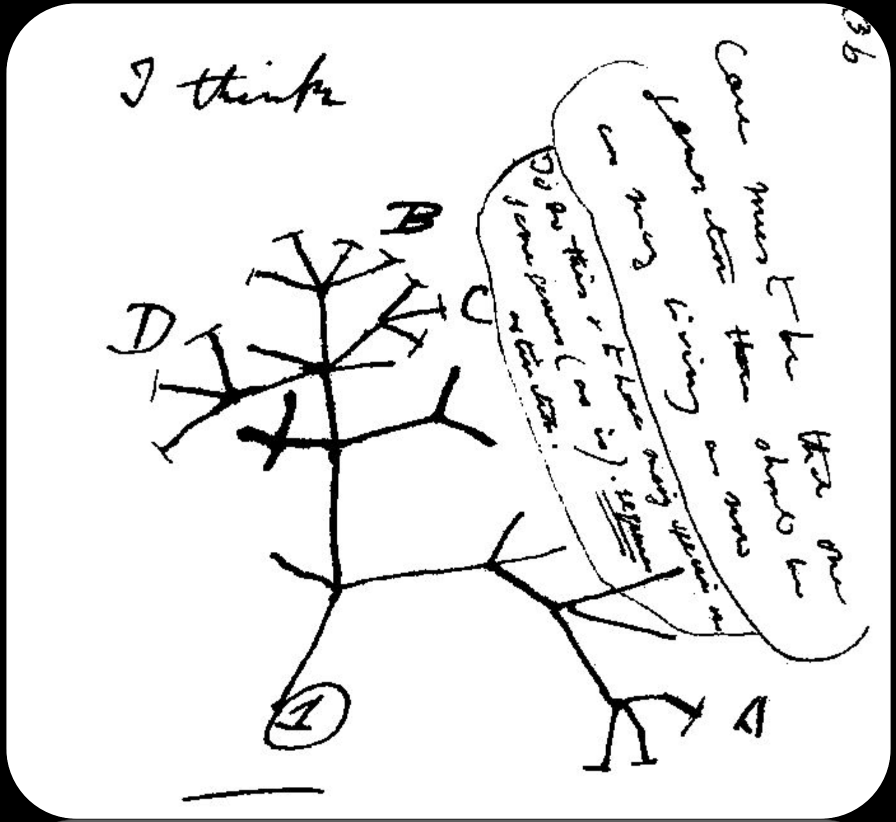


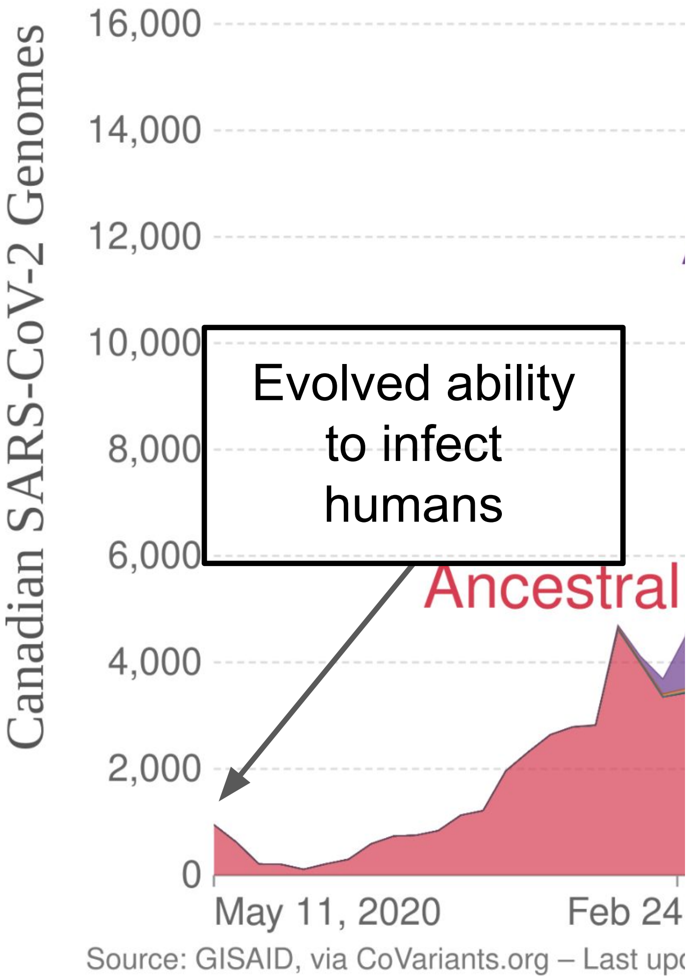
Sequence change and evolution

Overview

1. Mutations are changes that are passed from generation to generation
2. Many different types of mutation are possible, each can have different impacts on encoded proteins
3. Mutations can have consequences for the organism, leading to differential fitness, selection, and speciation

Why do we care about evolution?

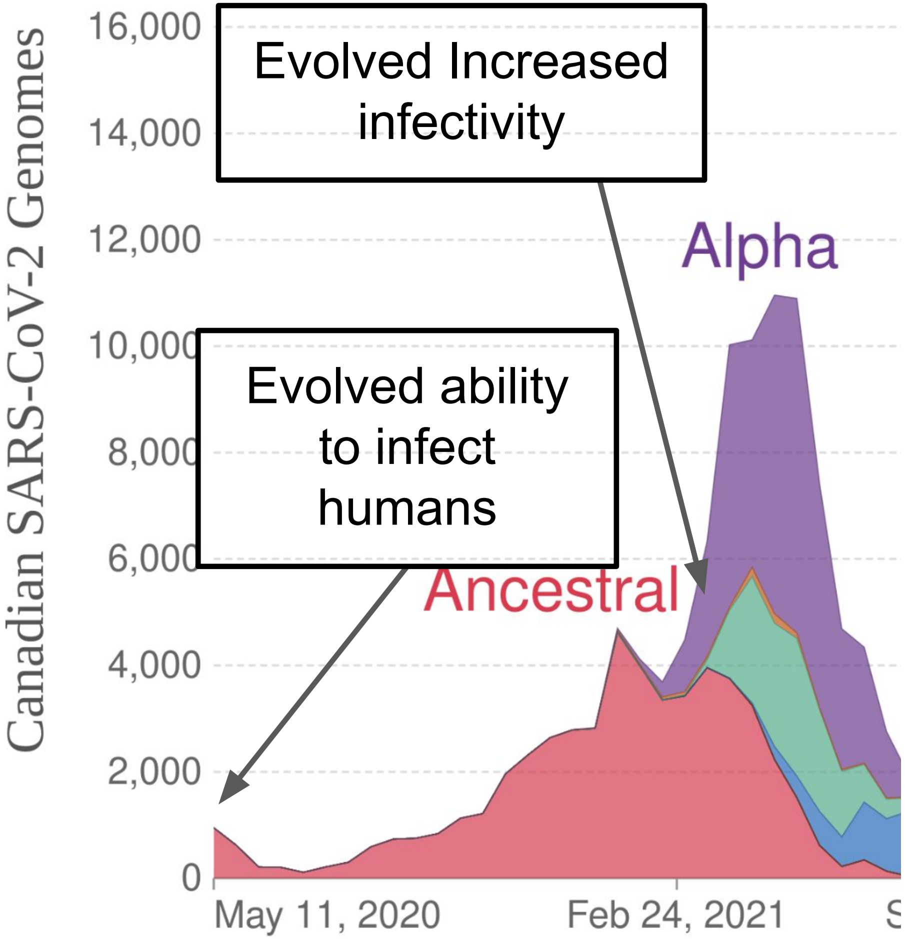
Evolution drives cases: SARS-CoV-2 waves



