980 2990 3000 3010 3020 3030 3040 3050
3060 3070 3080 3090 3100 3110 3120 3130 3140 3150 3160 3170 3180 3190 3200
3210 3220 3230 3240 3250 3260 3270 3280 3290 3300 3310 3320 3330 3340 3350
3360 3370 3380 3390 3400 3410 3420 3430 3440 3450 3460 3470 3480 3490 3500
3510 3520 3530 3540 3550 3560 3570 3580 3590 3600 3610 3620 3630 3640 3650
3660 3670 3680 3690 3700 3710 3720 3730 3740 3750 3760 3770 3780 3790 3800
3810 3820 3830 3840 3850 3860 3870 3880 3890 3900 3910 3920 3930 3940 3950
3960 3970 3980 3990 4000 4010 4020 4030 4040 4050 4060 4070 4080 4090 4100
4110 4120 4130 4140 4150 4160 4170 4180 4190 4200 4210 4220 4230 4240 4250
4260 4270 4280 4290 4300 4310 4320 4330 4340 4350 4360 4370 4380 4390 4400
4410 4420 4430 4440 4450 4460 4470 4480 4490 4500 4510 4520 4530 4540 4550
4560 4570 4580 4590 4600 4610 4620 4630 4640 4650 4660 4670 4680 4690 4700
4710 4720 4730 4740 4750 4760 4770 4780 4790 4800 4810 4820 4830 4840 4850
4860 4870 4880 4890 4900 4910 4920 4930 4940 4950 4960 4970 4980 4990 5000
5010 5020 5030 5040 5050 5060 5070 5080 5090 5100 5110 5120 5130 5140 5150
5160 5170 5180 5190 5200 5210 5220 5230 5240 5250 5260 5270 5280 5290 5300
5310 5320 5330 5340 5350 5360 5370 5380 5390 5400 5410 5420 5430 5440 5450
5460 5470 5480 5490 5500 5510 5520 5530 5540 5550 5560 5570 5580 5590 5600
5610 5620 5630 5640 5650 5660 5670 5680 5690 5700 5710 5720 5730 5740 5750
5760 5770 5780 5790 5800 5810 5820 5830 5840 5850 5860 5870 5880 5890 5900
5910 5920 5930 5940 5950 5960 5970 5980 5990 6000 6010 6020 6030 6040 6050
6060 6070 6080 6090 6100 6110 6120 6130 6140 6150 6160 6170 6180 6190 6200
6210 6220 6230 6240 6250 6260 6270 6280 6290 6300 6310 6320 6330 6340 6350
6360 6370 6380 6390 6400 6410 6420 6430 6440 6450 6460 6470 6480 6490 6500
6510 6520 6530 6540 6550 6560 6570 6580 6590 6600 6610 6620 6630 6640 6650
6660 6670 6680 6690 6700 6710 6720 6730 6740 6750 6760 6770 6780 6790 6800
6810 6820 6830 6840 6850 6860 6870 6880 6890 6900 6910 6920 6930 6940 6950
6960 6970 6980 6990 7000 7010 7020 7030 7040 7050 7060 7070 7080 7090 7100
7110 7120 7130 7140 7150 7160 7170 7180 7190 7200 7210 7220 7230 7240 7250
7260 7270 7280 7290 7300 7310 7320 7330 7340 7350 7360 7370 7380 7390 7400
7410 7420 7430 7440 7450 7460 7470 7480 7490 7500 7510 7520 7530 7540 7550
7560 7570 7580 7590 7600 7610 7620 7630 7640 7650 7660 7670 7680 7690 7700
7710 7720 7730 7740 7750 7760 7770 7780 7790 7800 7810 7820 7830 7840 7850
7860 7870 7880 7890 7900 7910 7920 7930 7940 7950 7960 7970 7980 7990 8000
8010 8020 8030 8040 8050 8060 8070 8080 8090 8100 8110 8120 8130 8140 8150
8160 8170 8180 8190 8200 8210 8220 8230 8240 8250 8260 8270 8280 8290 8300
8310 8320 8330 8340 8350 8360 8370 8380 8390 8400 8410 8420 8430 8440 8450
8460 8470 8480 8490 8500 8510 8520 8530 8540 8550 8560 8570 8580 8590 8600
8610 8620 8630 8640 8650 8660 8670 8680 8690 8700 8710 8720 8730 8740 8750
8760 8770 8780 8790 8800 8810 8820 8830 8840 8850 8860 8870 8880 8890 8900
8910 8920 8930 8940 8950 8960 8970 8980 8990 9000 9010 9020 9030 9040 9050
9060 9070 9080 9090 9100 9110 9120 9130 9140 9150 9160 9170 9180 9190 9200
9210 9220 9230 9240 9250 9260 9270 9280 9290 9300 9310 9320 9330 9340 9350
9360 9370 9380 9390 9400 9410 9420 9430 9440 9450 9460 9470 9480 9490 9500
9510 9520 9530 9540 9550 9560 9570 9580 9590 9600 9610 9620 9630 9640 9650
9660 9670 9680 9690 9700 9710 9720 9730 9740 9750 9760 9770 9780 9790 9800
9810 9820 9830 9840 9850 9860 9870 9880 9890 9900 9910 9920 9930 9940 9950
9960 9970 9980 9990 10000
```

Quasispecies effects

- Viral infections are initiated a population of particles, not a single virus particle
- The large number of progeny produced are complex products of selective forces inside the host
- The survivors that can re-infect a new host reflect the selective forces outside the host

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

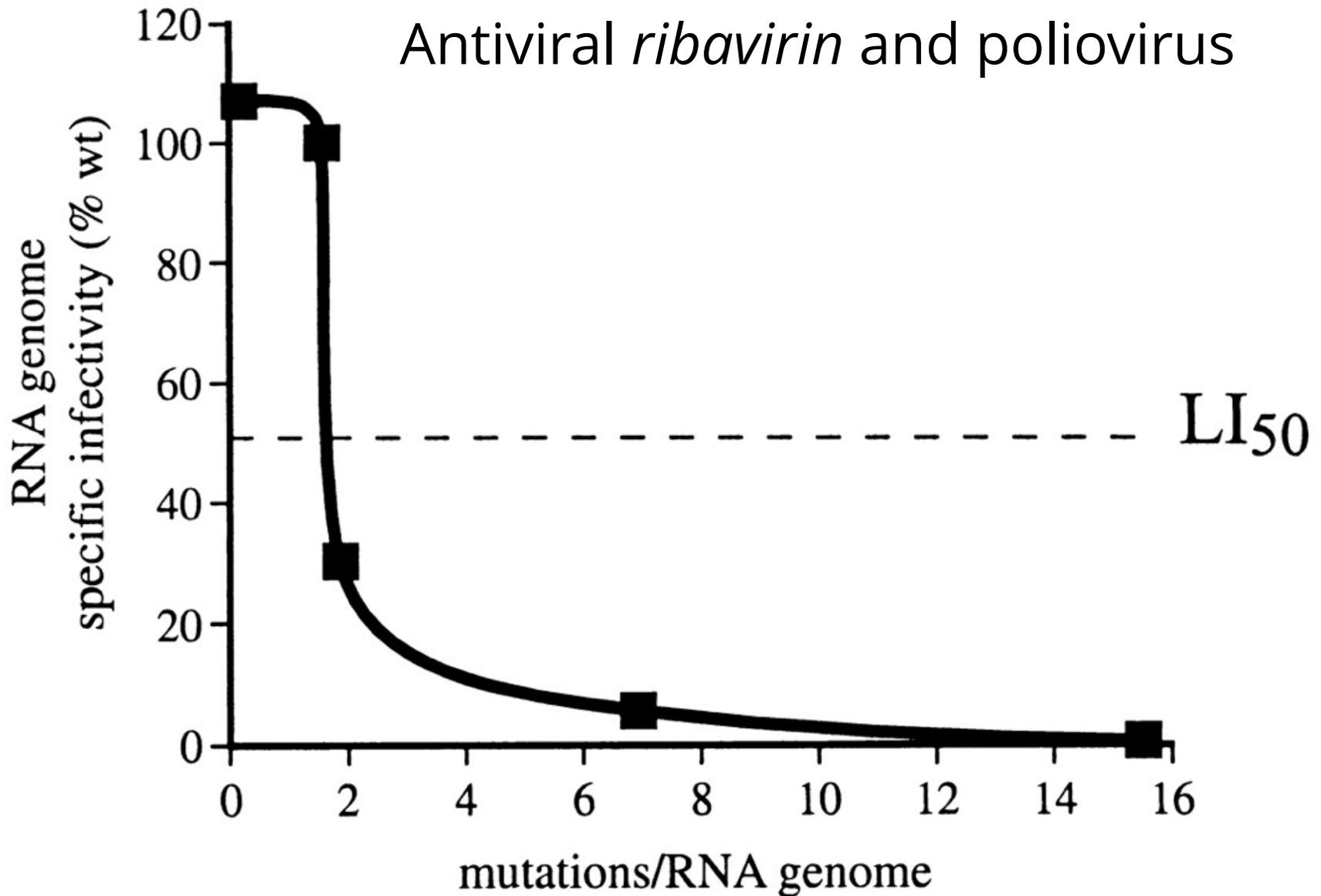
Error threshold

- 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

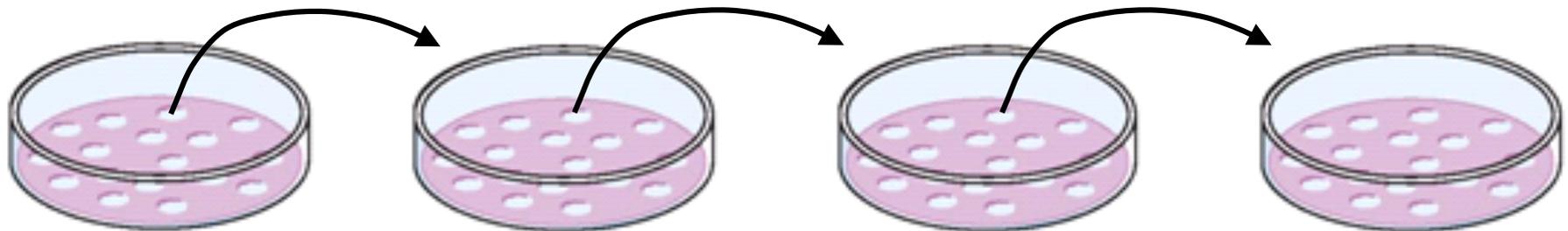
- Expose a cell culture infected with a DNA virus to a base analog such as 5-azacytidine
- 5-azacytidine is incorporated as a C, but templates as a T (G to A transitions)
- Mutation rate among viral progeny increases several orders of magnitude
- When a similar experiment is done with an RNA virus, the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

Error threshold



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

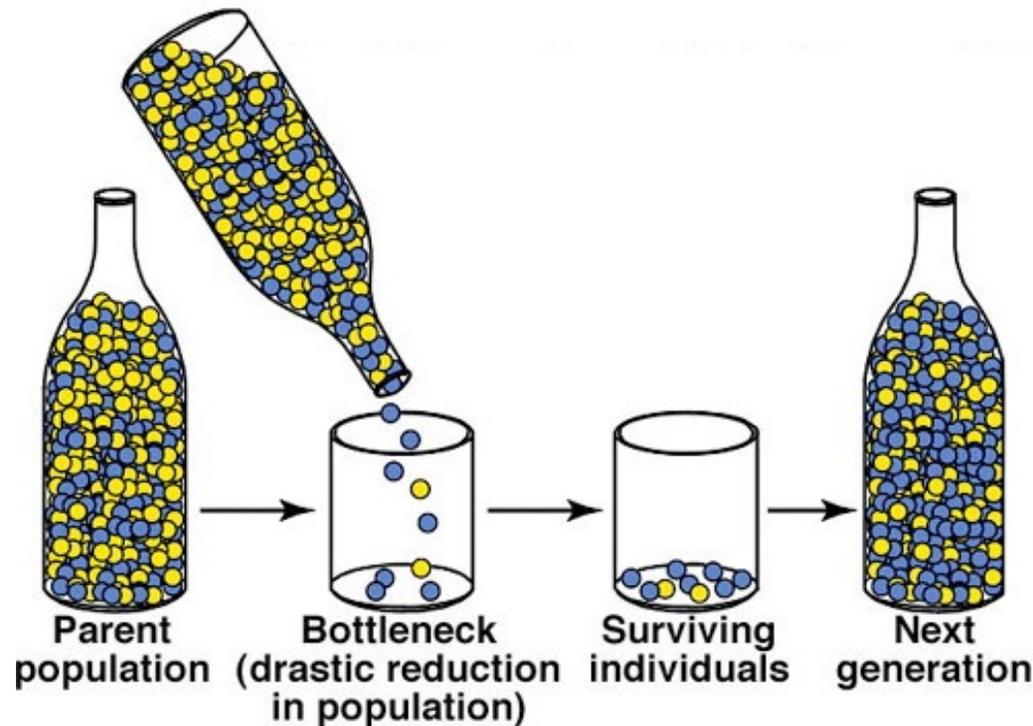


Genetic bottlenecks

- 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?

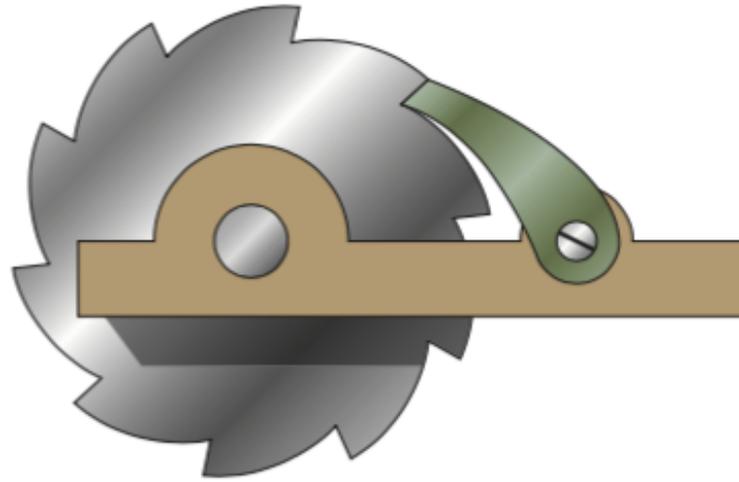
Genetic bottlenecks

- 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



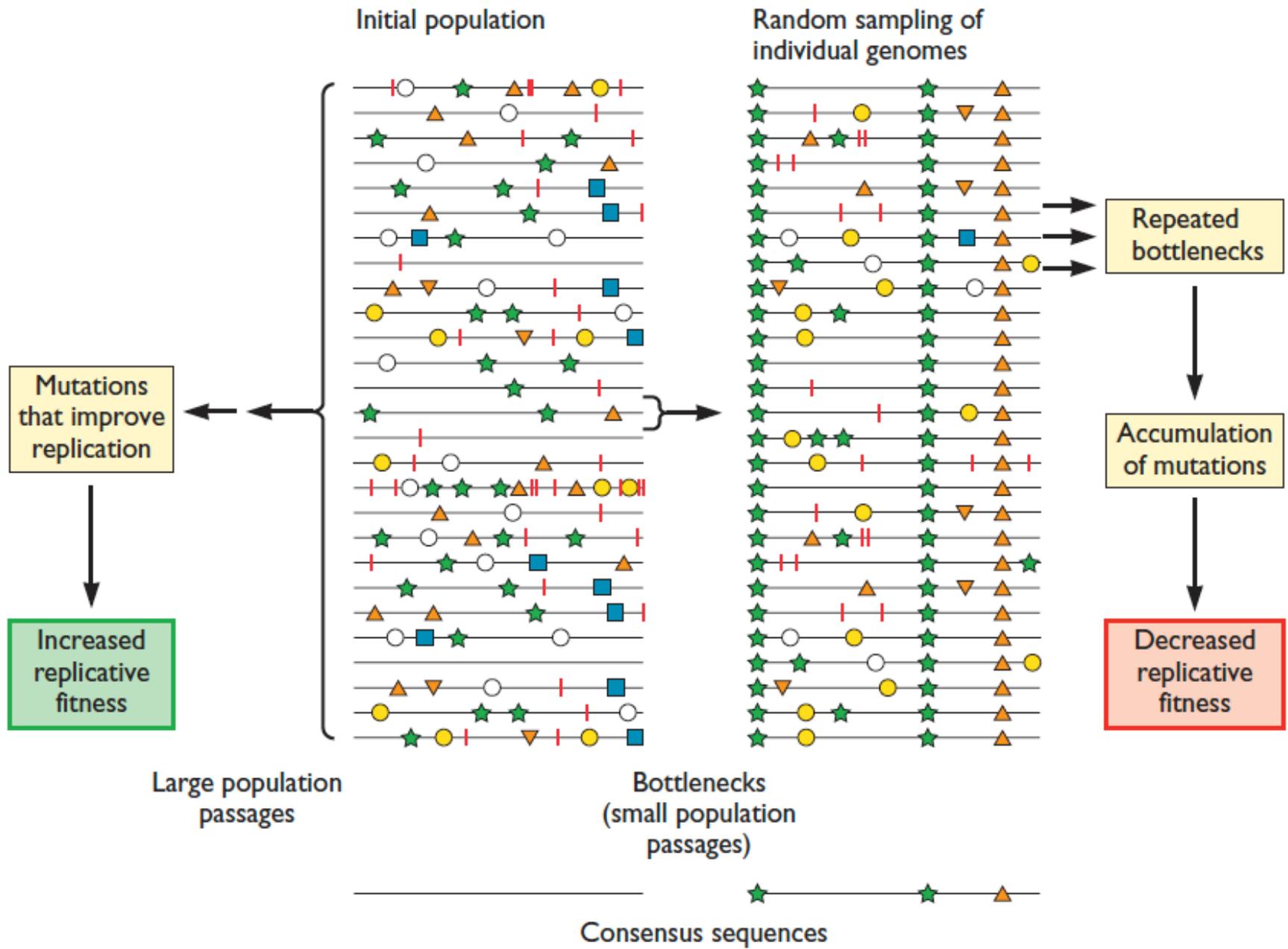
Genetic bottlenecks

- The answer lies in a phenomenon dubbed **Muller's ratchet**: Small, asexual populations decline in fitness over time if the mutation rate is high
- 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



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

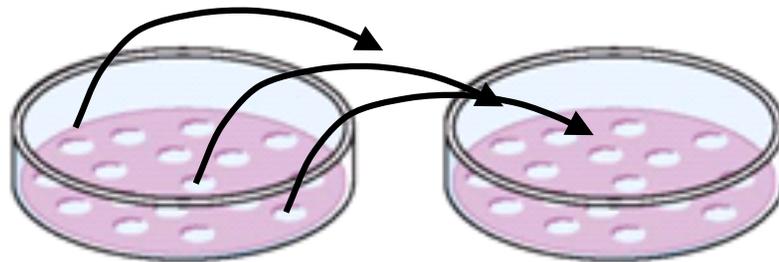
| Virus | # of bottleneck passages | % Decrease in fitness |
|------------------------------|--------------------------|-----------------------|
| Bacteriophage ϕ 6 | 40 | 22 |
| Vesicular stomatitis virus | 20 | 18 |
| Foot-and-mouth disease virus | 30 | 60 |
| HIV | 15 | 94 |
| Bacteriophage MS2 | 20 | 17 |

Bottlenecks in the real world?