Evolved ability to infect humans

Source: GISAID, via CoVariants.org – Last up

Evolution drives cases: SARS-CoV-2 waves

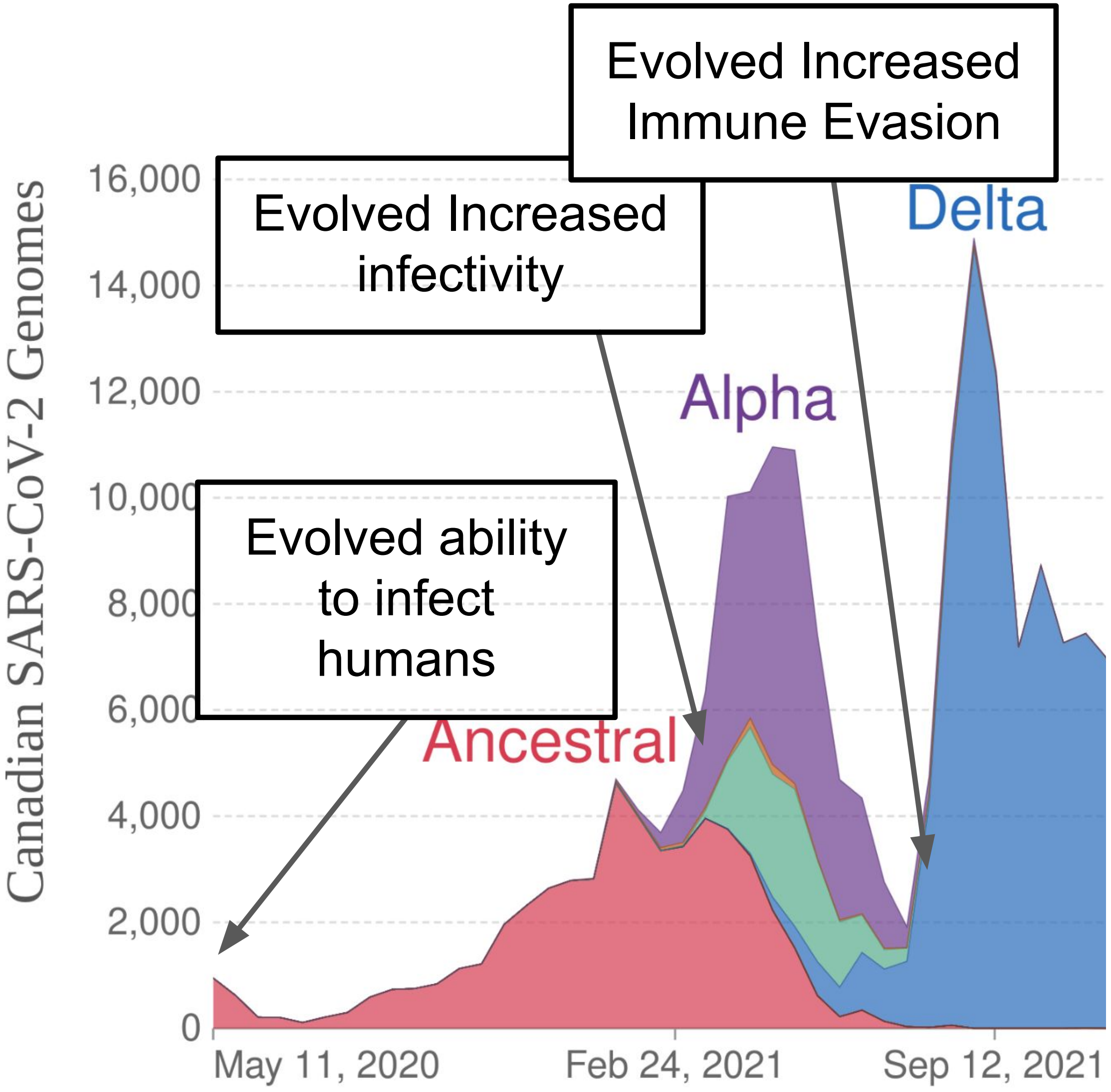


Evolved Increased infectivity

Evolved ability to infect humans

Source: GISAID, via CoVariants.org – Last updated 1 Decembe

Evolution drives cases: SARS-CoV-2 waves



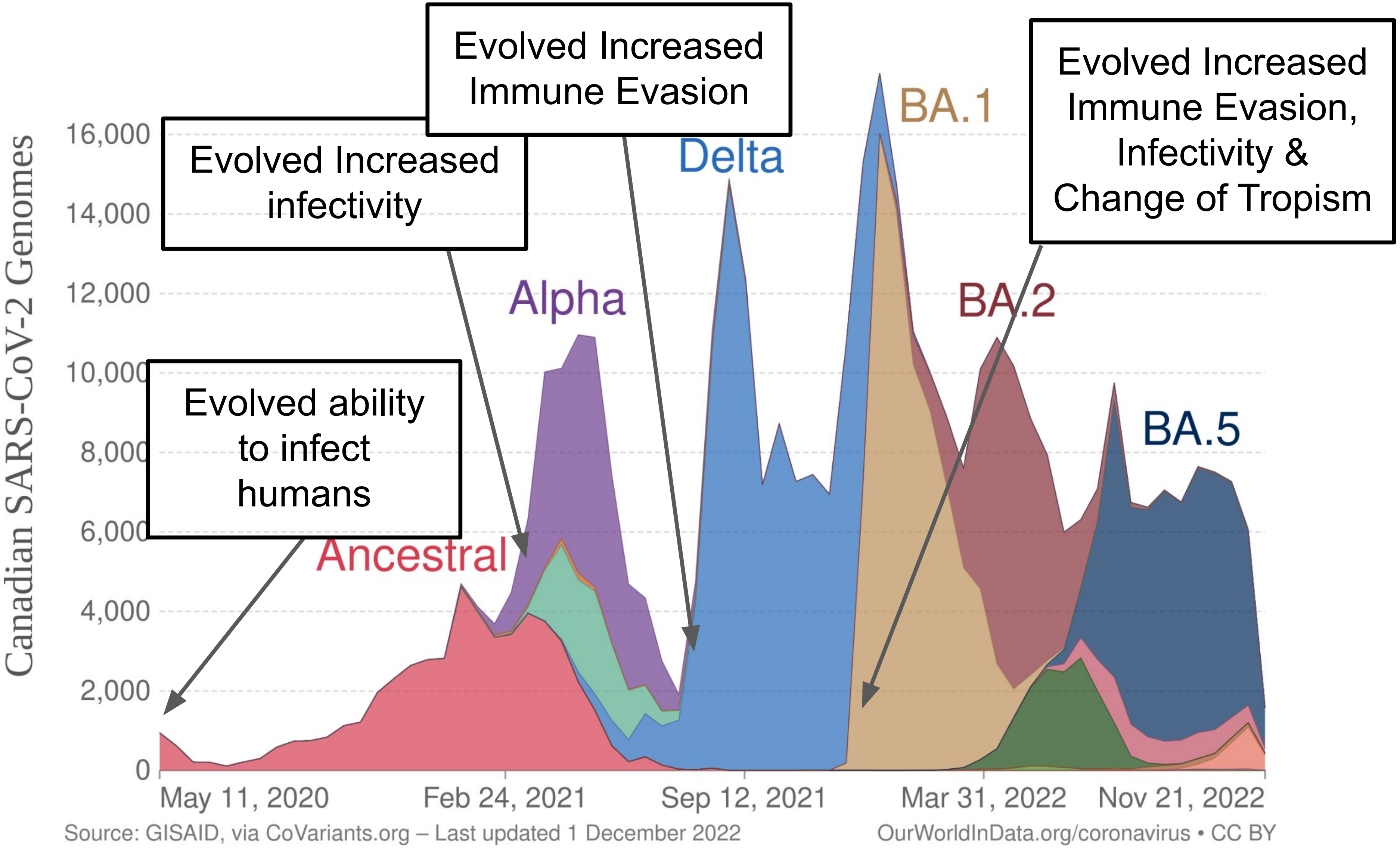
Evolved Increased Immune Evasion

Evolved Increased infectivity

Evolved ability to infect humans

Source: GISAID, via CoVariants.org – Last updated 1 December 2022

Evolution drives cases: SARS-CoV-2 waves



Evolved Increased Immune Evasion

Evolved Increased infectivity

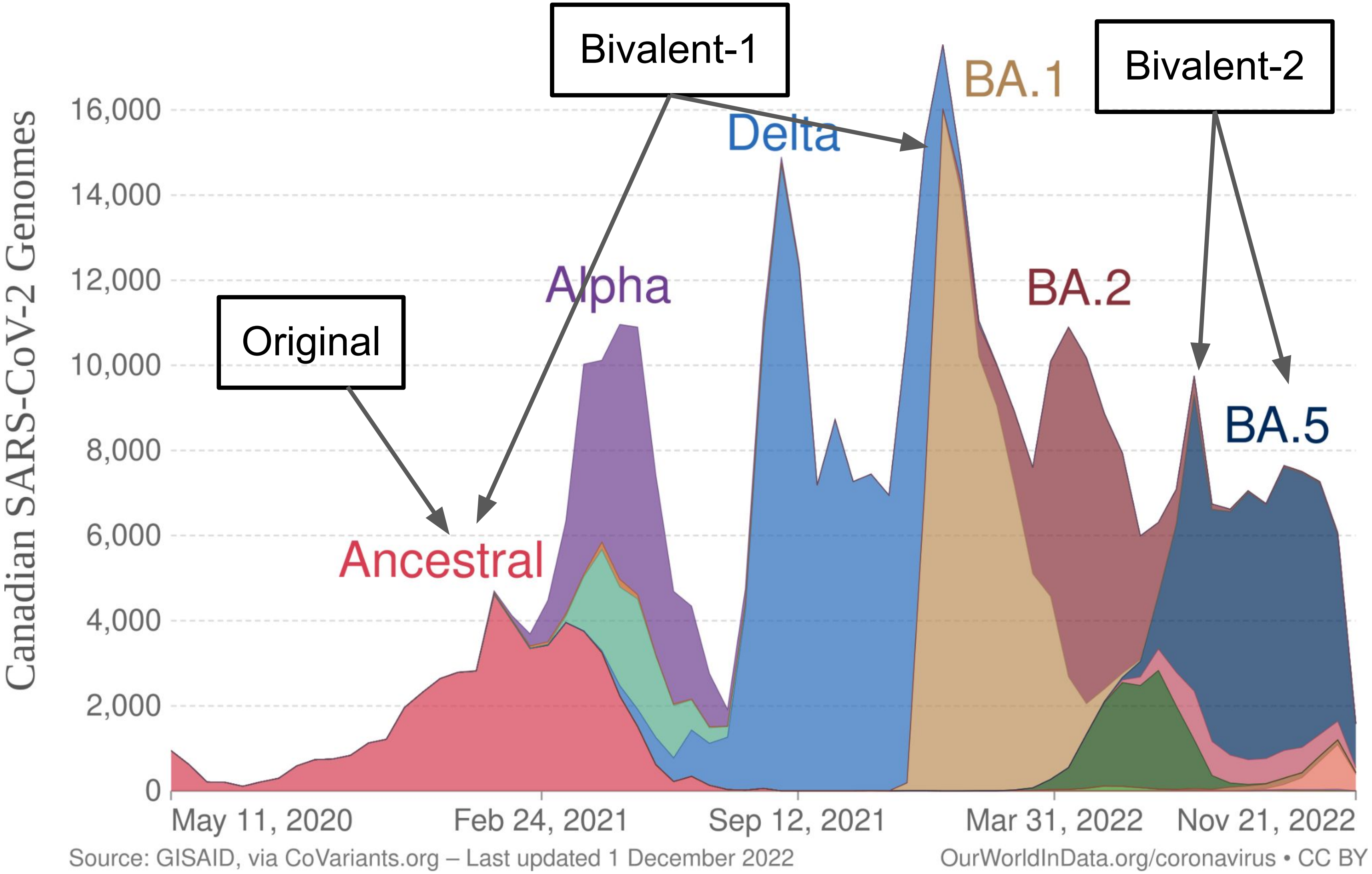
Evolved ability to infect humans

Evolved Increased Immune Evasion, Infectivity & Change of Tropism

Source: GISAID, via CoVariants.org – Last updated 1 December 2022

OurWorldInData.org/coronavirus • CC BY

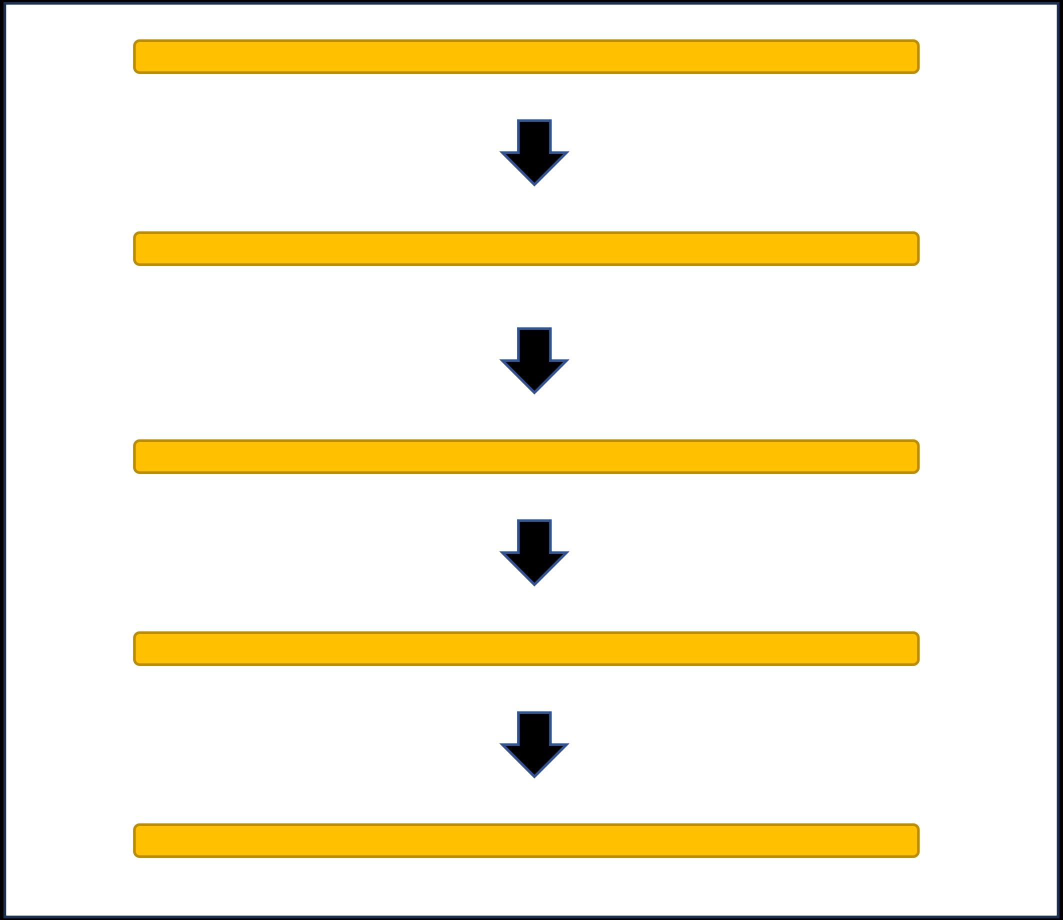
Evolution drives vaccine design/effectiveness



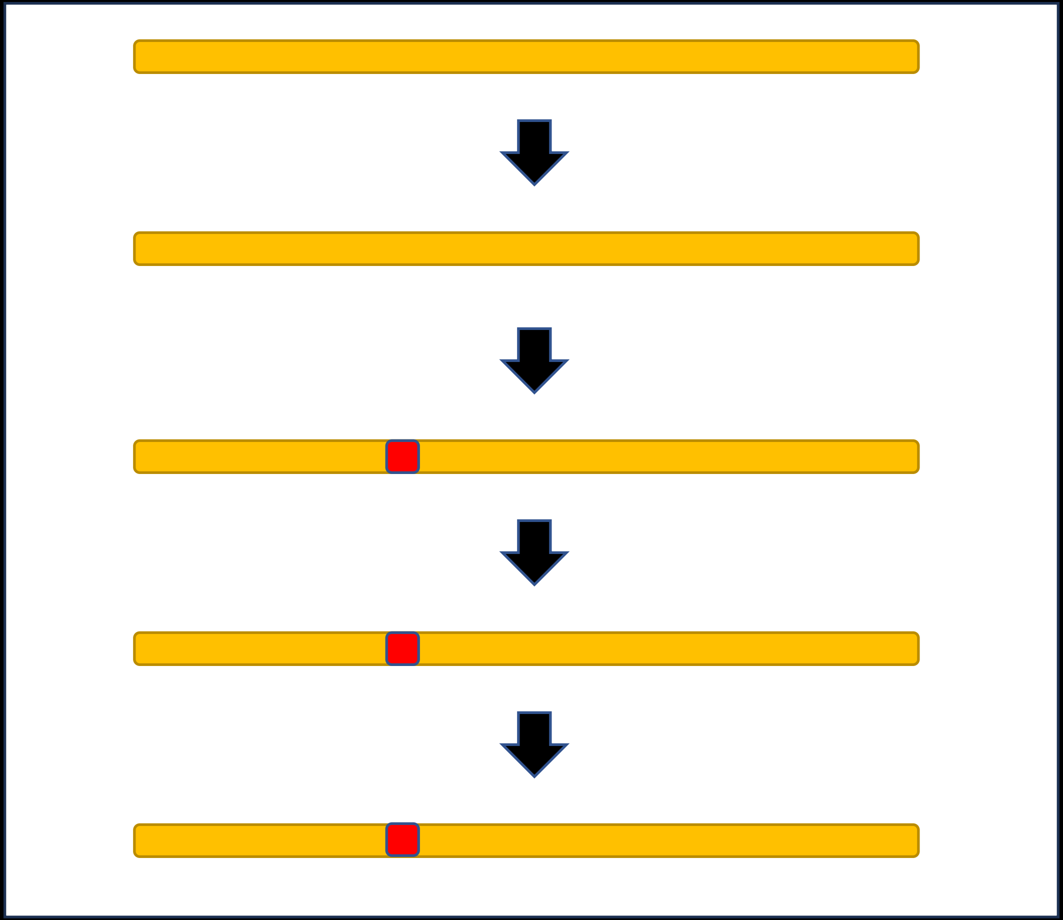
Source: GISAID, via CoVariants.org – Last updated 1 December 2022

OurWorldInData.org/coronavirus • CC BY

DNA is passed from parent to offspring through *replication*



Replication is not always 100% faithful – *mutations* can occur when mistakes are made



Point mutations: One nucleotide replaced by another

- May be SILENT/synonymous (remember codons!)

GCC (Ala) → GCA (Ala)

- Or may change a single amino acid (*missense mutation*)

GCC (Ala) → CCC (Pro)

- *Or* may introduce a stop codon (*nonsense mutation*)

CAA (Gln) → TAA (Stop)

Other types of non-synonymous change

- Insertions

  AAA ACC GAA TCA CCG GAT → AAA ACC G**C**A ATC ACC GGA T..

- Deletions

  AAA ACC GAA TC**A** CCG GAT → AAA ACC GAA TCC CGG AT.

- Inversions

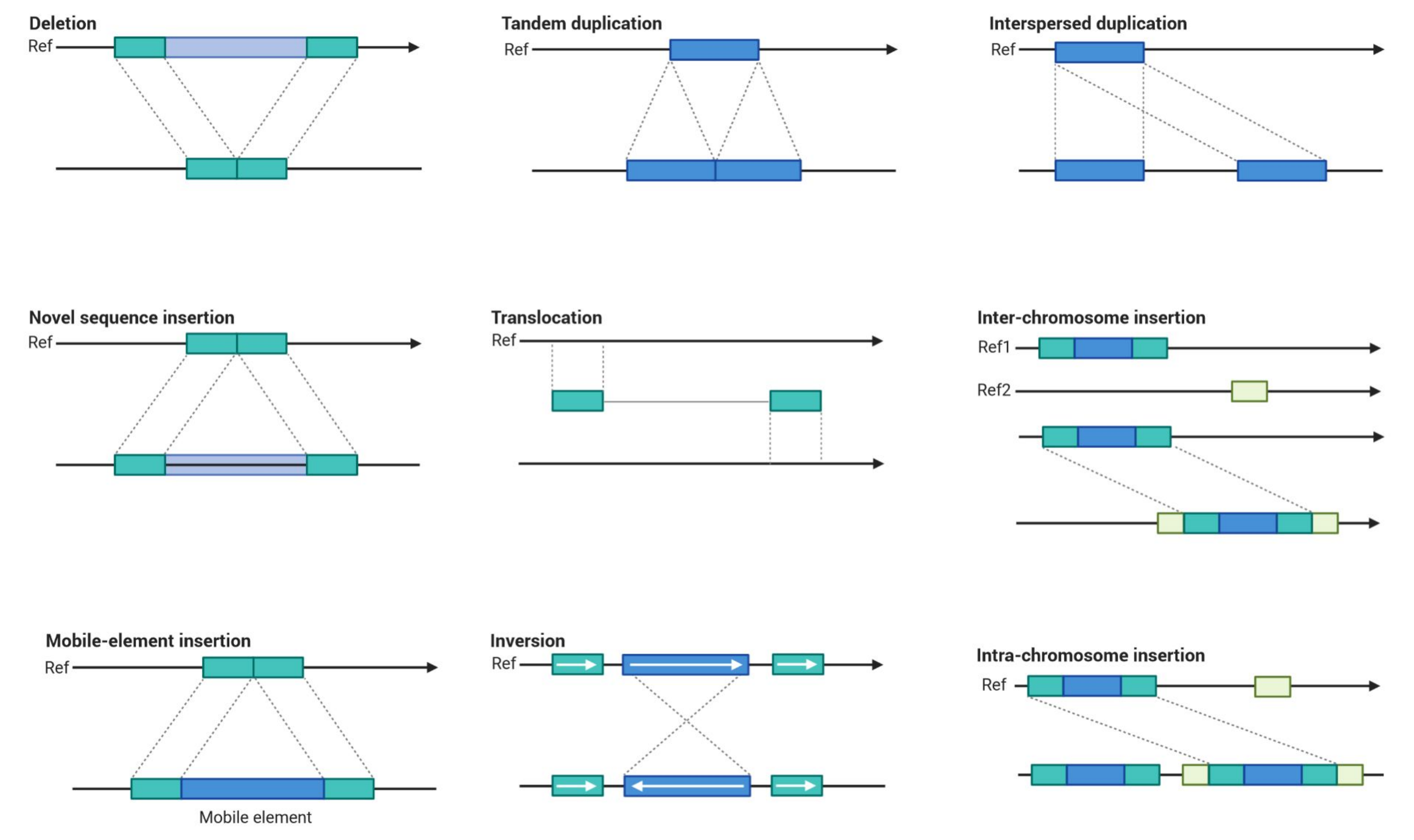
  AAA ACC **GAA TCA** CCG GAT → AAA ACC **TGA TTC** CCG GAT

- Duplications (up the whole genome!)

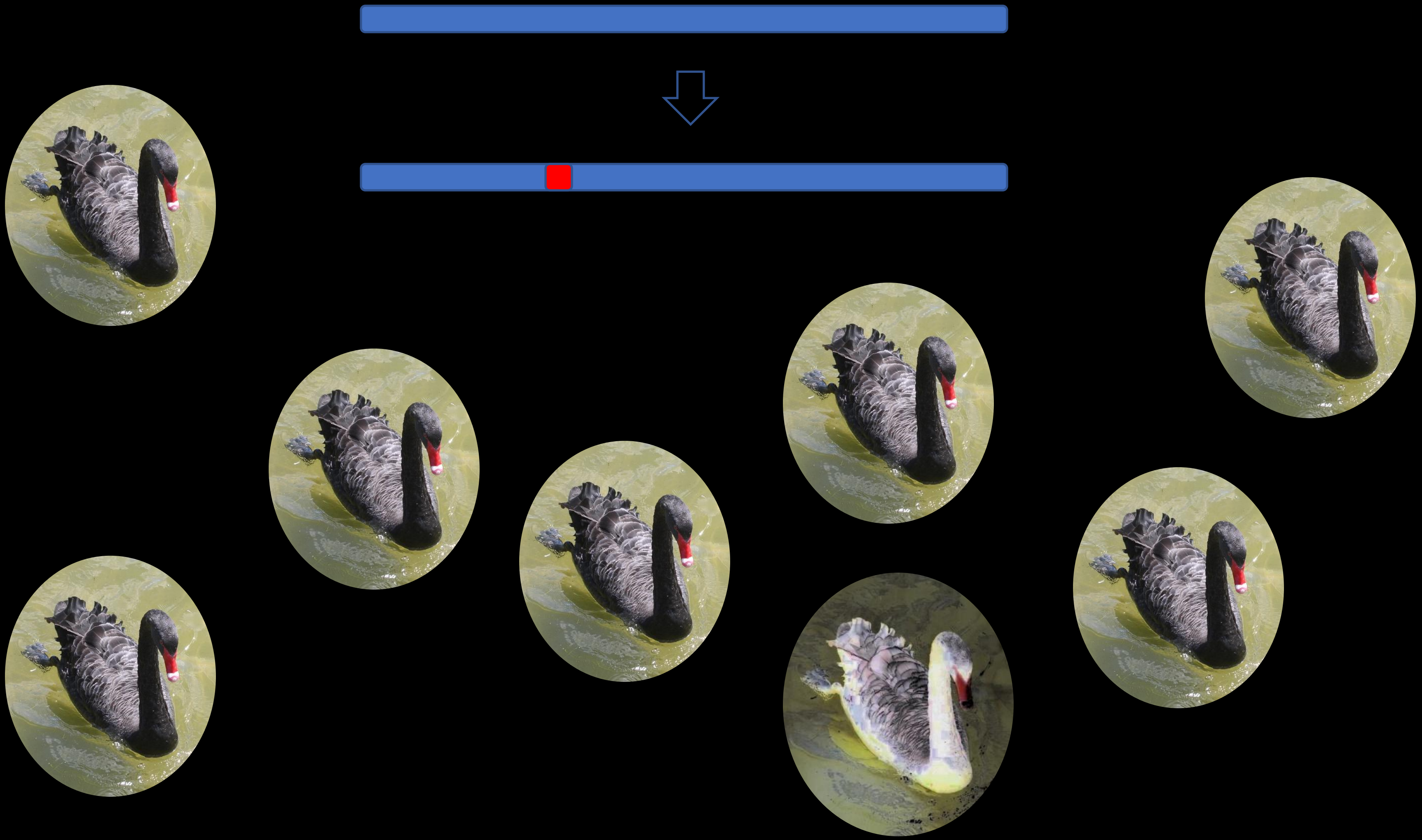
  AAA ACC **GAA TC**A CCG GAT → AAA ACC **GAA TCG AAT C**AC CGG AT.