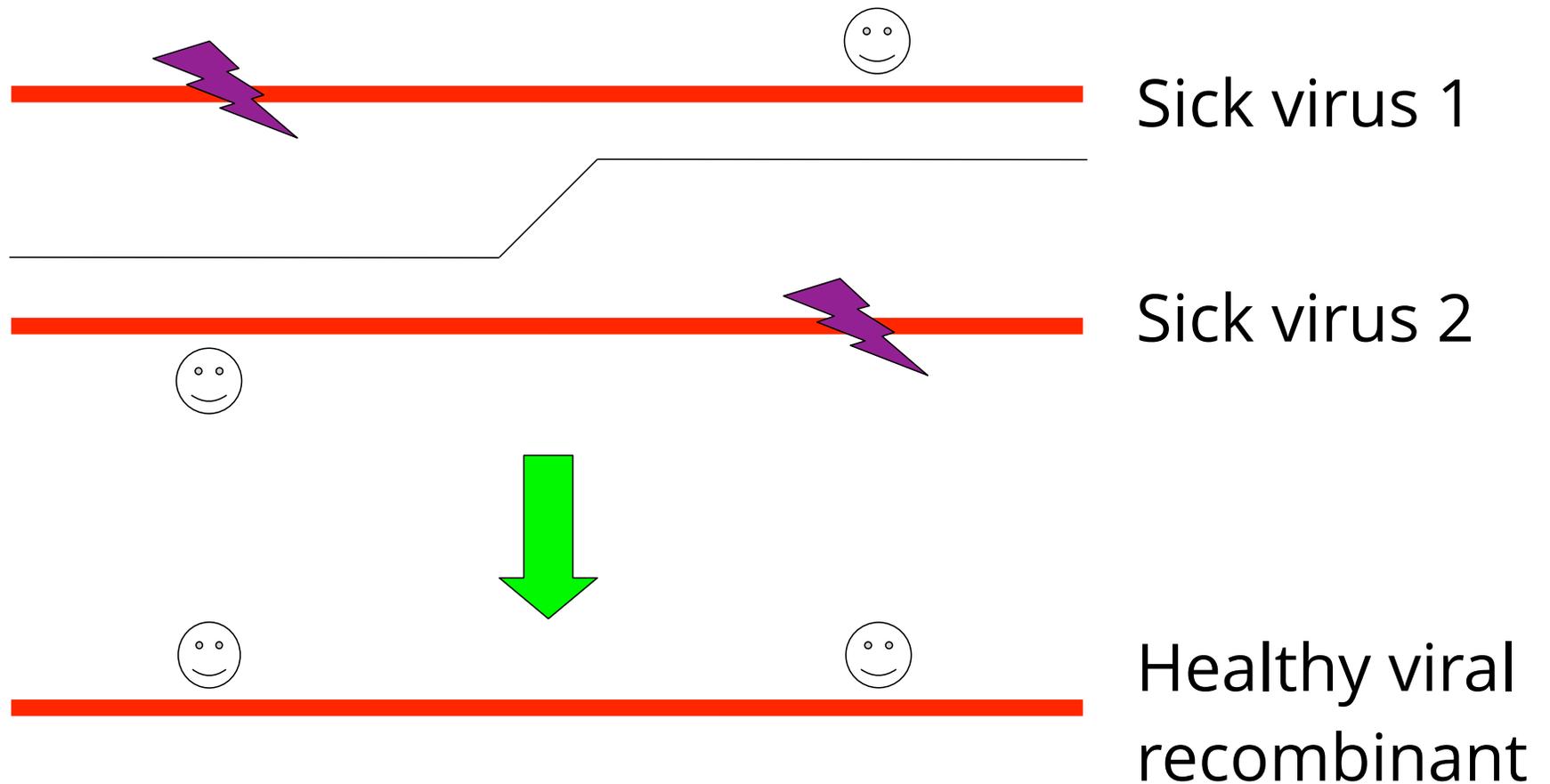
- Infection by a limited virus population and subsequent amplification are often found in nature
 - *Small droplets of suspended virus during aerosol transmission*
 - *Activation of a latent virus from a limited population of cells*
 - *Small volume of inoculum introduced in infection by insect bites*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



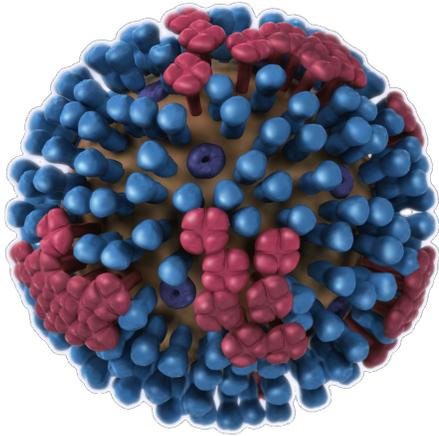
Reassortment

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*

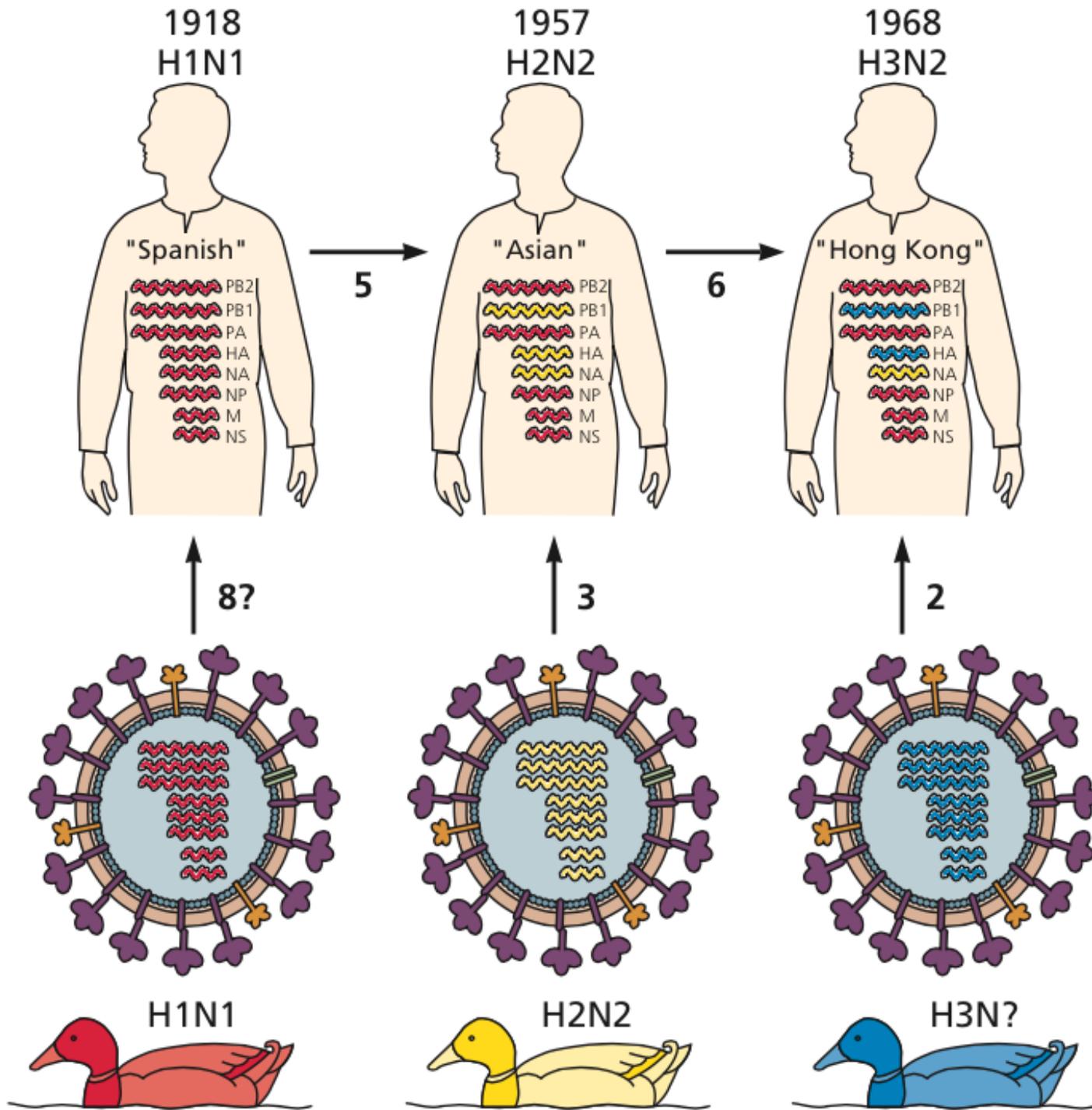
Selection: Genetic shift & drift

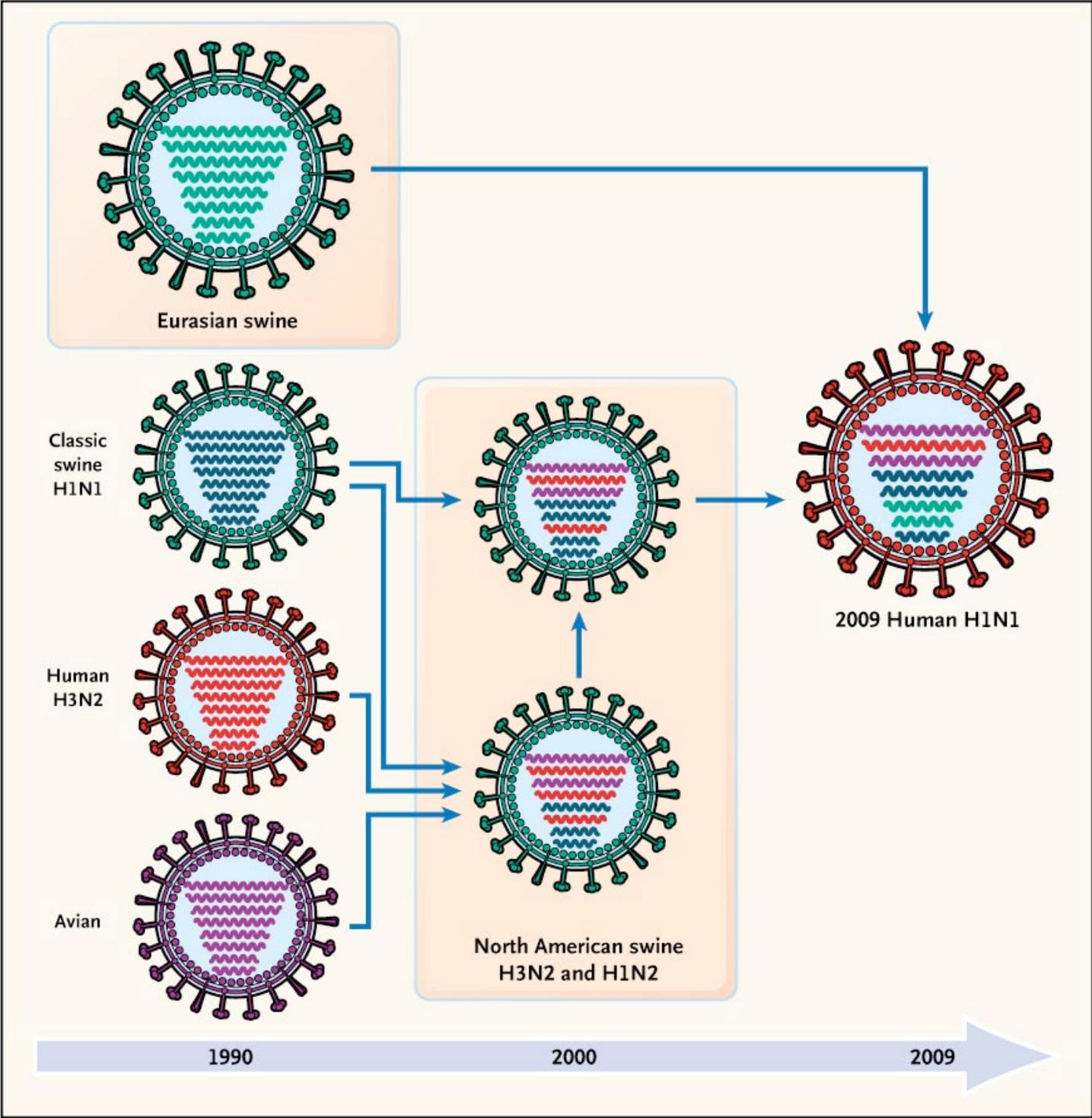
- Selection of viral mutants resistant to elimination by antibodies or cytotoxic T cells inevitable when sufficient virus replication occurs in an immunocompetent individual
- Drift - diversity arising from copying errors and immune selection - may occur each time a genome replicates
- Shift - diversity arising after recombination or reassortment - is relatively rare



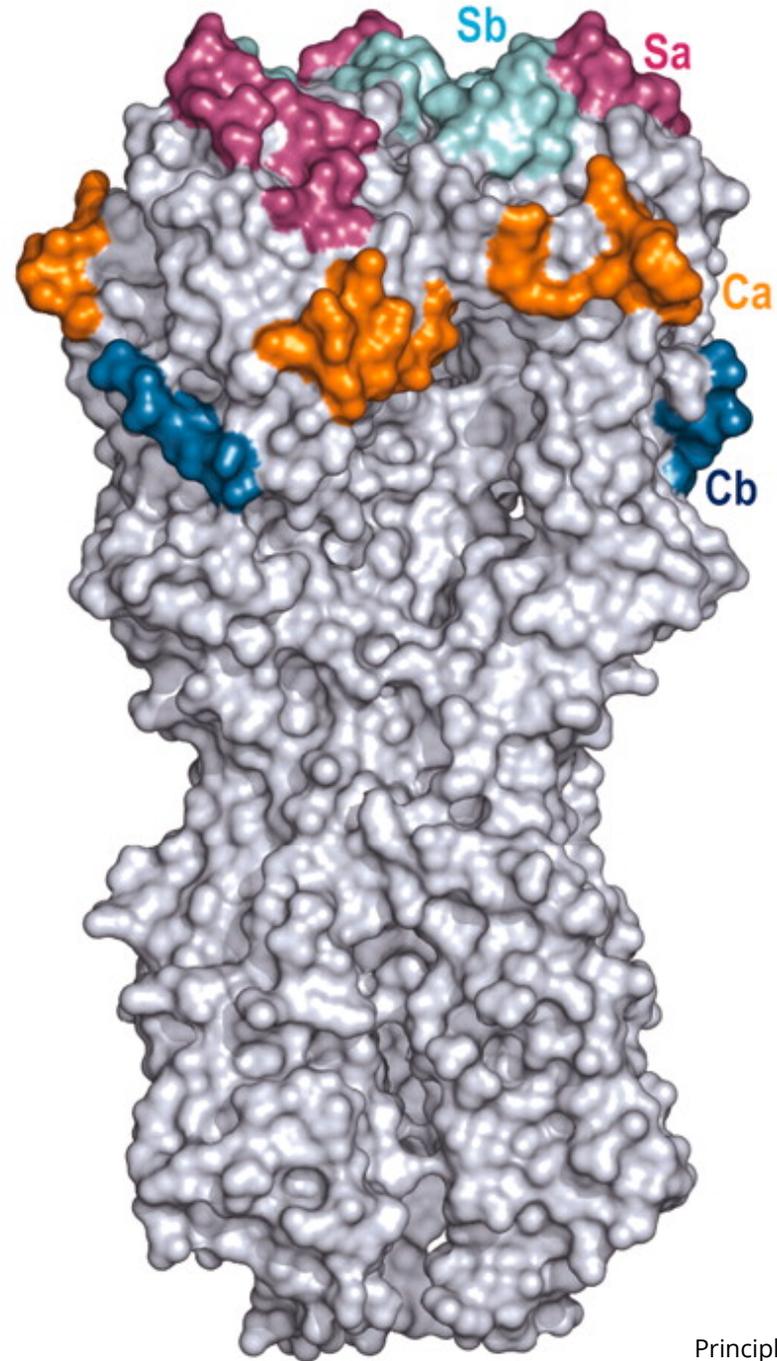
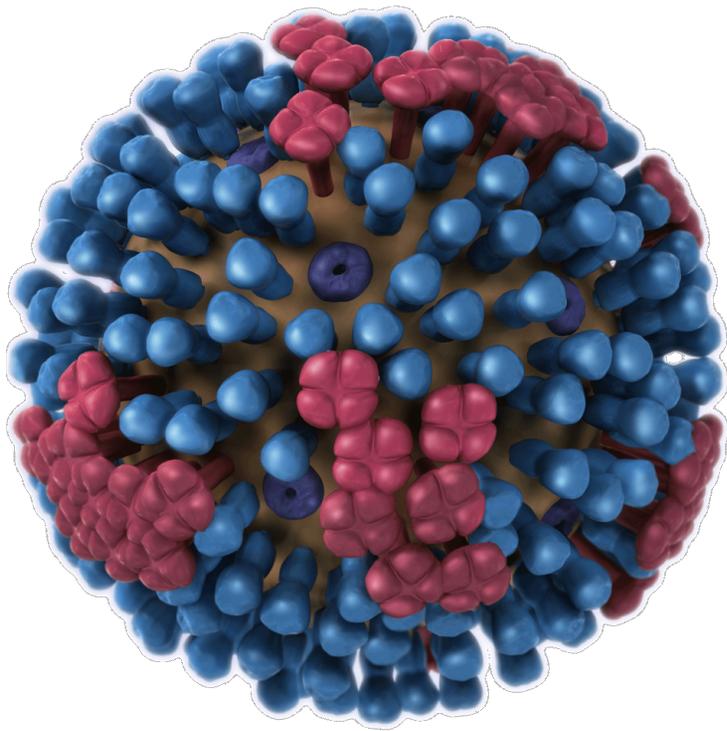
Influenza viruses

- 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

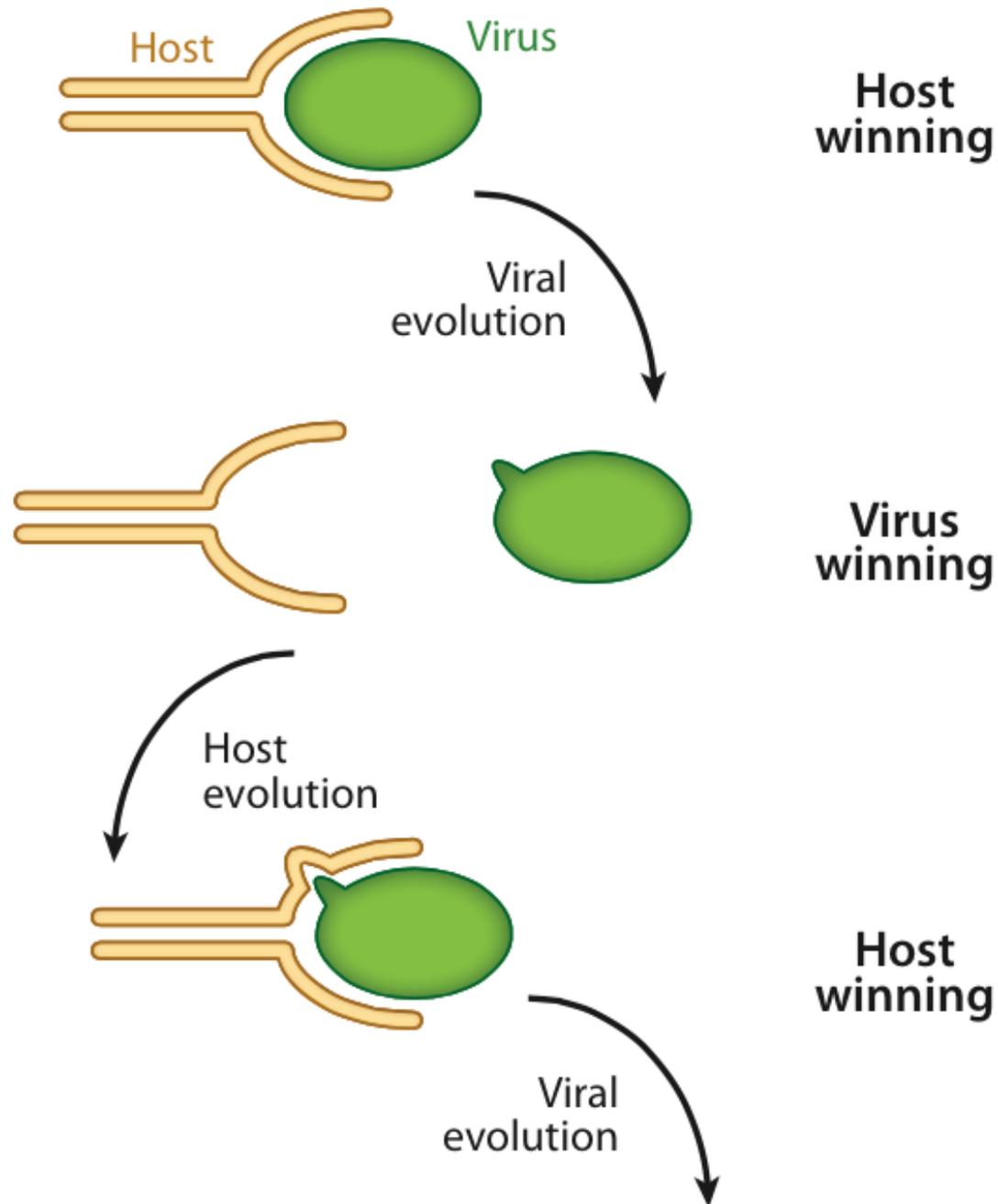




Antigenic drift: Influenza virus

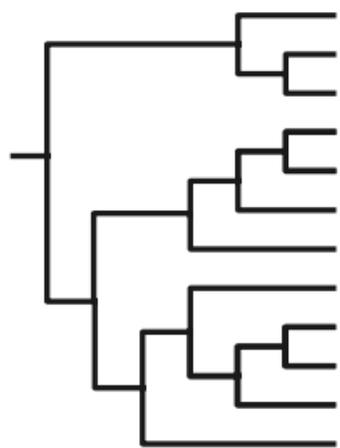


Host-virus arms race





Phase 1: Analyze sequences for positive selection



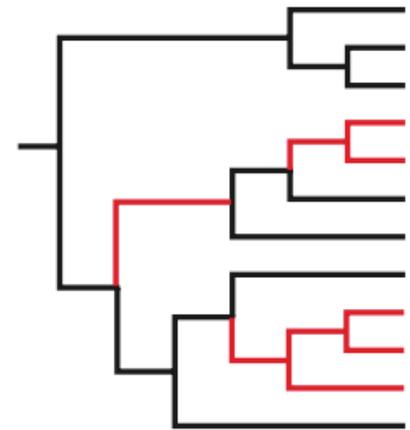
Primate phylogeny

| | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|
| ... | ACA | TGG | GCT | CCG | GAG | ... |
| ... | ACT | TGG | GGT | CCT | GAG | ... |
| ... | ACA | TGG | GGT | CCG | GAA | ... |
| ... | ACA | TGG | AGT | CCG | GAG | ... |
| ... | ACA | TGG | GGT | CCG | GAG | ... |
| ... | ACT | TGG | GAT | CCG | GAG | ... |
| ... | ACA | TGG | GGT | CCG | GGG | ... |
| ... | ACA | TGG | GAT | CCT | GAG | ... |
| ... | ACA | TGG | GGT | CCG | GAG | ... |
| ... | ACA | TGG | GGT | CCG | GAG | ... |
| ... | ACA | TGG | GAT | CCA | GAG | ... |
| ... | ACG | TGG | GGA | CCG | GAA | ... |

Ortholog sequences

Positively selected lineages

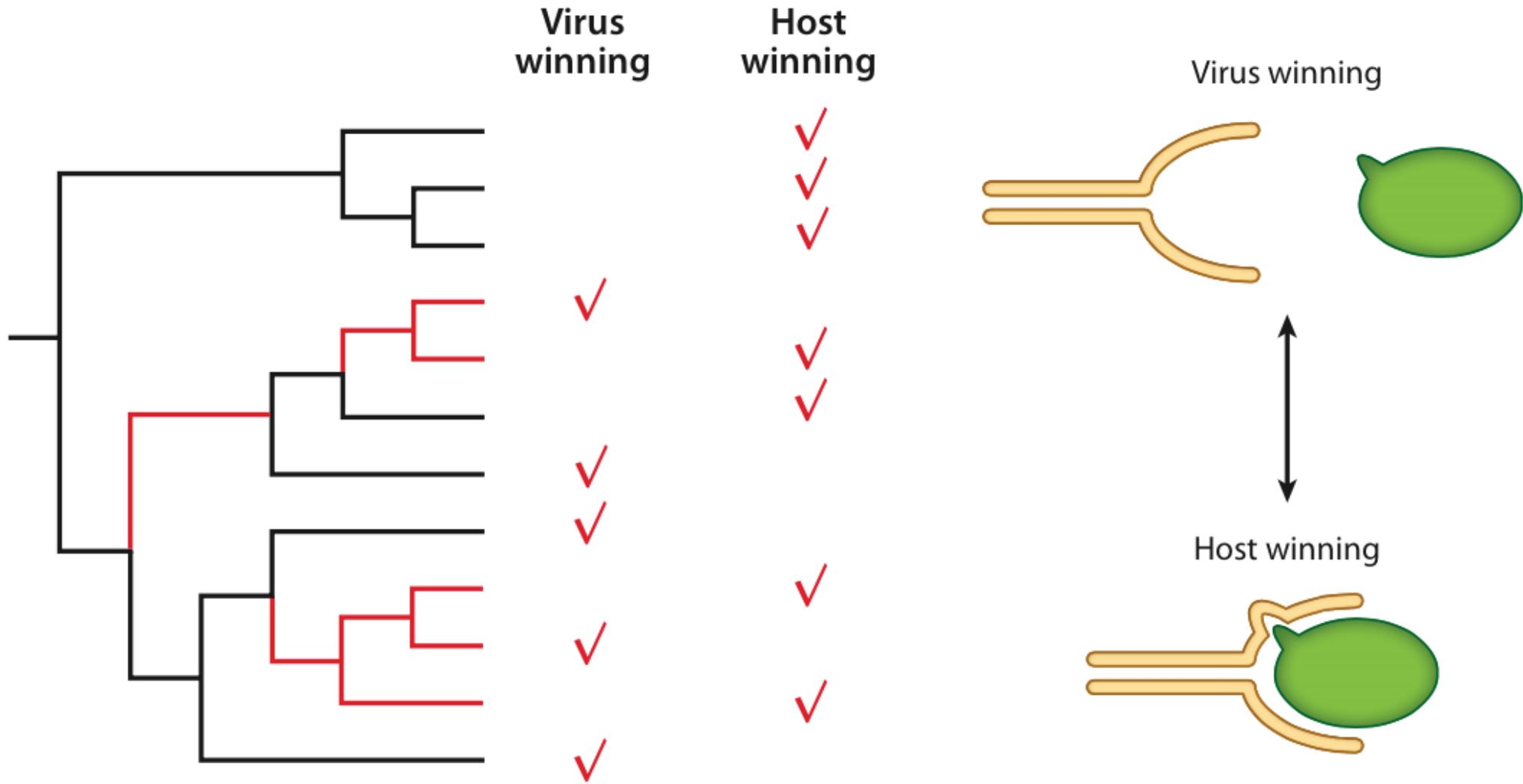
Positively selected codons



| | | |
|-------------|---|---------------|