(this is why we need sequence alignment – coming in Module 2)

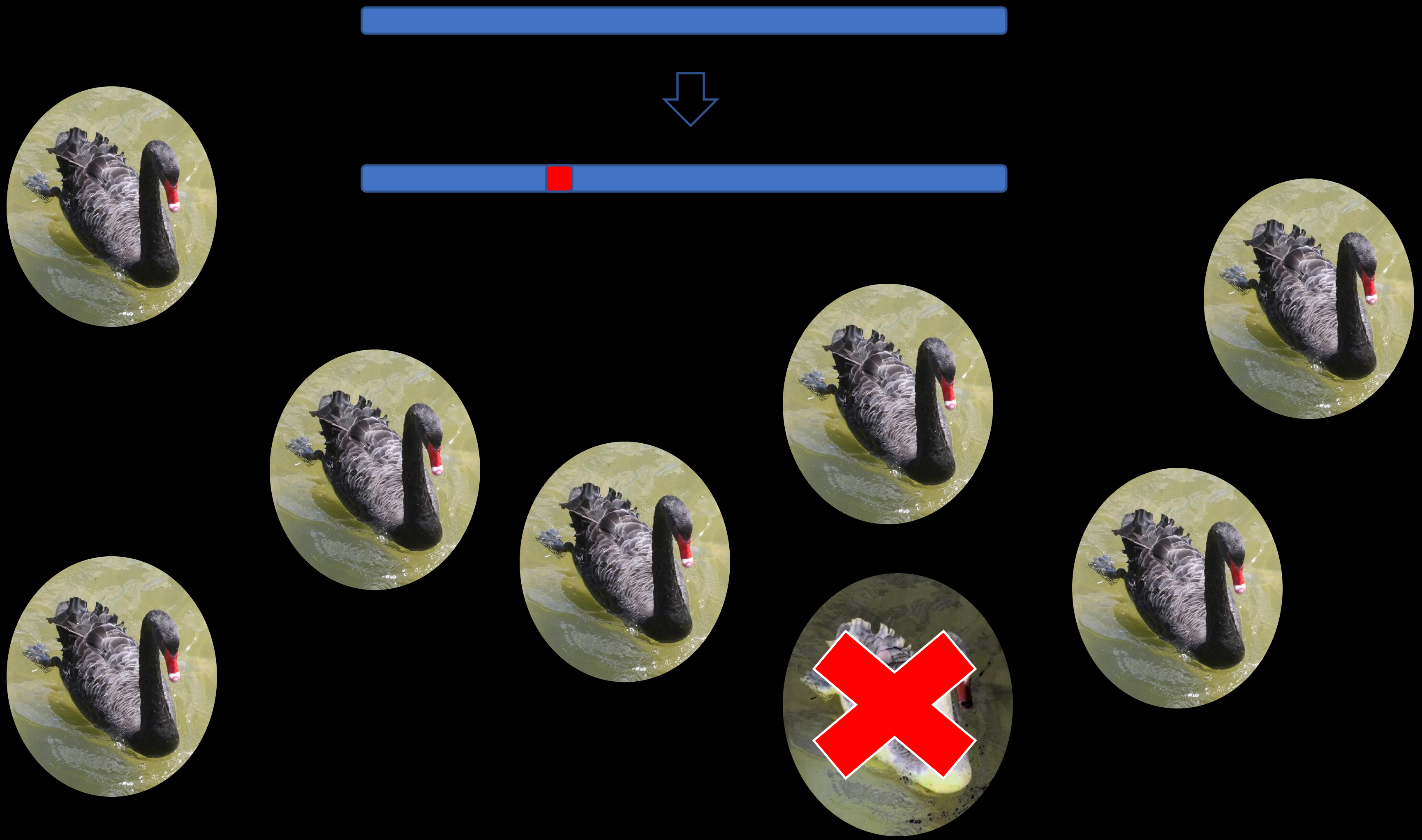
Larger scale structural variations



The population context of a mutation

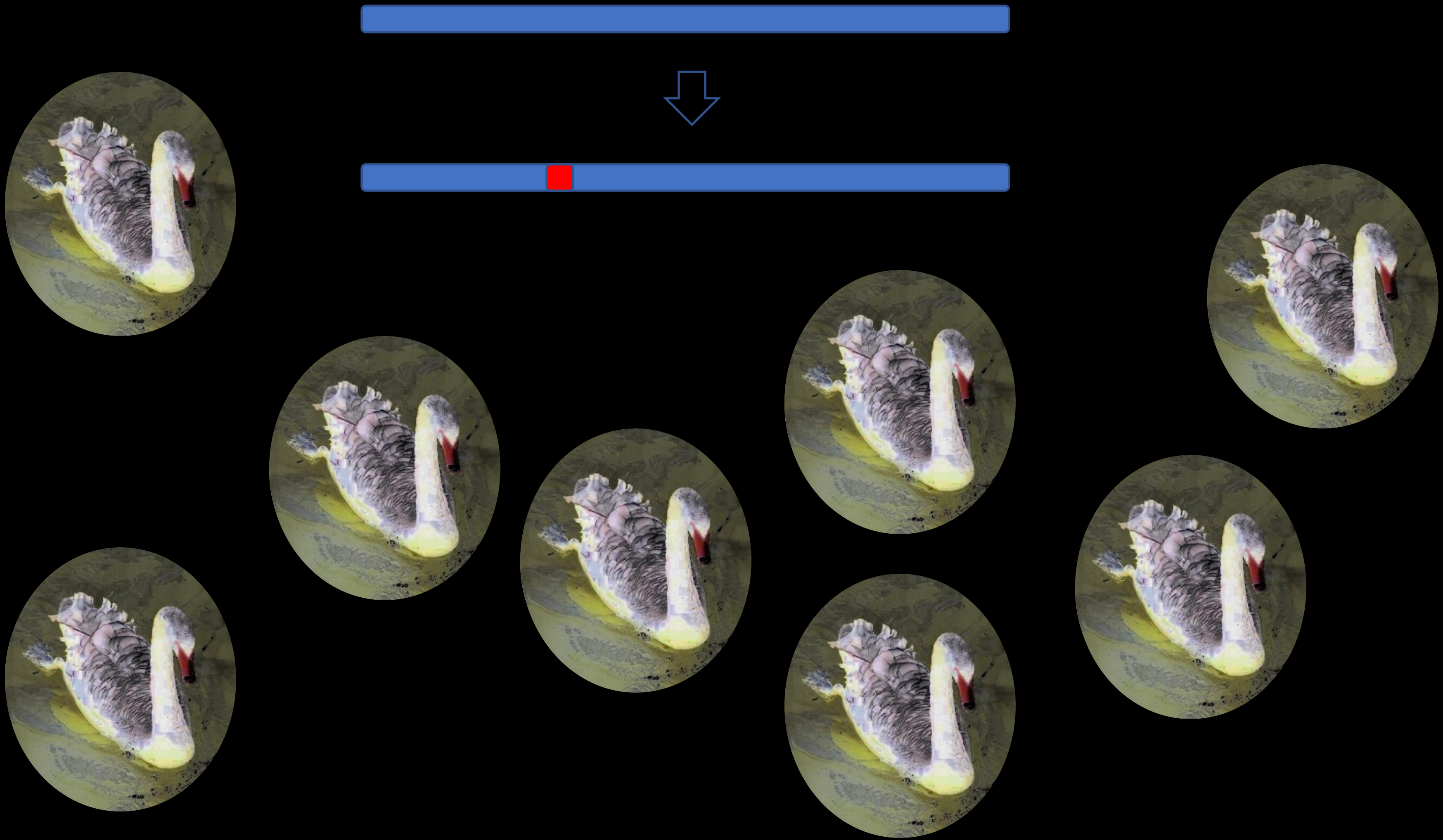


The population context of a mutation



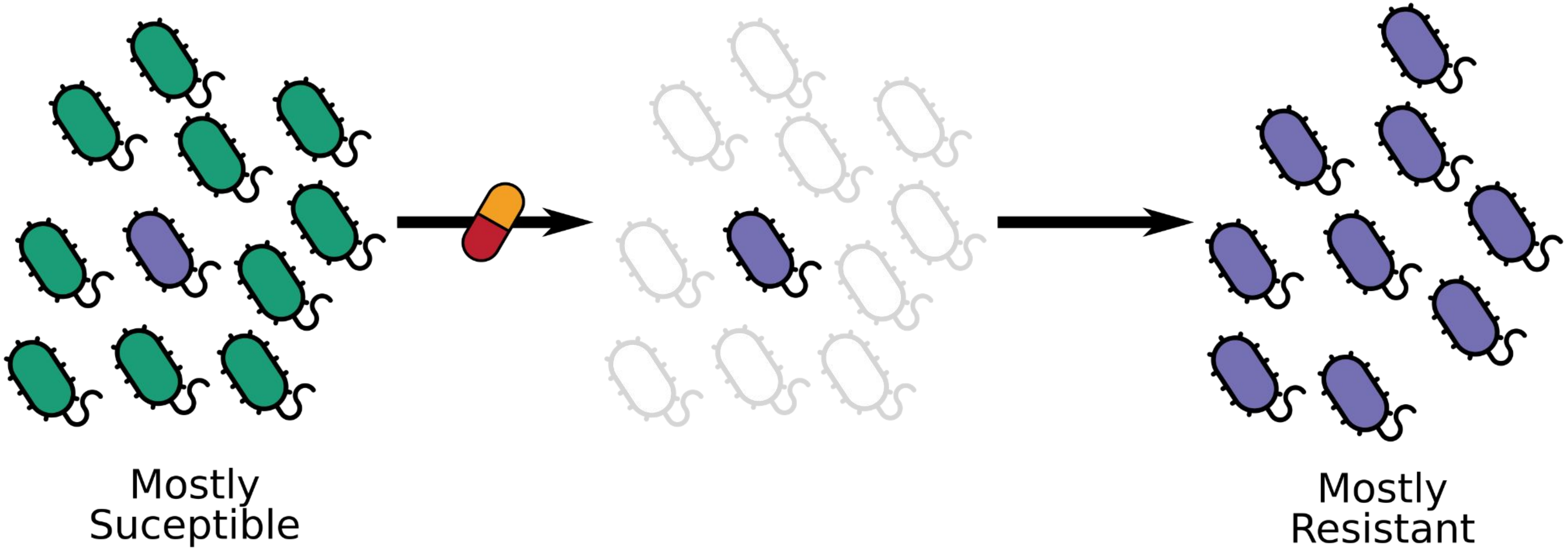
DELETERIOUS

The population context of a mutation

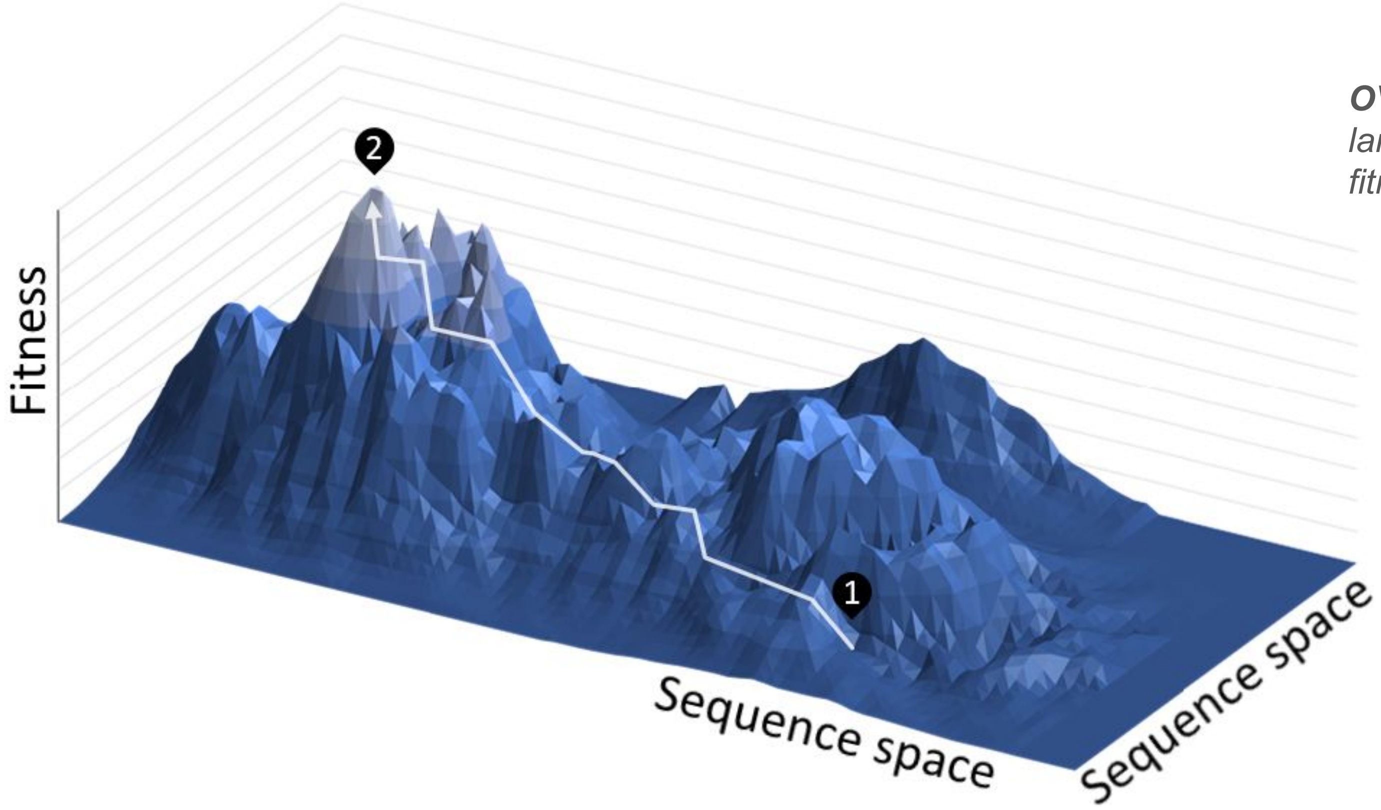


ADVANTAGEOUS

Strong selective force rapidly changes genotype



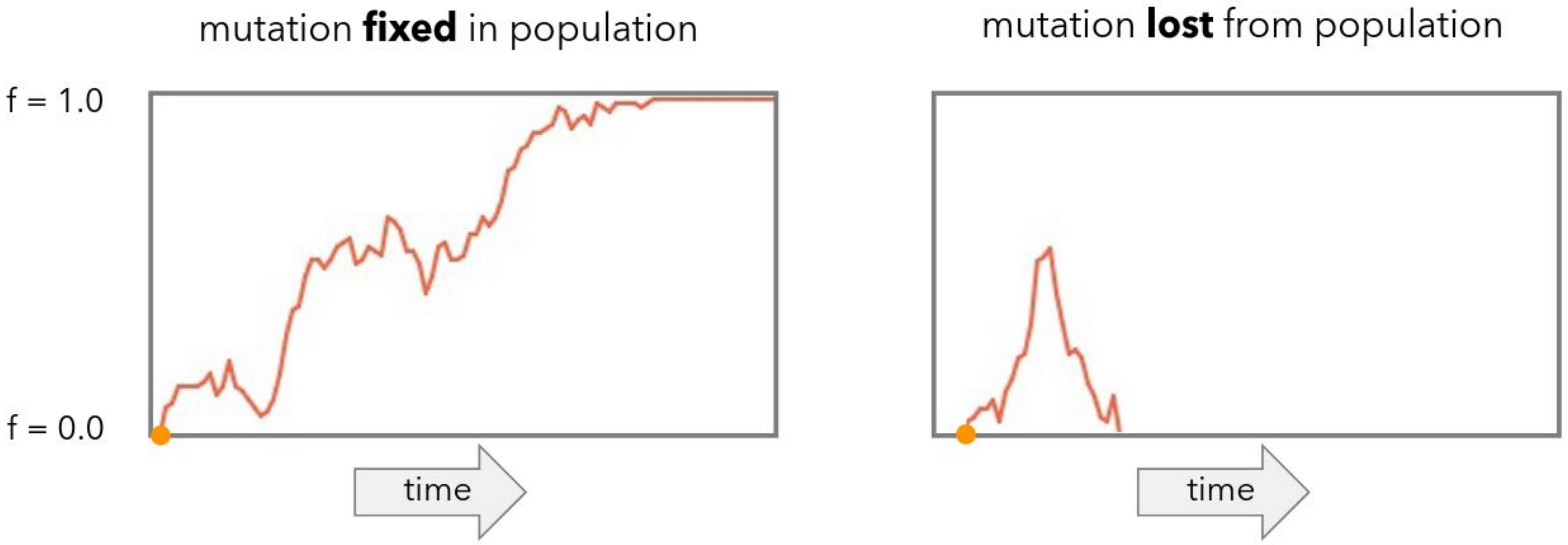
Mutations as a random walk across a fitness landscape



OVERSIMPLIFICATION - *actual landscapes are dynamic/changing & fitness is hard to measure*

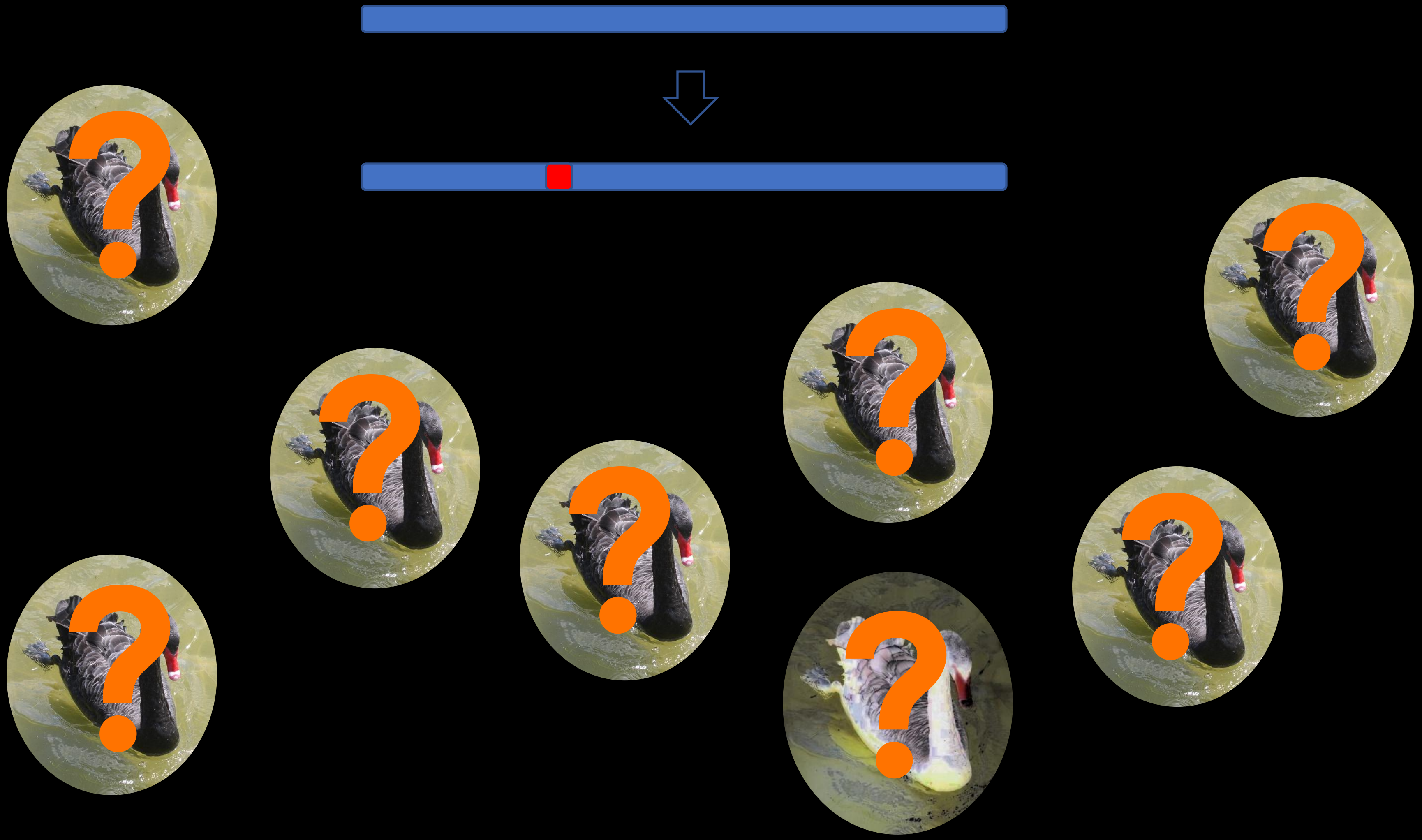
- *Fitness = quantitative representation of individual reproductive success*
- *High Fitness = more descendants (larger proportion of circulating population)*
- *Low Fitness = fewer descendants (higher chance of dying out)*