| GLELHPDYKTW | S | PEQVCSFLRRGGF |
| GPELHPDHKTW | G | PEQVCSFLRRGGF |
| GLELHPDYKTW | G | PEQVCSFLRRGGF |
| GLELHPDYKTW | D | PEQVCSFLRRGGF |
| GLELHPDYKTW | G | PEQVCFFLRRGGF |
| GLELHLDYKTW | D | PEQVCFFLRRGGF |
| GLELHPDYKTW | G | PEQVCFFLRRGGF |
| GLELHPDYKTW | G | PEQVCFFLRRGGF |
| GLELHPDYKTW | D | PEQVCFFLRRGGF |
| GLELDPDYKTW | D | PEQVCSFLGRGGF |



Phase 2: Determine phenotypic consequences of ortholog variation

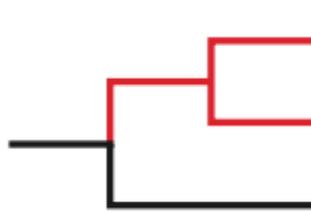




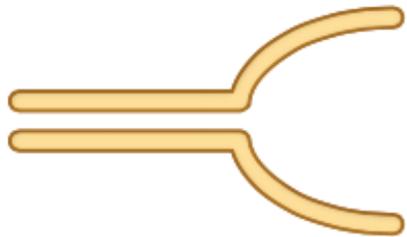
Phase 3: Map positively selected changes onto host-virus interaction

**Virus
winning**

**Host
winning**



GLELHPDYKTW**S**PEQVCSFLRRGGF
GPELHPDHKTW**G**PEQVCSFLRRGGF
GLELHPDYKTW**G**PEQVCSFLRRGGF



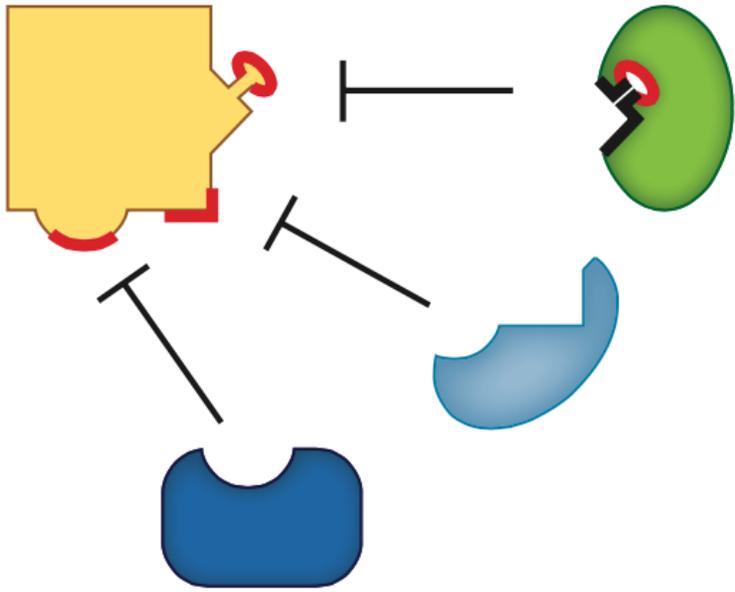
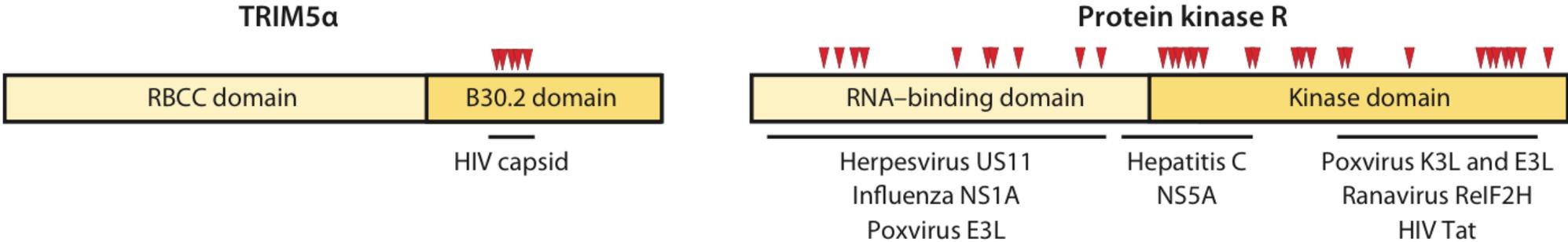
Virus winning

Mutate S → G

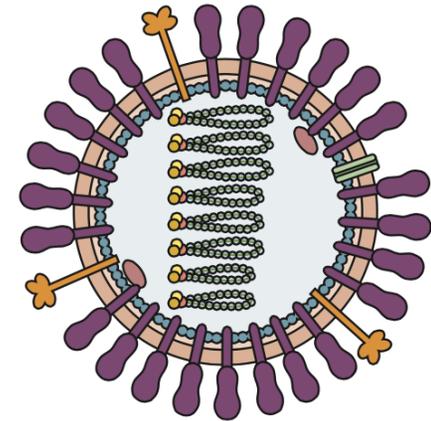
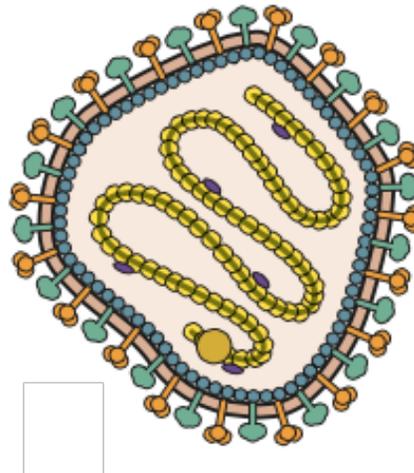
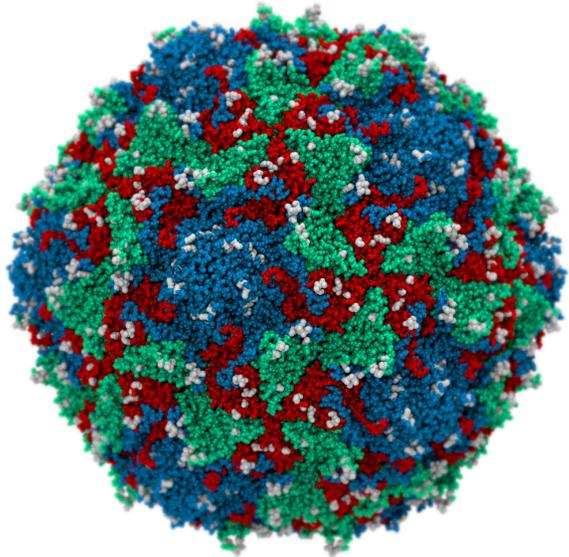


Host winning

Virus-host conflicts have driven evolution of the immune system



Despite this genome diversity...



- There are only 3 serotypes of poliovirus, but >150 of rhinoviruses
- One measles serotype, continuous influenza variation
- Why?

Selection: Is virulence a positive or negative trait?

- Idea: increased virulence reduces transmissibility because hosts die faster, reducing exposure to uninfected hosts
- Expectation: all viruses evolve to be maximally infectious and avirulent
- But this is not observed - there are many virulent viruses

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

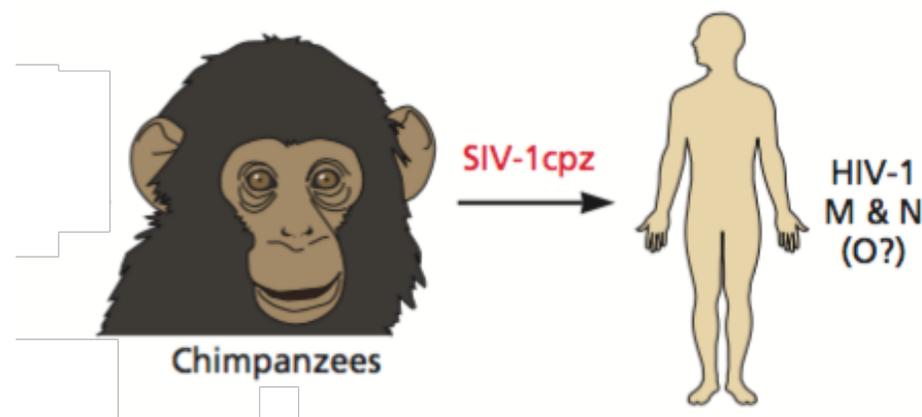
Evolution of virulence

- 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

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, the press is obsessed with increased viral virulence

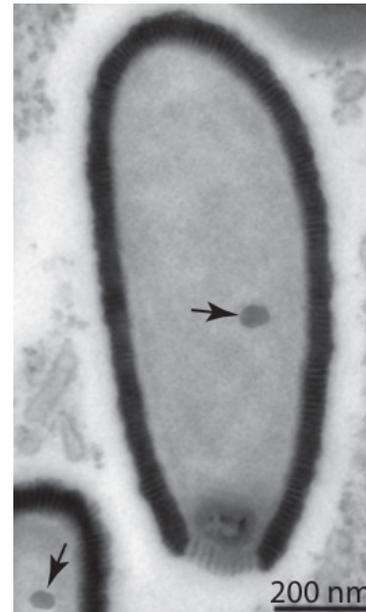
- Ebolavirus is mutating, will go airborne (Osterholm, NY Times, 11 Sept 2014)
- Ebolavirus, Lisa Henley: “Is it getting better at replicating as it goes from person to person?” (*Ebola Wars*, Richard Preston, *New Yorker*)
- Peter Hotez, *NYT* Op-Ed 8 April 2016: “There are many theories for Zika’s 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.”