Plotting proportion of population with a specific mutation

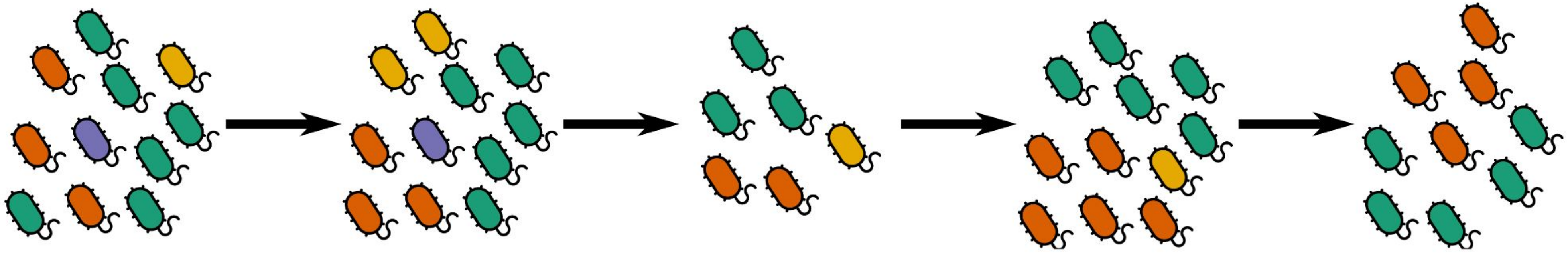


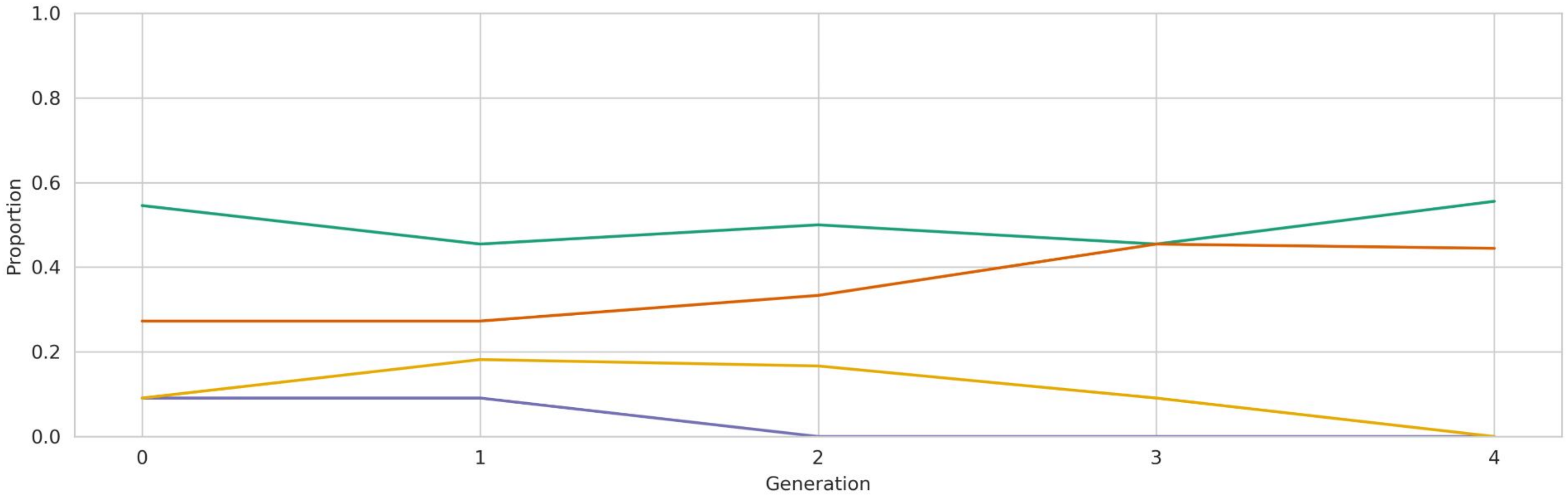
https://awarnach.mathstat.dal.ca/~joeb/PAML_lab/slides/Bielawski_lecture_PART_1.pdf

The population context of a mutation

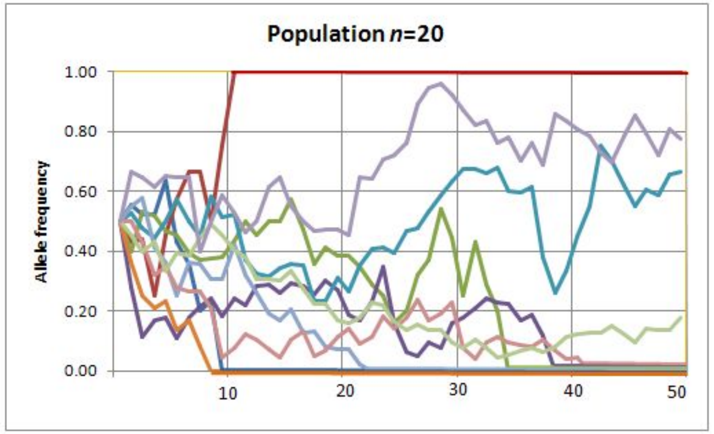


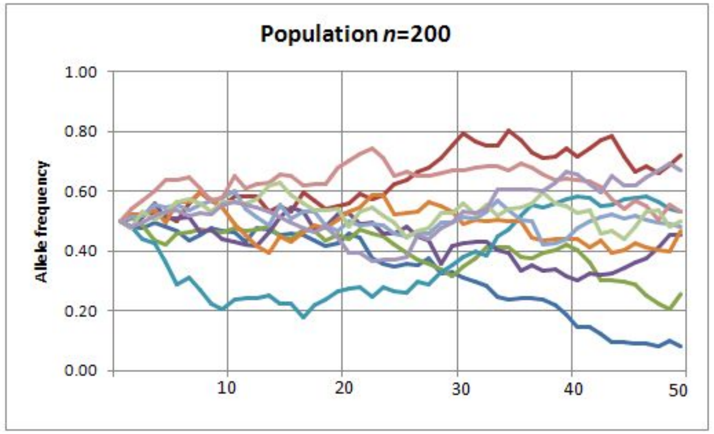
Genotypes can also change due to random sampling: “drift”

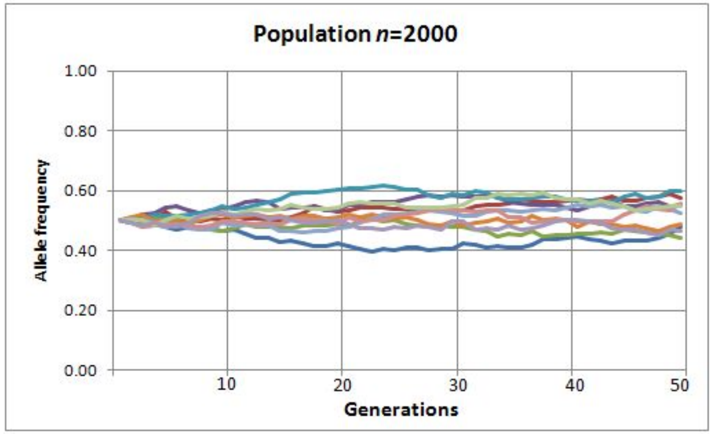


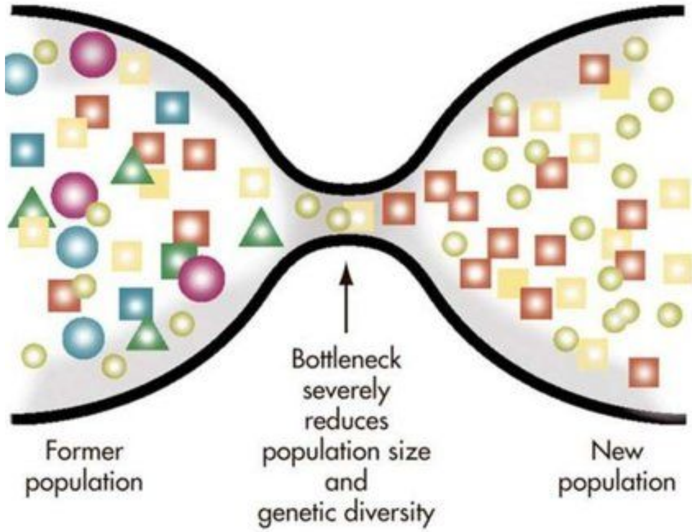


Smaller populations are more impacted by drift


Population *n*=20

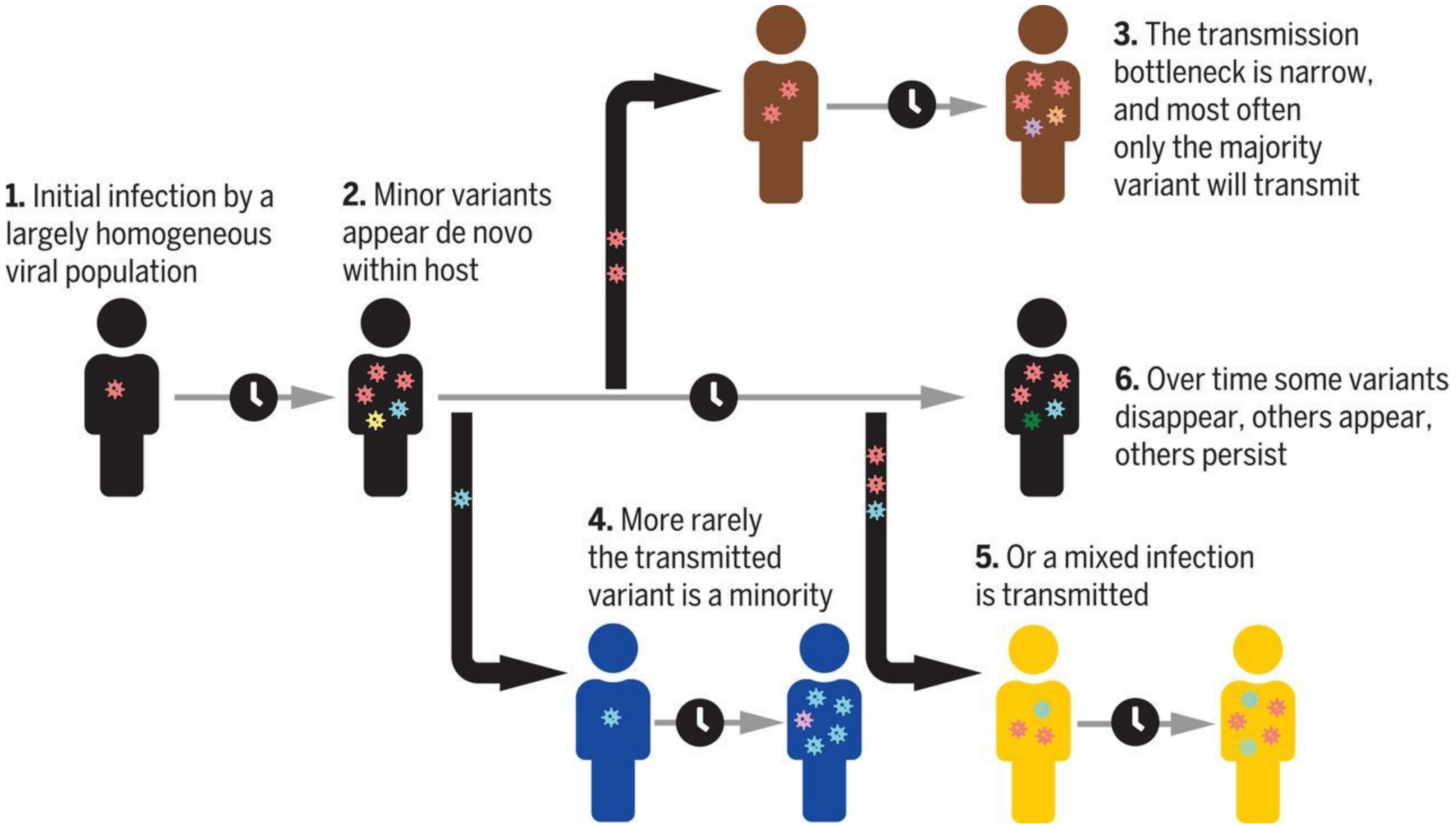

Population *n*=200


Population *n*=2000



Transient reductions in population size also drive drift: “bottlenecks”

Complicated sampling of a (within-host) population of a (between host) population



https://www.science.org/doi/10.1126/science.abg0821

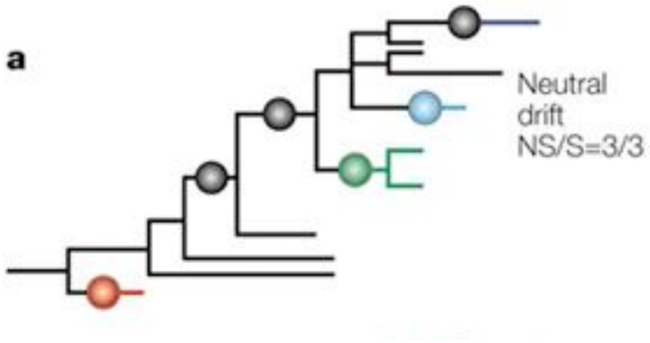
How do we estimate how much of our change is drift and how much is selection?

dN/dS is one way to detect selection

dN = non-synonymous mutations (normalised)

dS = synonymous mutations (normalised)

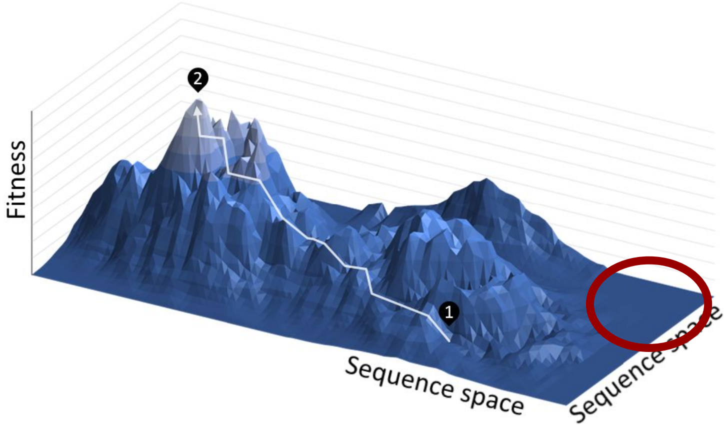
dN/dS is one way to detect selection



dN = non-synonymous mutations (normalised)

dS = synonymous mutations (normalised)

dN/dS ~ 1 : drift/neutral selection



Bush, R. Predicting adaptive evolution. *Nat Rev Genet* **2**, 387–392 (2001).
https://doi.org/10.1038/35072023

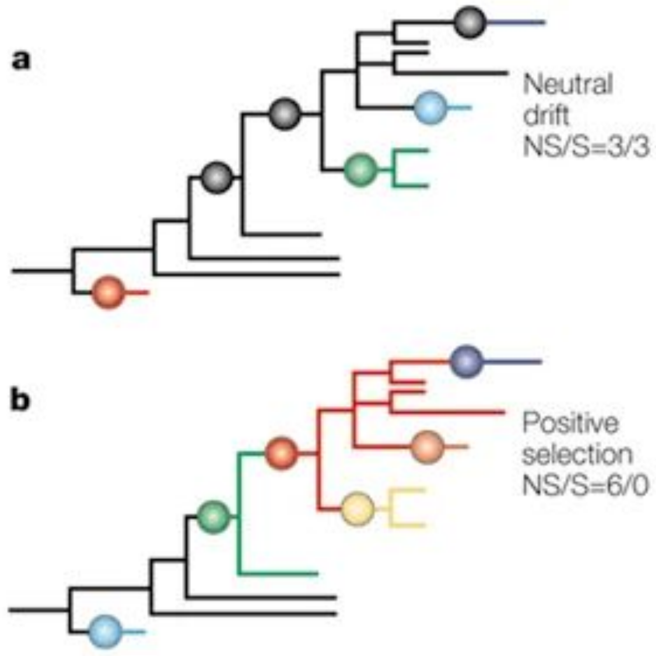
dN/dS is one way to detect selection

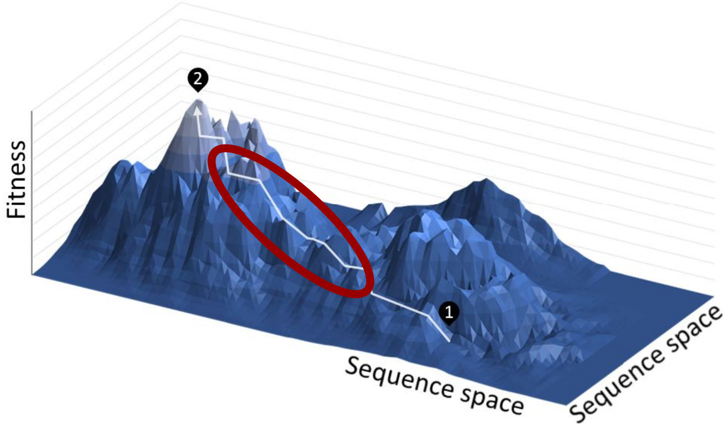
dN = non-synonymous mutations (normalised)

dS = synonymous mutations (normalised)

dN/dS > 1 : adaptive/positive selection

dN/dS ~ 1 : drift/neutral selection





Bush, R. Predicting adaptive evolution. *Nat Rev Genet* **2**, 387–392 (2001). https://doi.org/10.1038/35072023

dN/dS is one way to detect selection

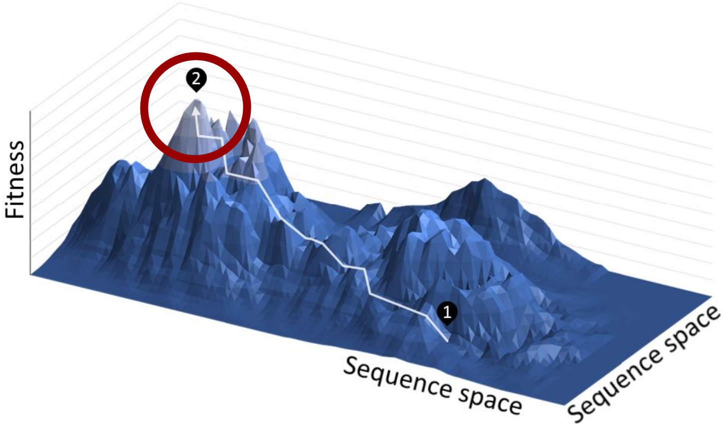
dN = non-synonymous mutations (normalised)