We have no data on the effect of evolution on virulence of human viruses

The origin of viruses

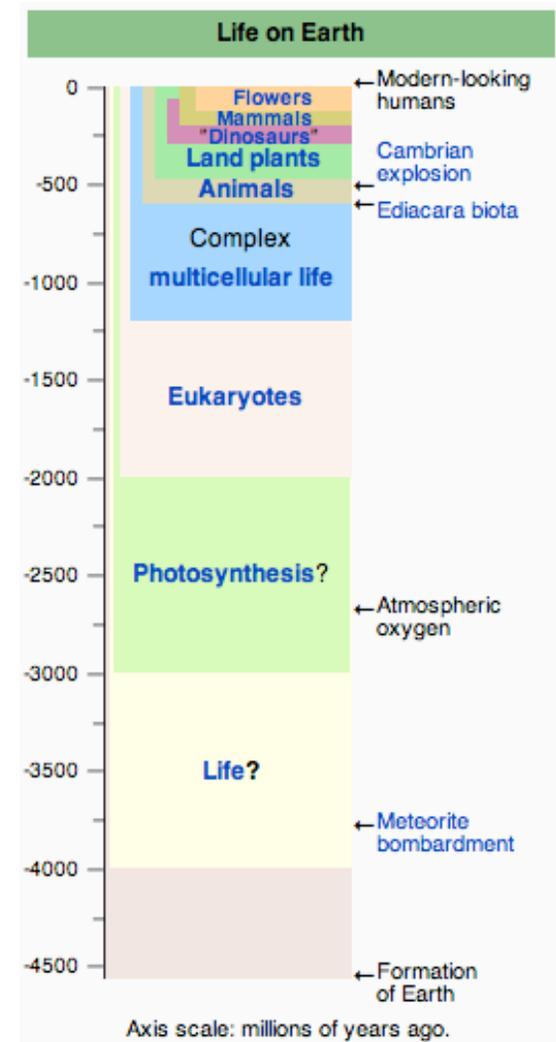


Oldest viral stocks:
1918 influenza virus
Pithovirus sibericum
(30,000 y)



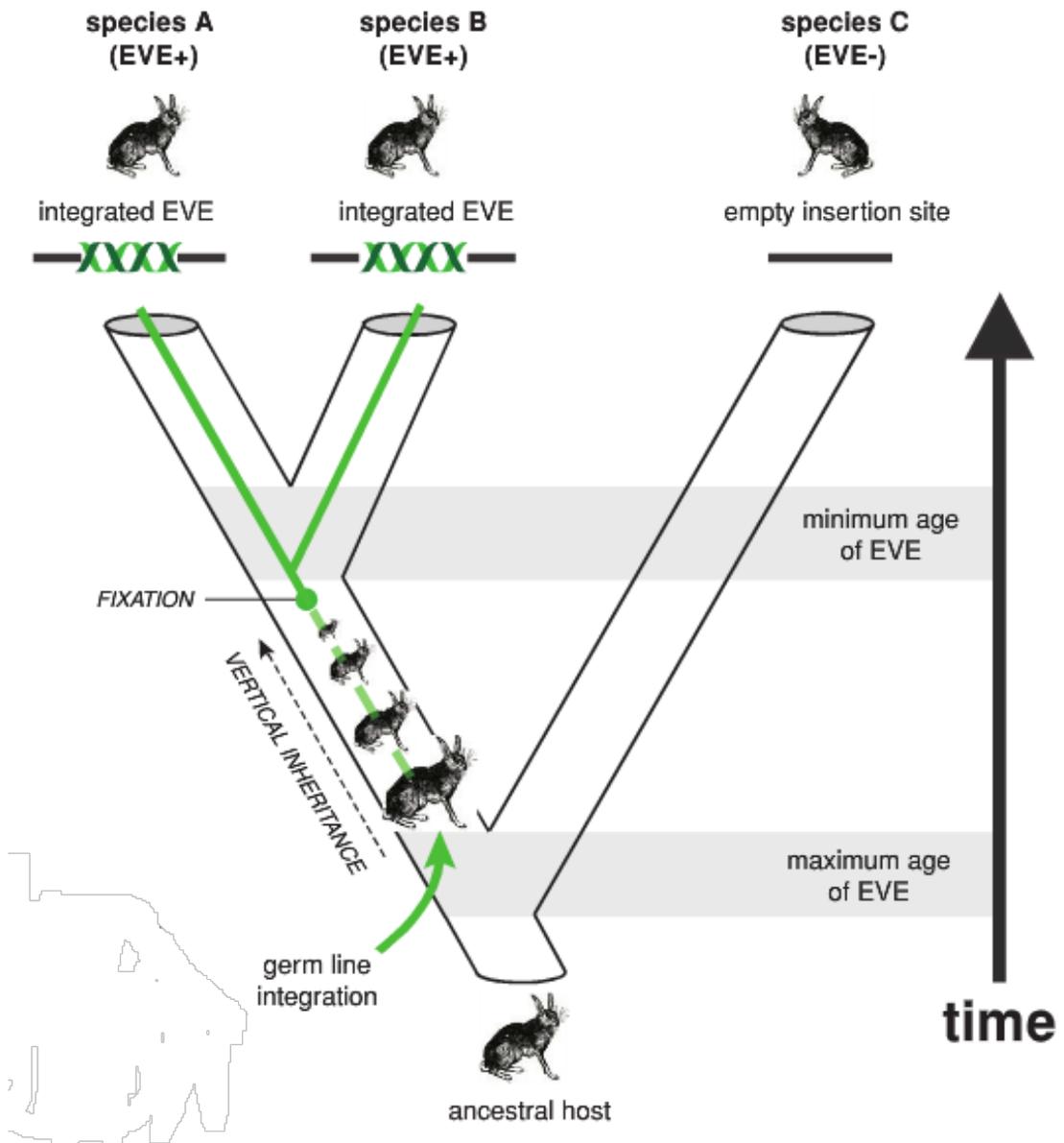
Origins of DNA viruses

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

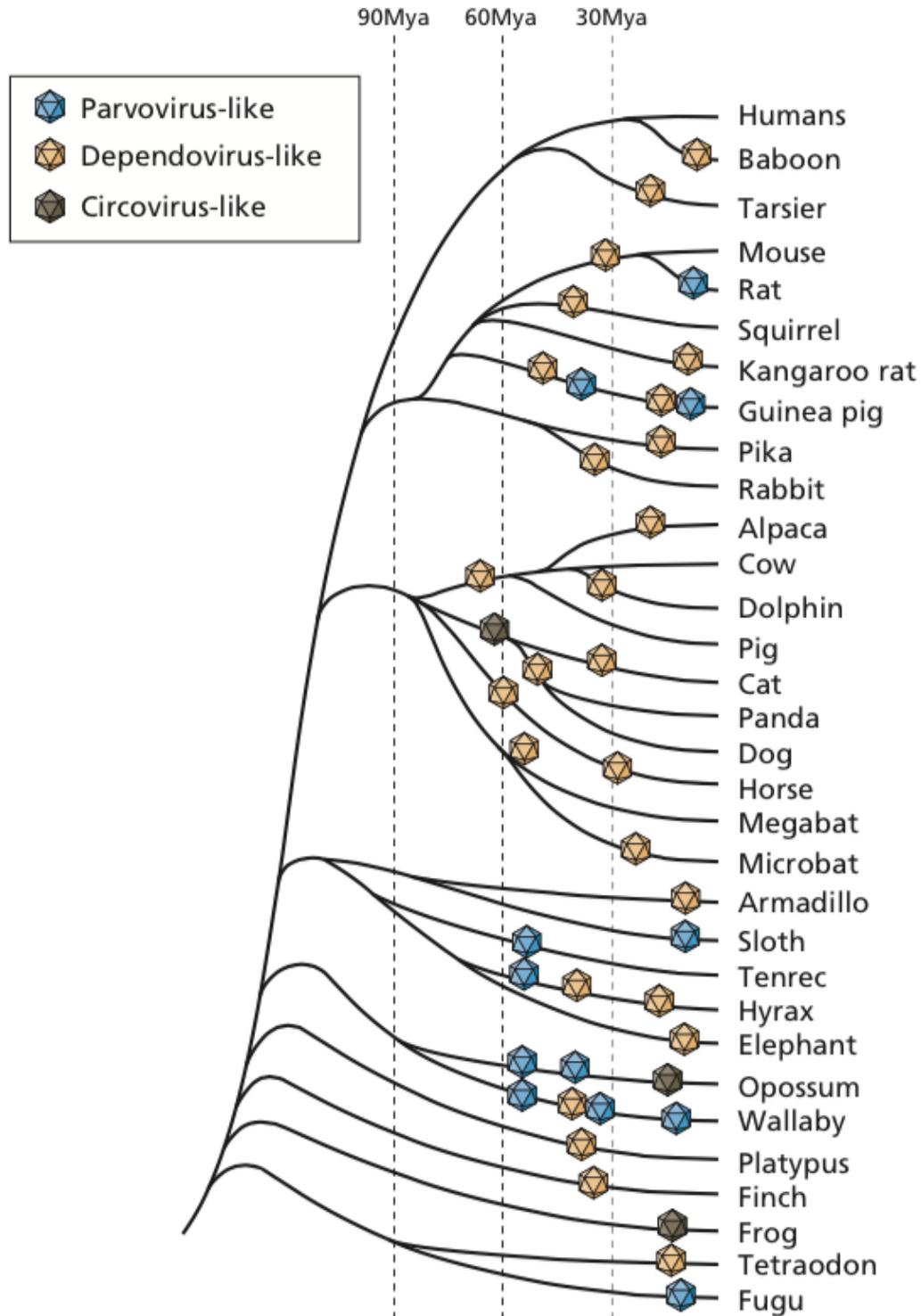


Endogenous viruses - retrovirus and otherwise

Phylogenomics

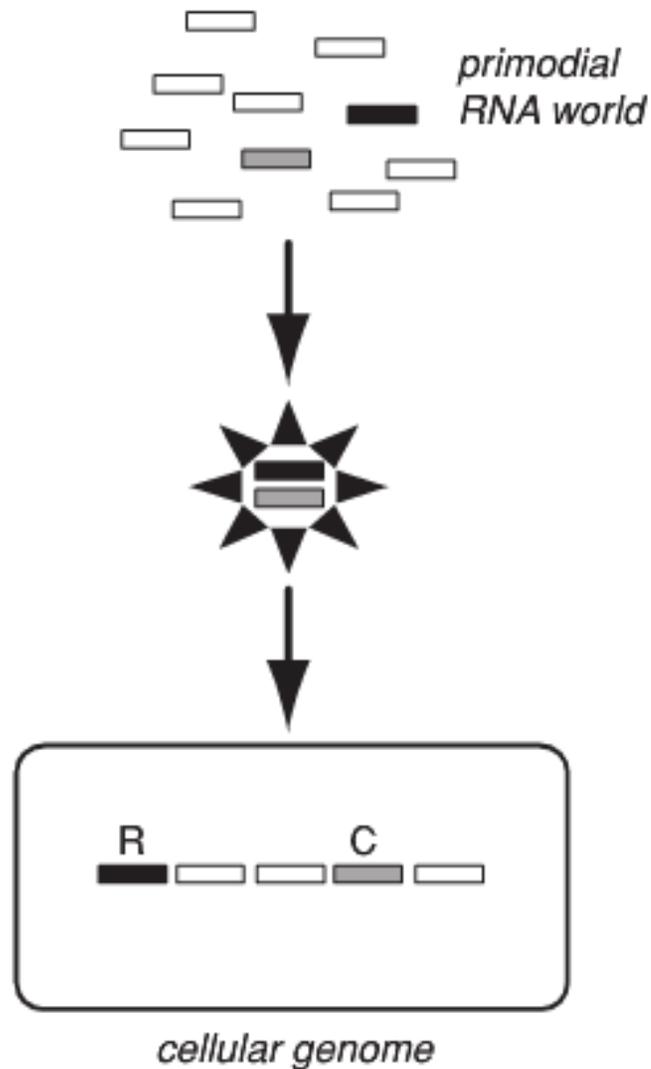


History of ssDNA virus integrations

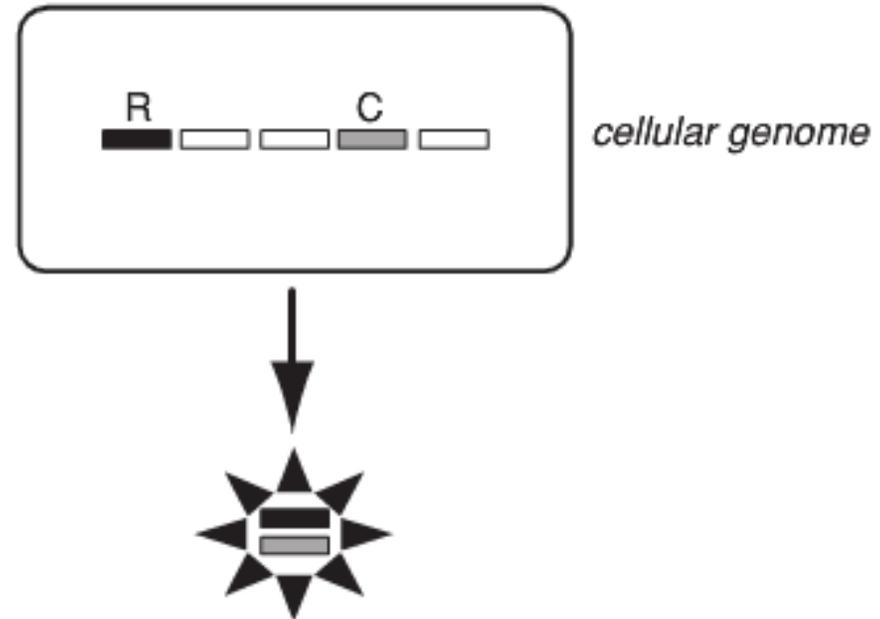


Origins of viruses

A Pre-cellular origin



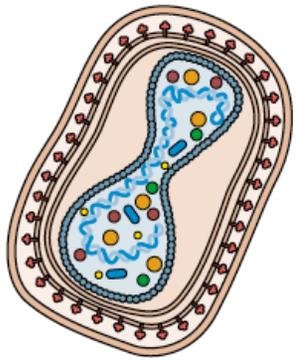
B Escaped host gene



Supported by discovery of protein structures conserved among diverse viruses with little sequence similarity

Human viruses

- All known types of viruses likely evolved long before humans appeared on Earth
- All human viruses have therefore evolved from animal viruses



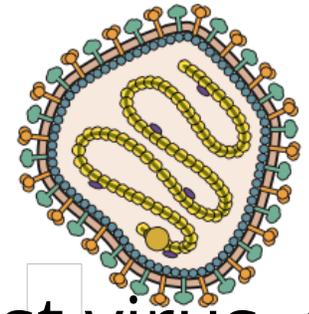
Origins of smallpox virus



- Sequence analysis of 45 isolates: genomes can be organized into 3 clades, which cluster geographically (West Africa, South America, Asia)
- Gene content is constant; lack of diversity indicates recent introduction into humans
- Perhaps human smallpox virus arose after a zoonotic infection from infected 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
- Established in the Middle East ~5,000 years ago, when human populations began to congregate in cities
- 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 increases the numbers
- $\sim 4^{135}$ atoms in the visible universe

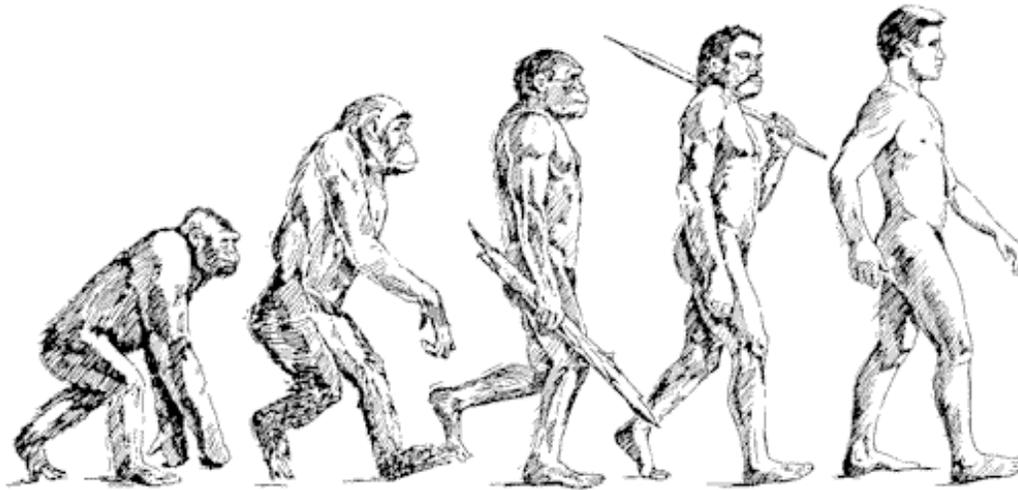
The fundamental properties of viruses constrain and drive evolution

- 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?

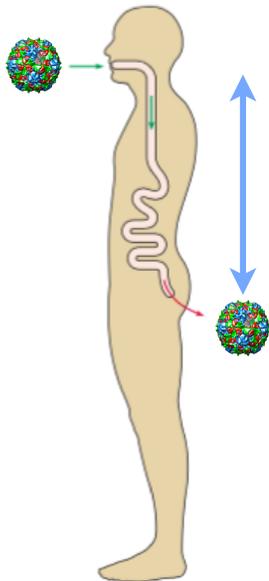
Constraining viral evolution

- Extreme alterations in viral consensus genome do not survive selection
- The viral genome is one constraint
 - *DNA cannot become RNA, or vice versa*
 - *Replication strategy - cannot change; consider interaction with host proteins*
- 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, suffer the same fate as one that does not replicate efficiently enough*

Food for thought



8 million years



5 days

Imagine what a virus can do with 8 million years