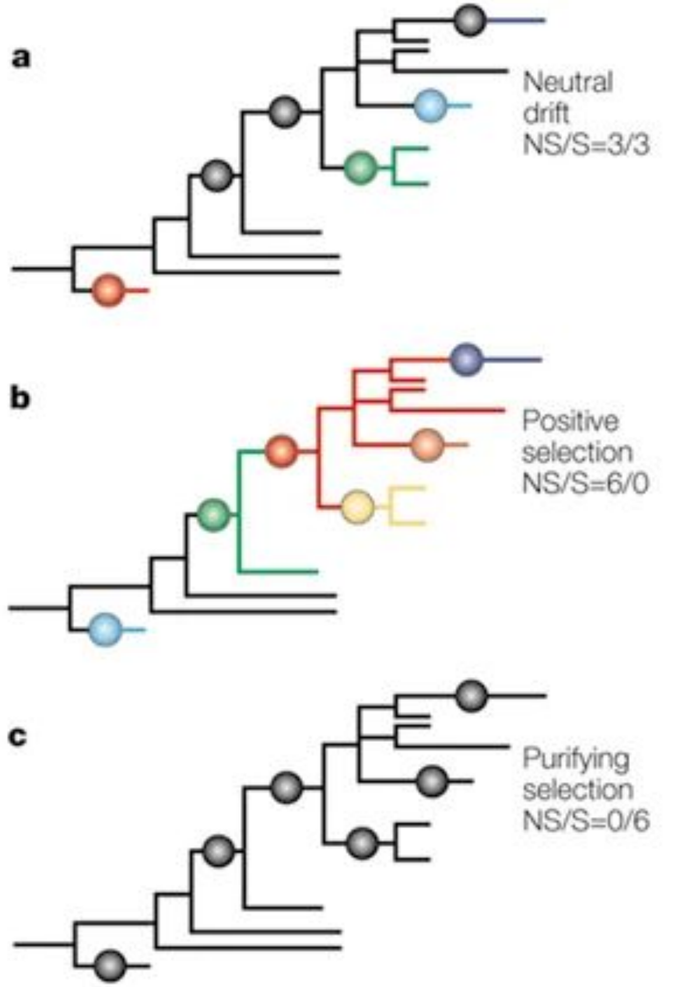
dS = synonymous mutations (normalised)

dN/dS > 1 : adaptive/positive selection

dN/dS ~ 1 : drift/neutral selection

dN/dS < 1: purifying/negative selection



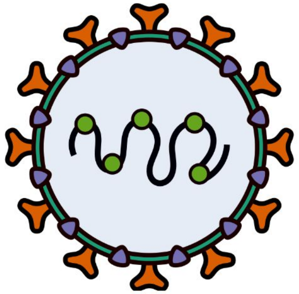


Nature Reviews | Genetics

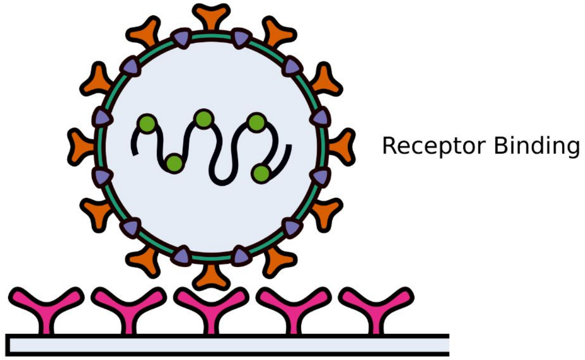
Bush, R. Predicting adaptive evolution. *Nat Rev Genet* **2**, 387–392 (2001). https://doi.org/10.1038/35072023

Using these approaches to investigate if shortening treatment causes problems of antiviral resistance

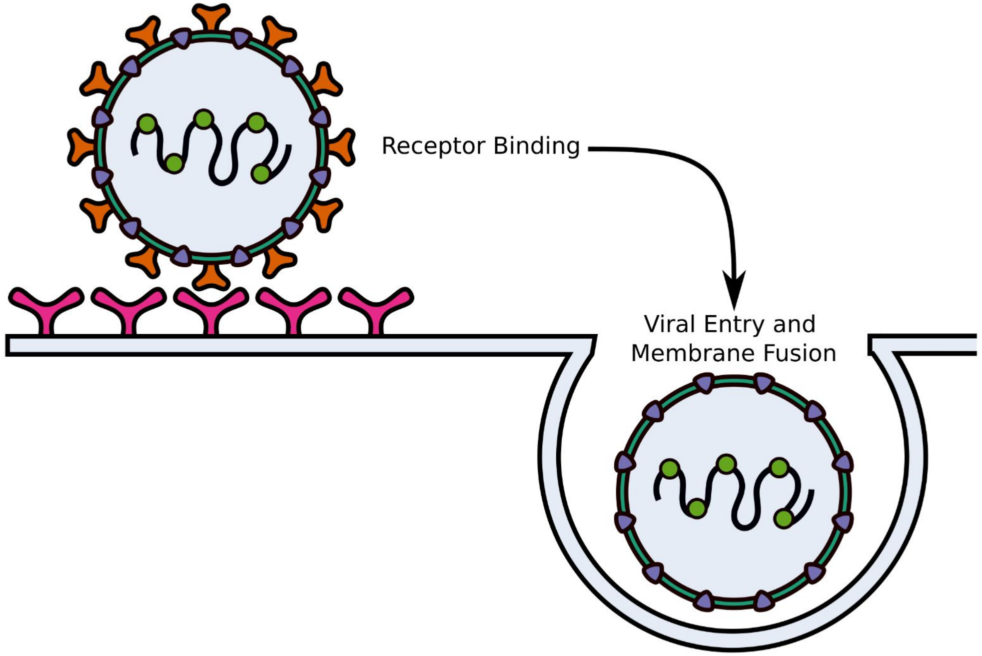
Remdesivir is an important antiviral



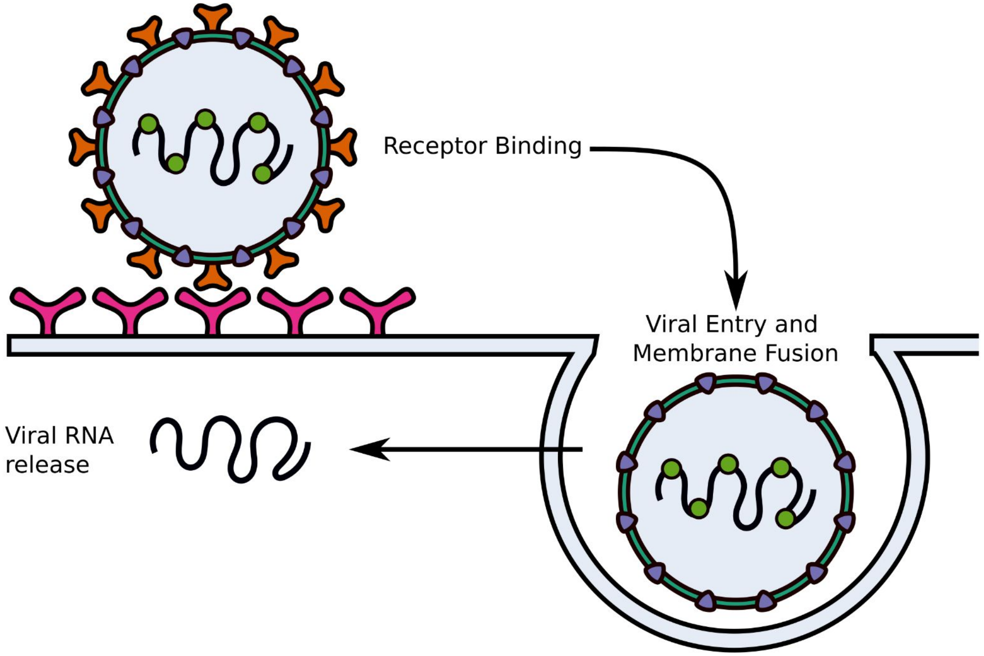
Remdesivir is an important antiviral



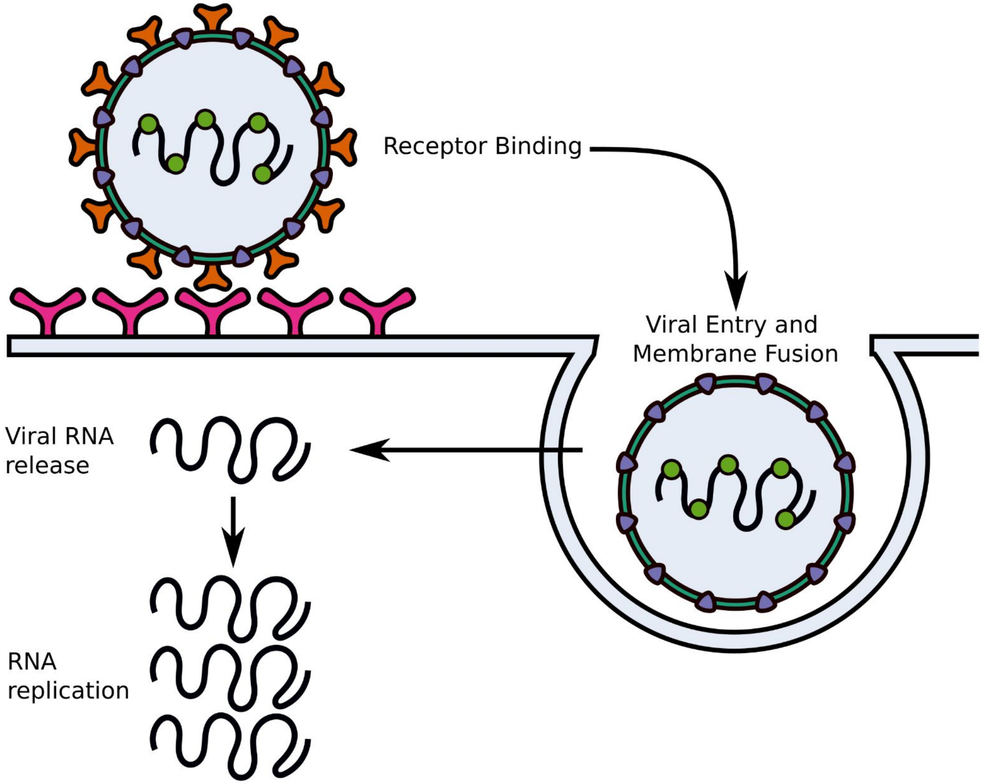
Remdesivir is an important antiviral



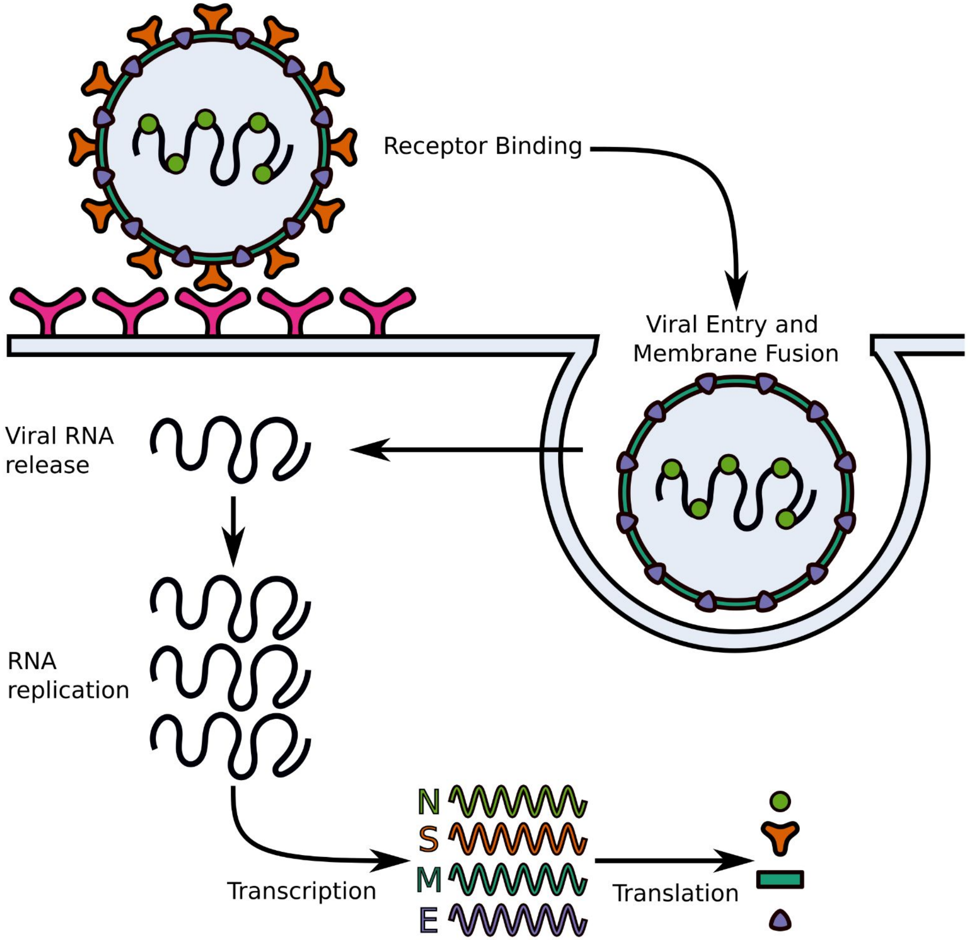
Remdesivir is an important antiviral



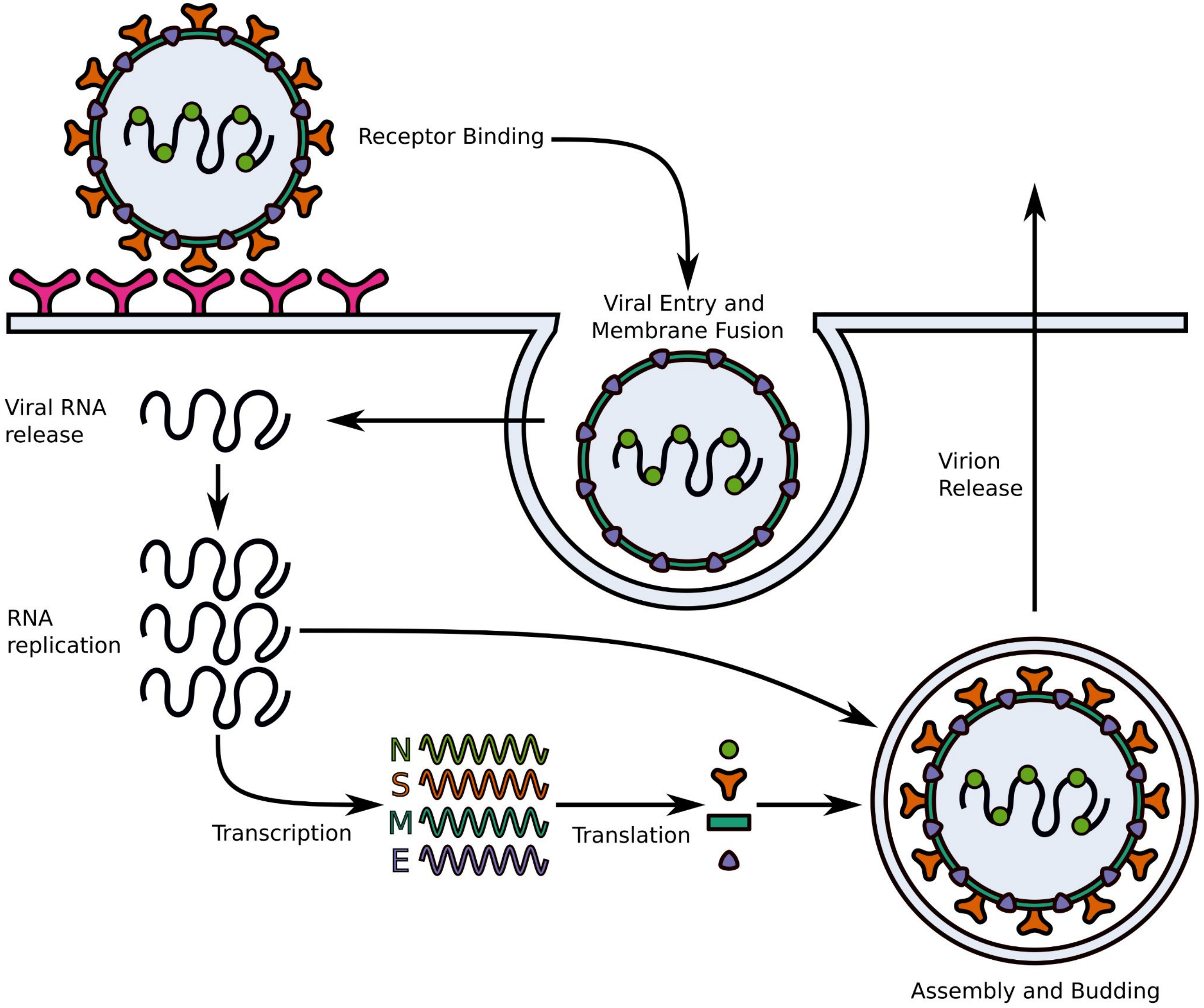
Remdesivir is an important antiviral



Remdesivir is an important antiviral

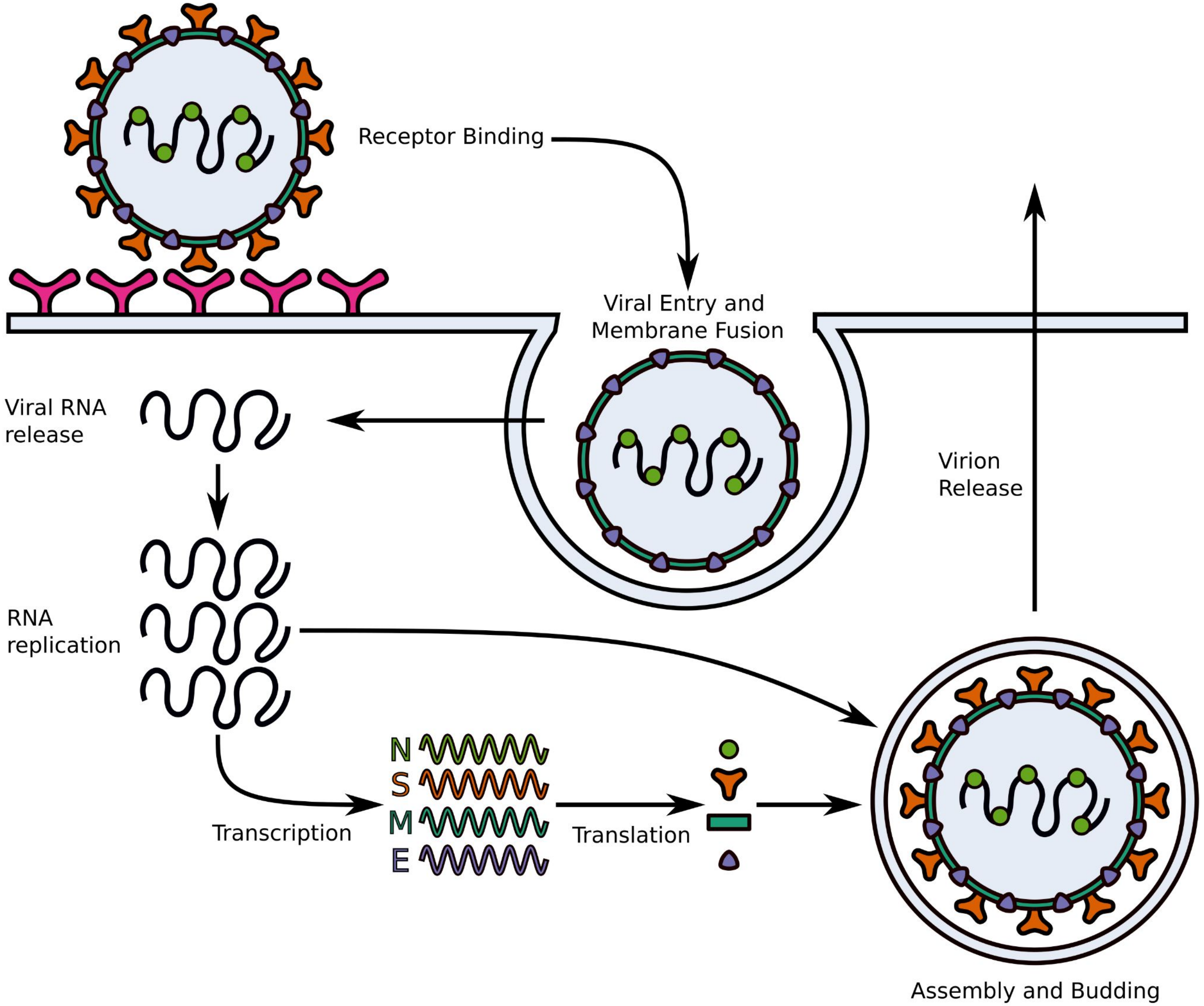


Remdesivir is an important antiviral



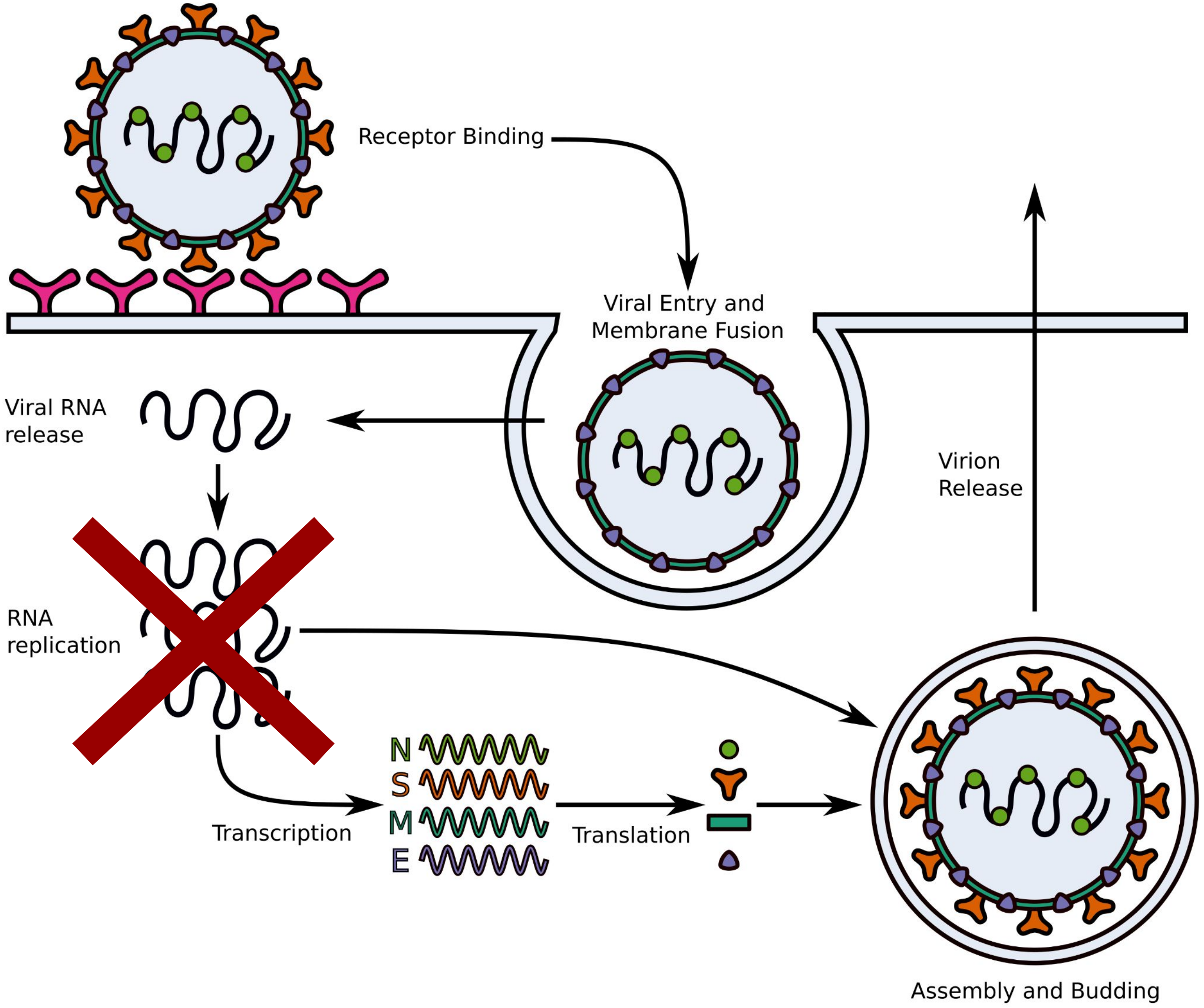
Remdesivir is an important antiviral

Remdesivir

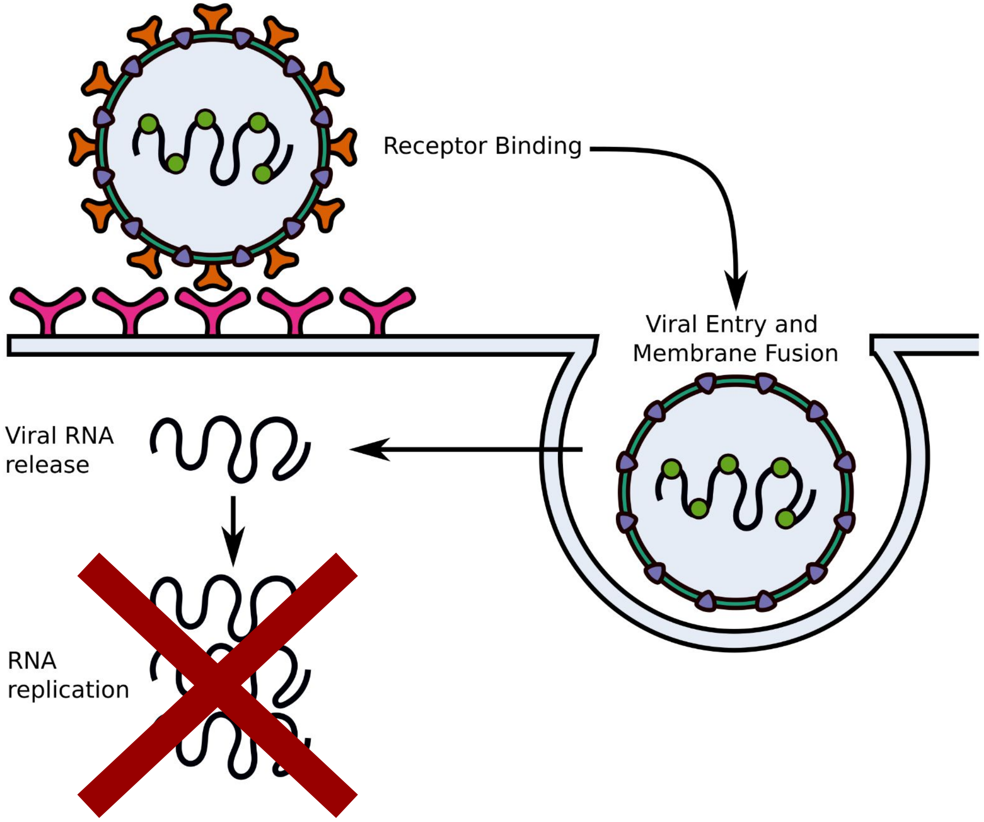


Remdesivir is an important antiviral

Remdesivir



Remdesivir is an important antiviral



Receptor Binding

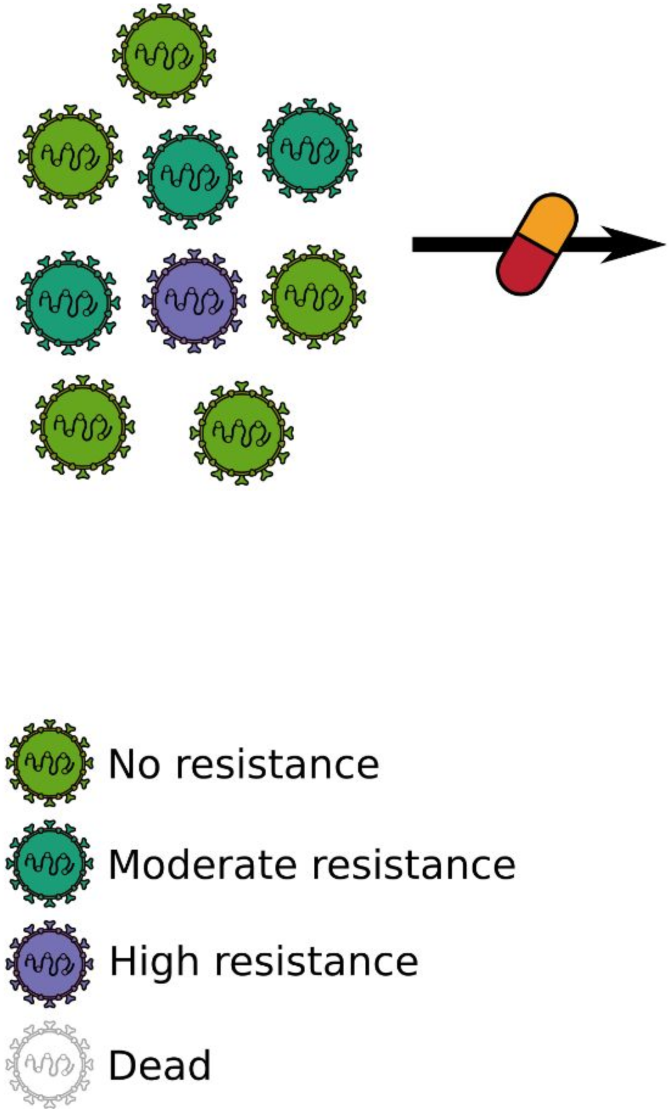
Viral Entry and
Membrane Fusion

Viral RNA
release

RNA
replication

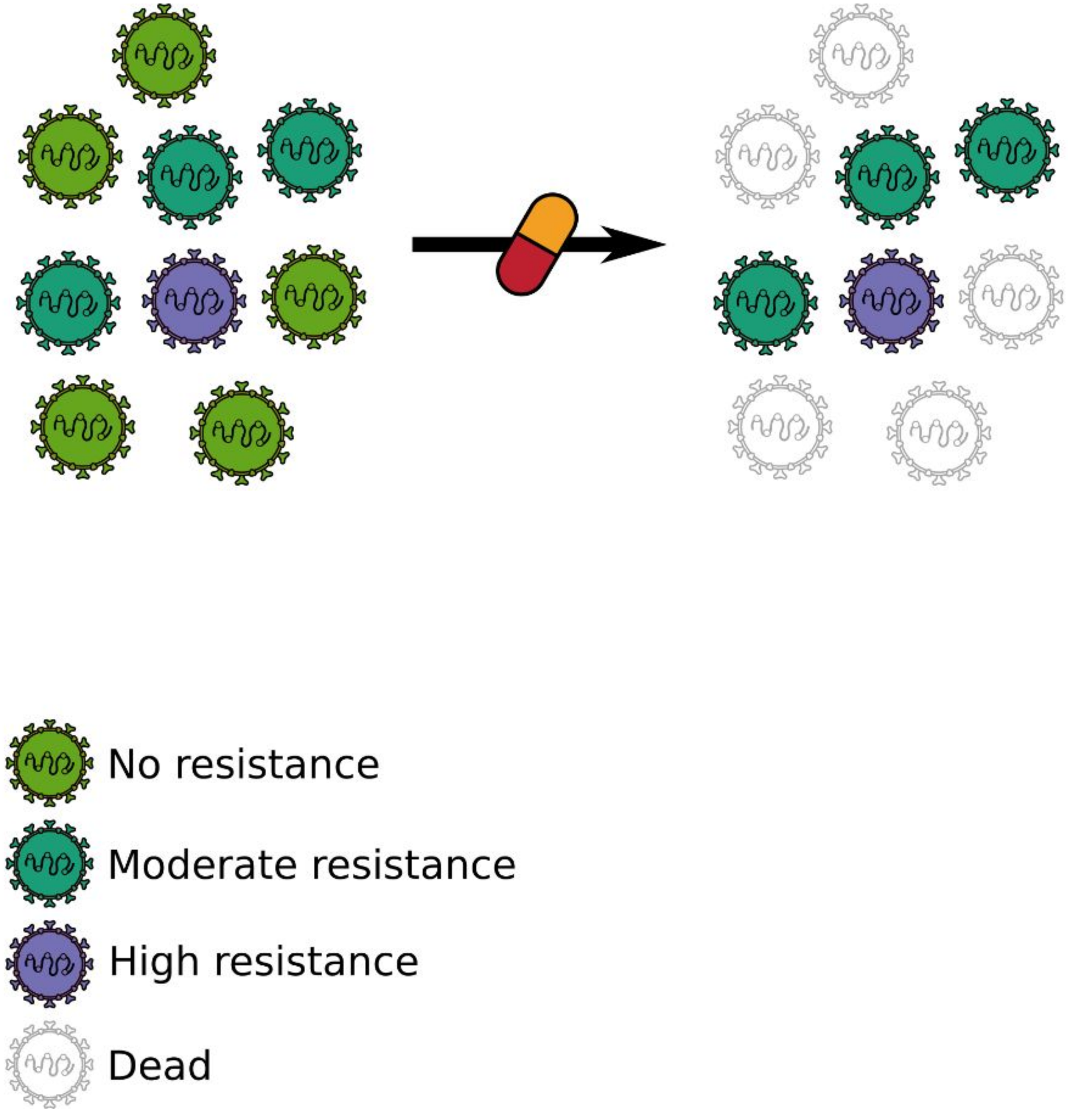
Remdesivir

Shortened treatment can promote resistance



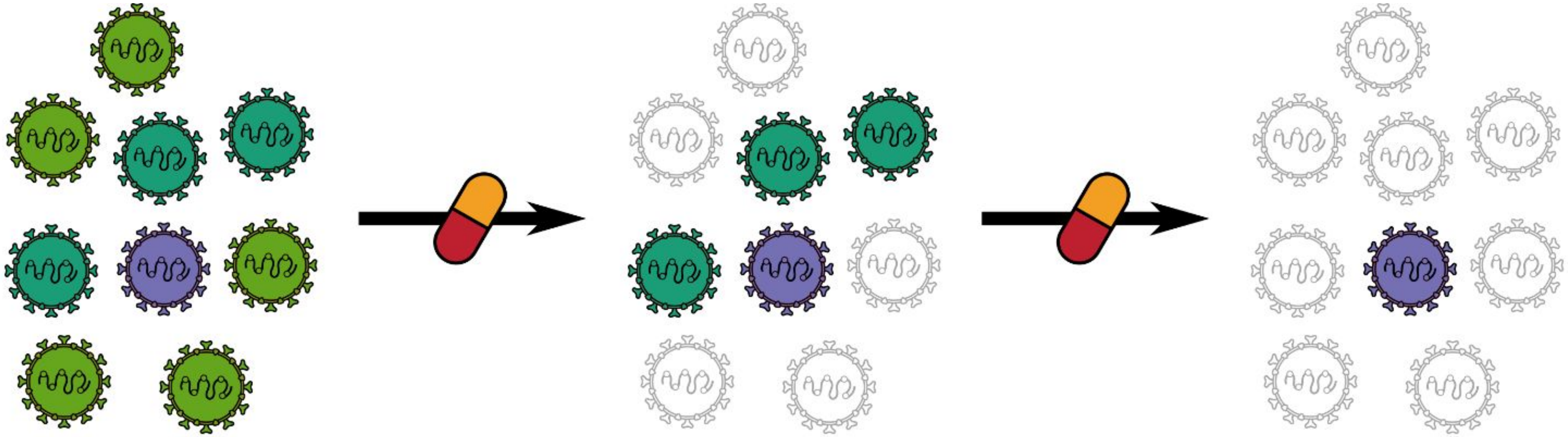
Caveat: lots of other dynamics/forces contribute to emergence of resistance (e.g., relative fitness, infection dynamics, use-patterns/contact demographics)

Shortened treatment can promote resistance



Caveat: lots of other dynamics/forces contribute to emergence of resistance (e.g., relative fitness, infection dynamics, use-patterns/contact demographics)

Shortened treatment can promote resistance



🟢 No resistance

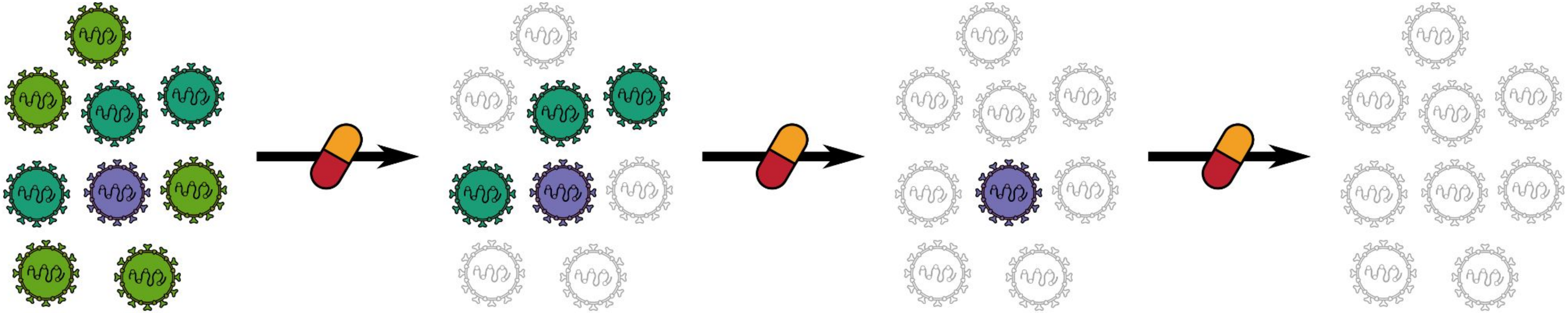
🟩 Moderate resistance

🟣 High resistance

⚪ Dead

Caveat: lots of other dynamics/forces contribute to emergence of resistance (e.g., relative fitness, infection dynamics, use-patterns/contact demographics)

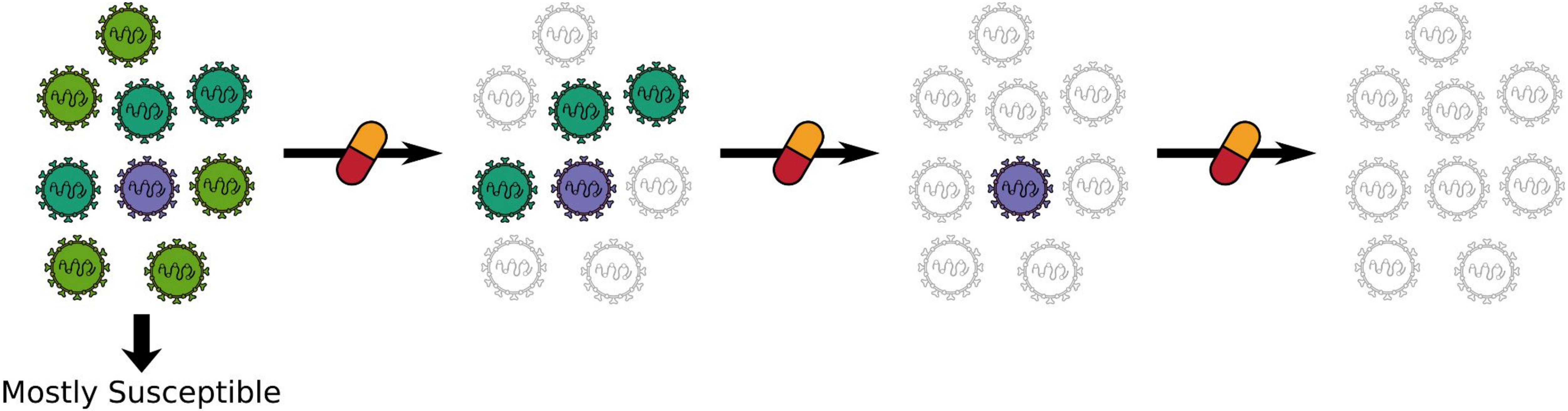
Shortened treatment can promote resistance



- 🟩 No resistance
- 🟩 Moderate resistance
- 🟪 High resistance
- ⚪ Dead

Caveat: lots of other dynamics/forces contribute to emergence of resistance (e.g., relative fitness, infection dynamics, use-patterns/contact demographics)

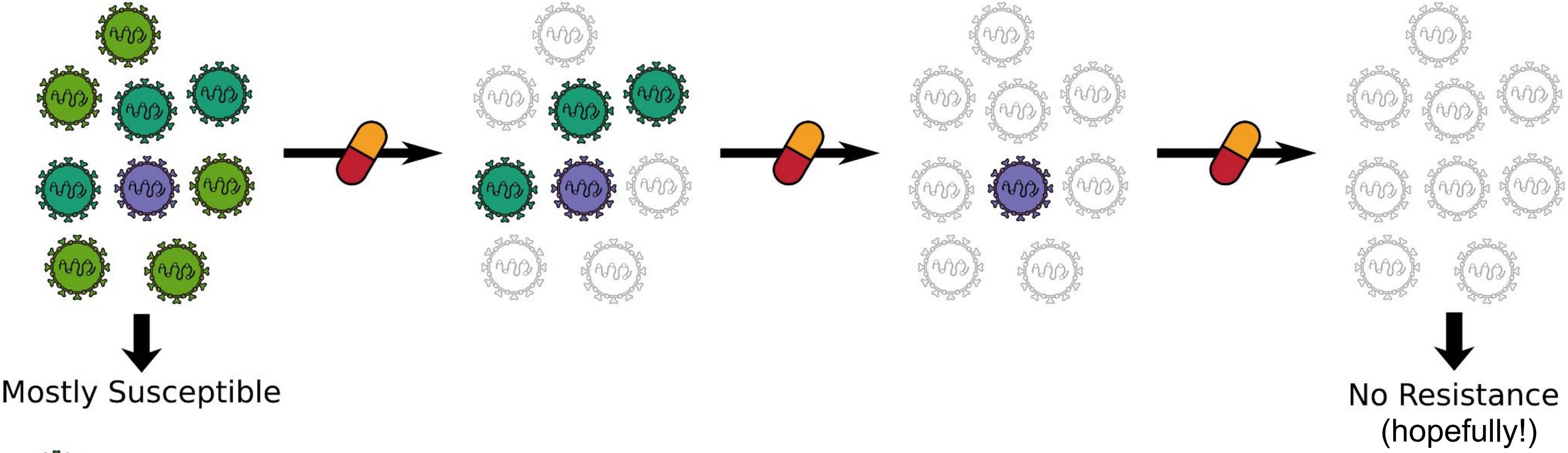
Shortened treatment can promote resistance



- 🟢 No resistance
- 🟢 Moderate resistance
- 🟣 High resistance
- ○ Dead

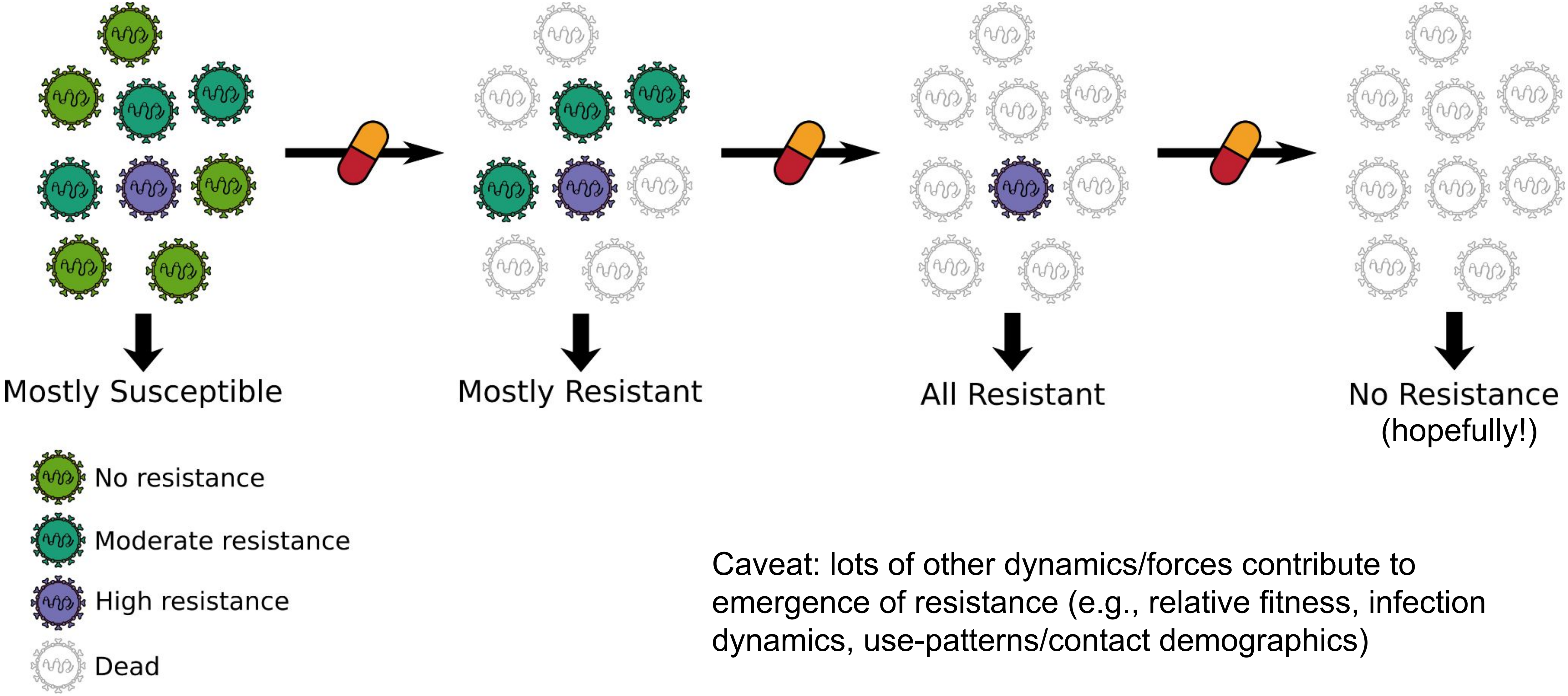
Caveat: lots of other dynamics/forces contribute to emergence of resistance (e.g., relative fitness, infection dynamics, use-patterns/contact demographics)

Shortened treatment can promote resistance



Caveat: lots of other dynamics/forces contribute to emergence of resistance (e.g., relative fitness, infection dynamics, use-patterns/contact demographics)

Shortened treatment can promote resistance



Mostly Susceptible

Mostly Resistant

All Resistant

No Resistance
(hopefully!)

No resistance

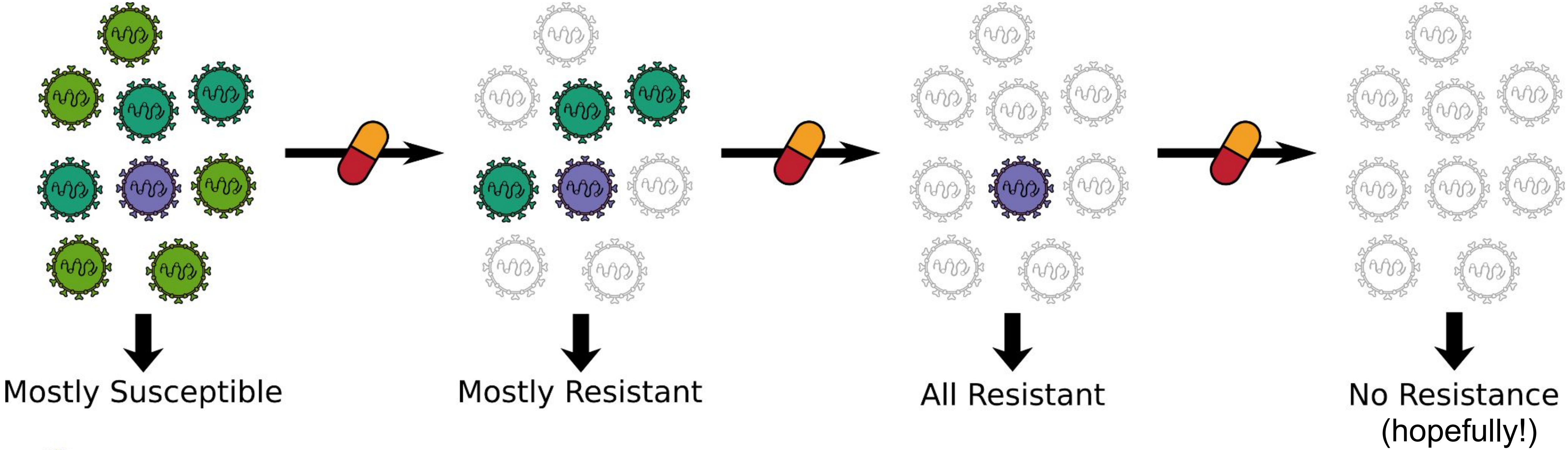
Moderate resistance

High resistance

Dead

Caveat: lots of other dynamics/forces contribute to emergence of resistance (e.g., relative fitness, infection dynamics, use-patterns/contact demographics)

Shortened treatment can promote resistance



Mostly Susceptible

Mostly Resistant

All Resistant

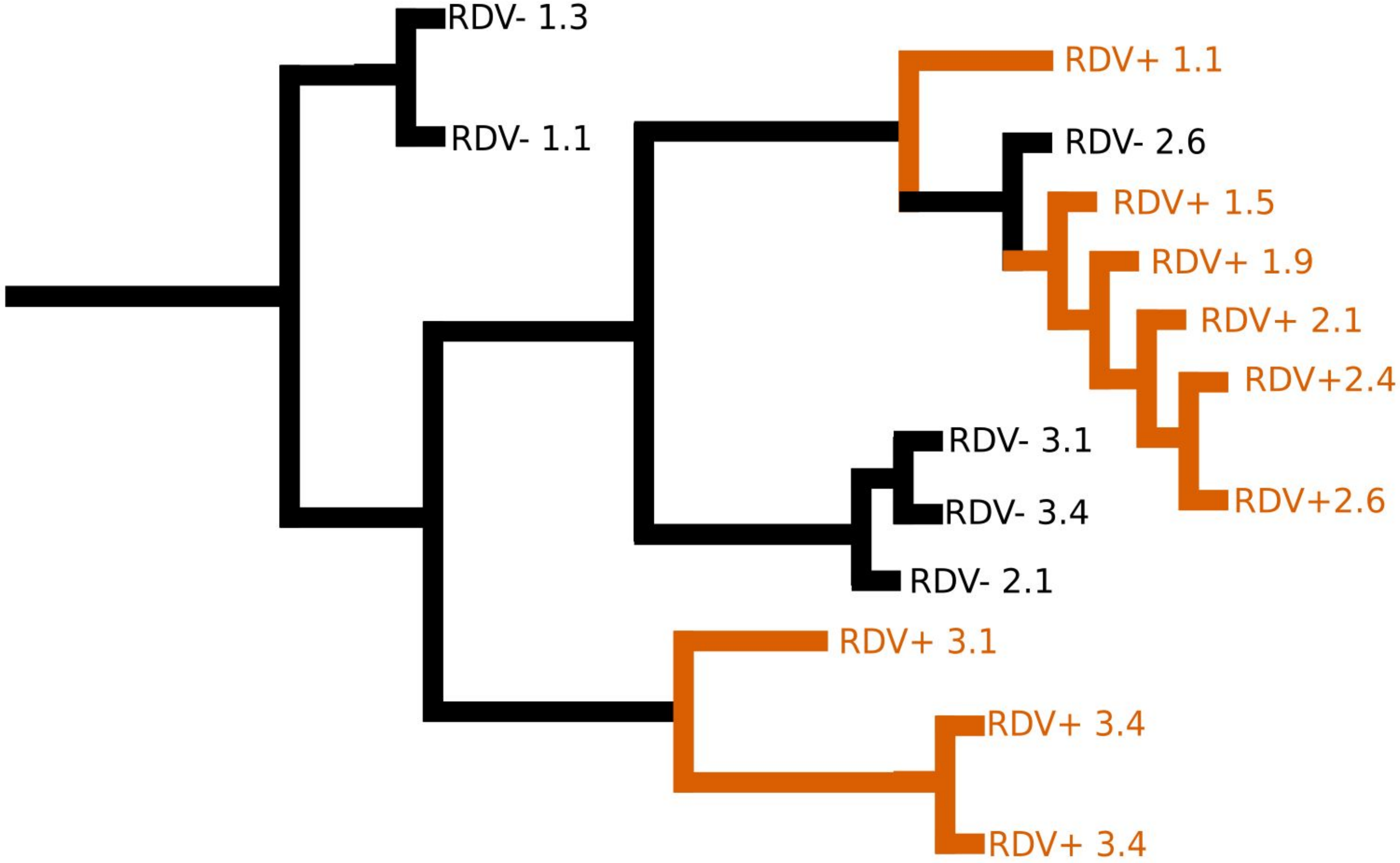
No Resistance
(hopefully!)

- No resistance
- Moderate resistance
- High resistance
- Dead

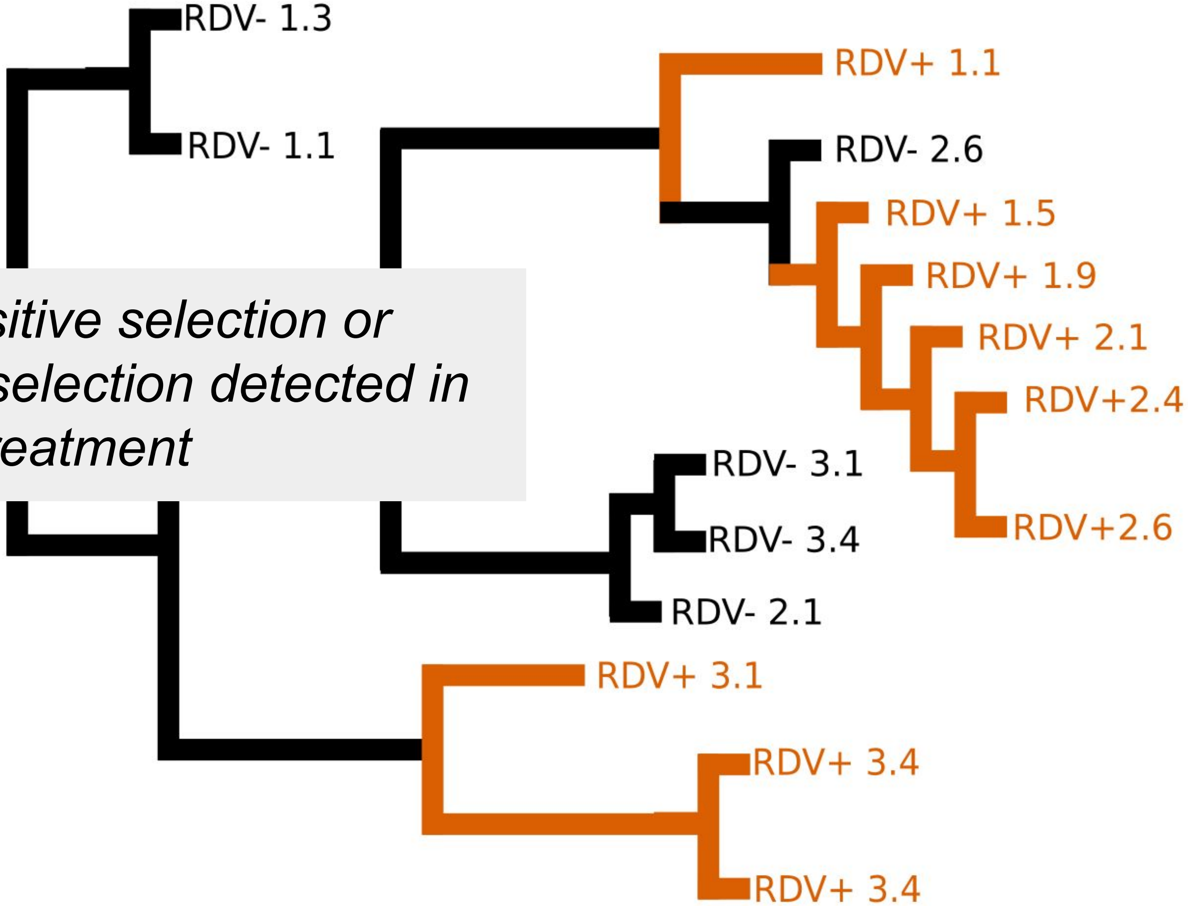
10 day course recommendation for remdesivir

Caveat: lots of other dynamics/forces contribute to emergence of resistance (e.g., relative fitness, infection dynamics, use-patterns/contact demographics)

dN/dS ratio comparison in short vs long RDV treatment

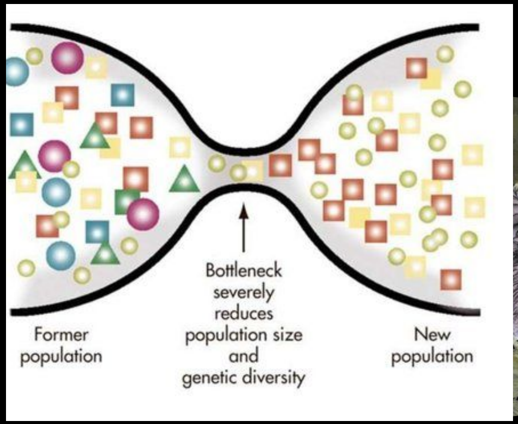


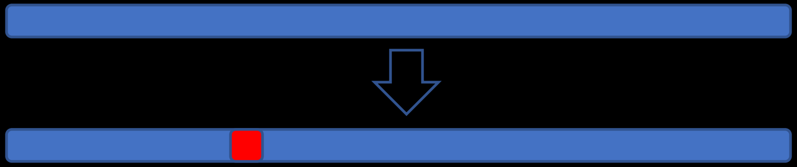
dN/dS ratio comparison in short vs long RDV treatment



No significant positive selection or intensification of selection detected in shortened RDV treatment

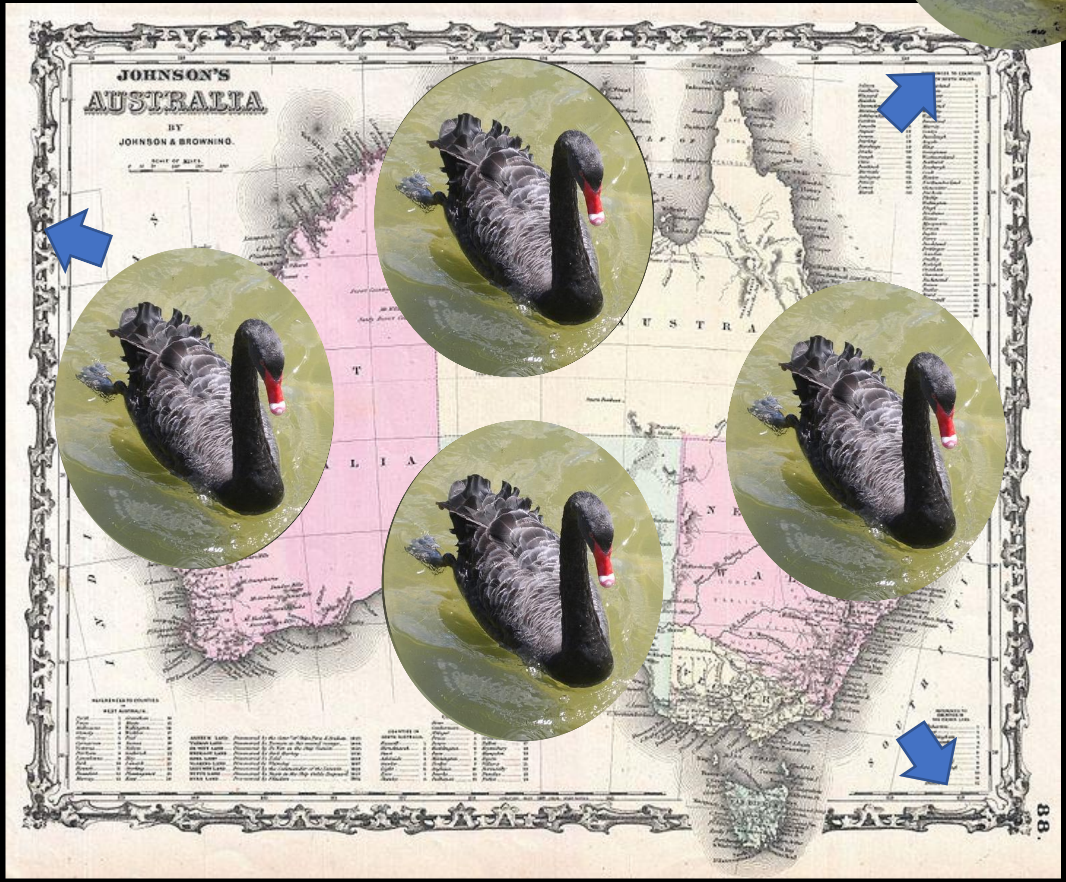
Without gene flow populations diverge



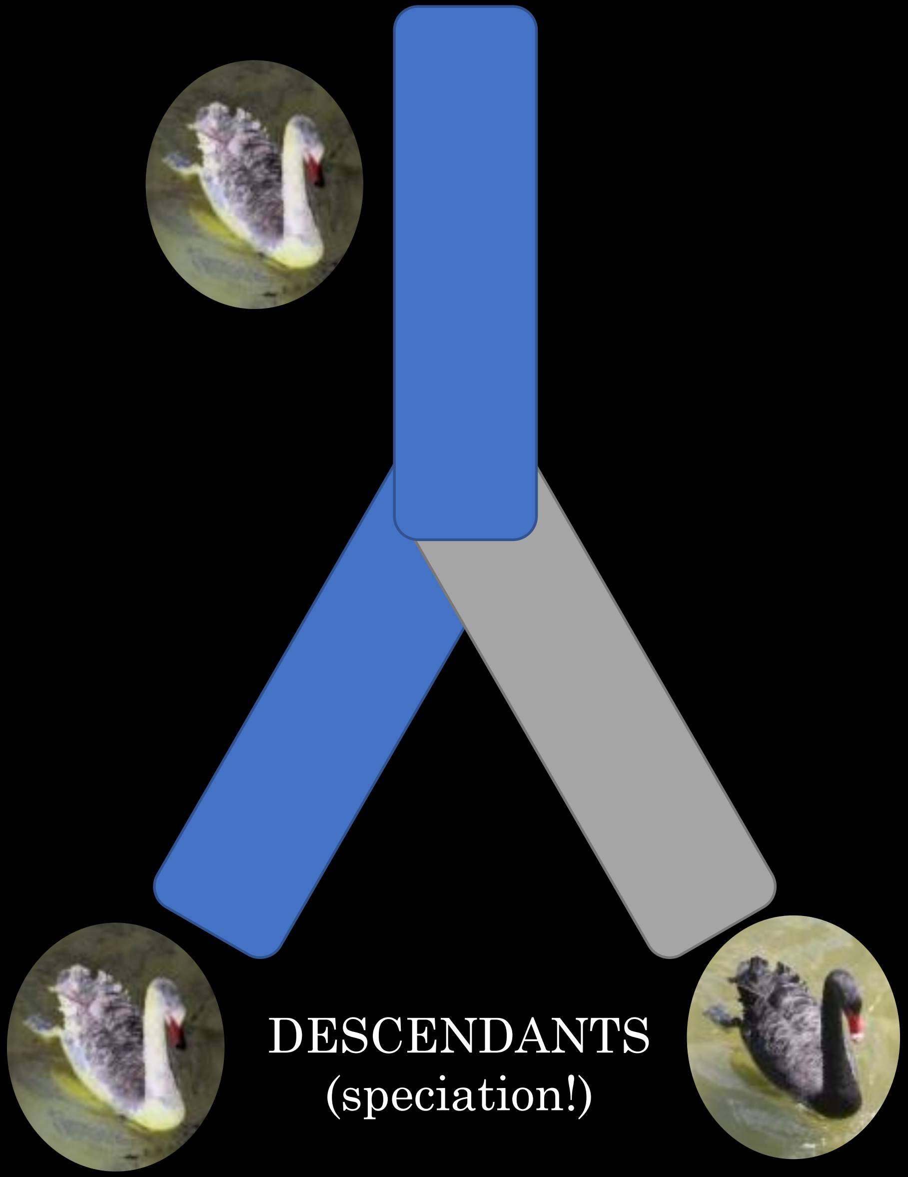


Different environment can have different selection

Migration can act as bottleneck



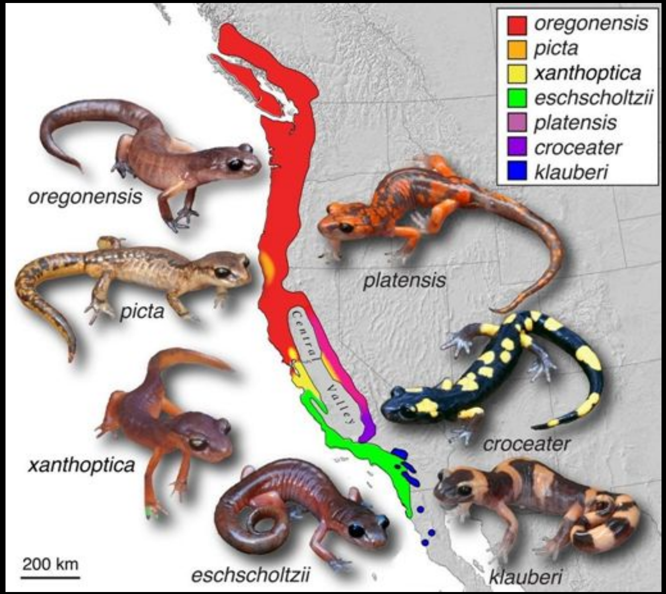
https://www.geographicus.com/P/AntiqueMap/Australia-johnson-1864



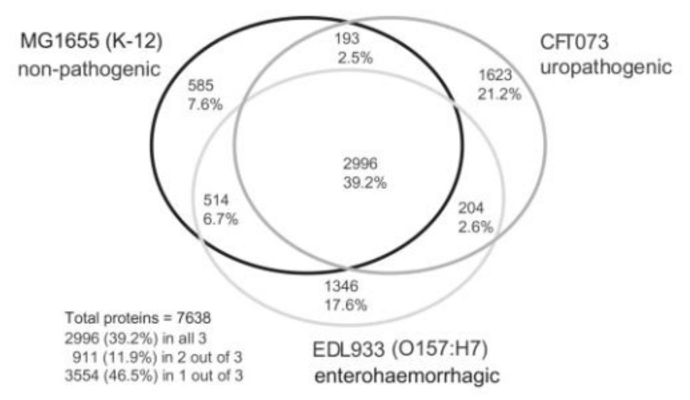
ANCESTOR

DESCENDANTS
(speciation!)

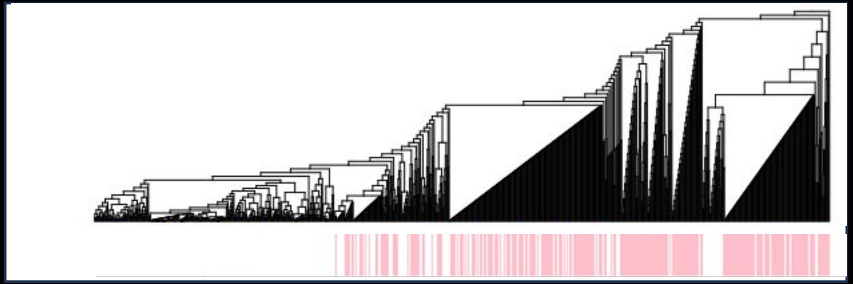
Speciation ain't pretty



Ring species: most neighbours can interbreed, more-distant populations cannot



Fewer than half of genes in these *E. coli* are found in all three of them



The erythromycin resistance gene *ErmB* is found in a scattered subset of *Enterococcus* genomes
(pink = present in genome)

Devitt et al (2011) *BMC Evolutionary Biology*
Welch et al (2002) *PNAS*
Sanderson, Gray, et al. (2022)

Relationships among sequences

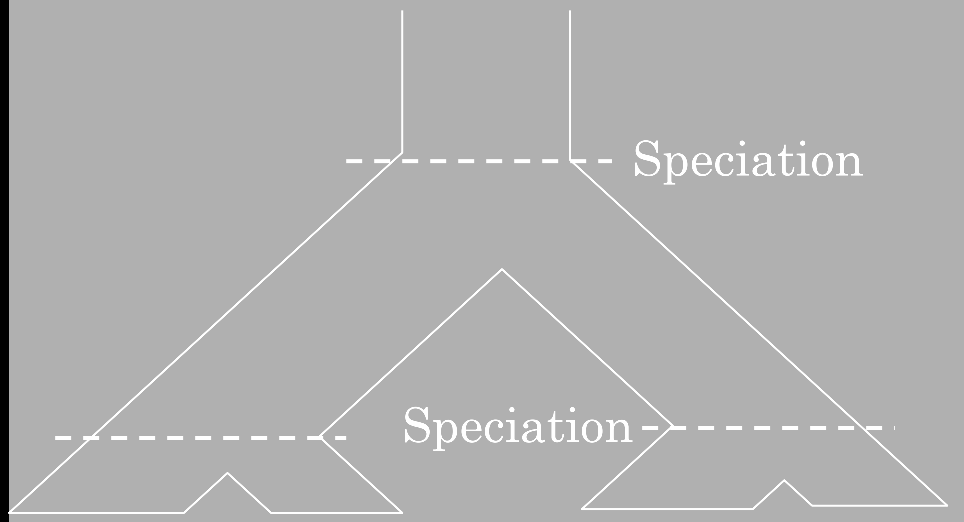
- HOMOLOGY
- ORTHOLOGY
- PARALOGY

Remember: duplications

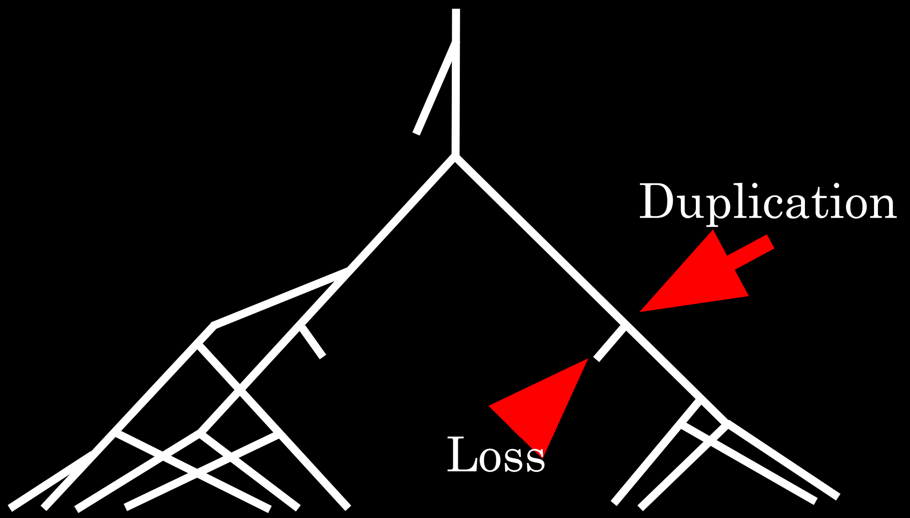
AAA ACC **GAA TC**A CCG GAT → AAA ACC **GAA TCG AAT C**A CCG GAT

Bifurcations at the gene and species level

When speciation occurs, most genes are inherited by both descendants

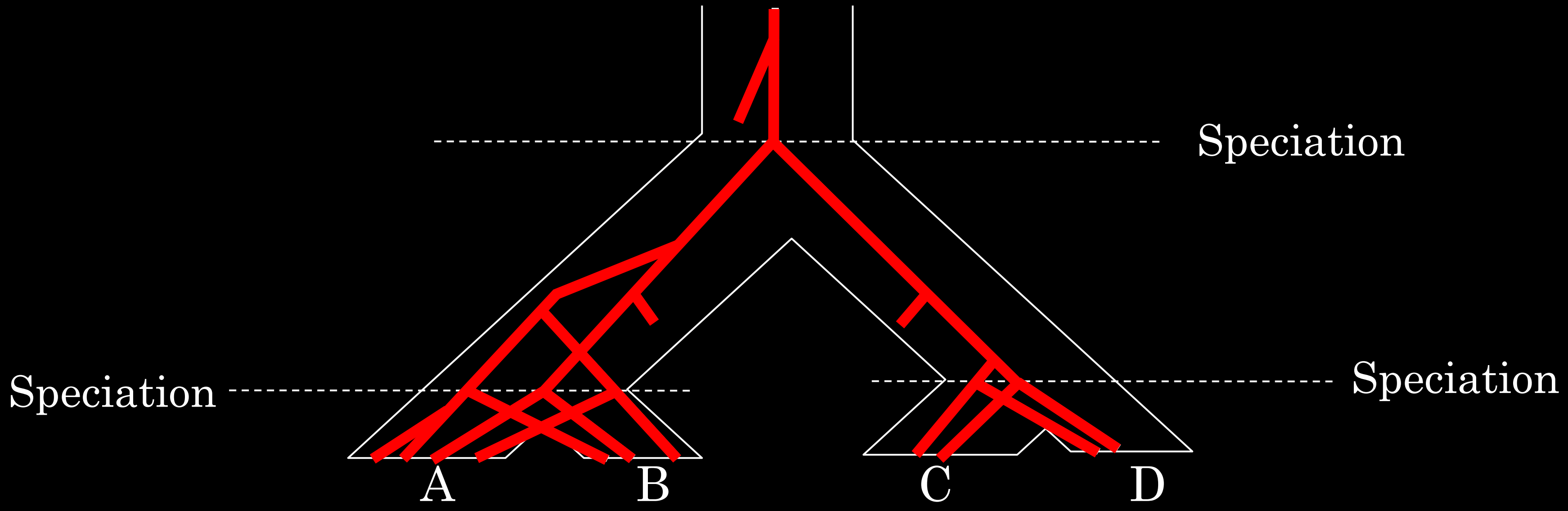


Genes can duplicate and disappear at any time, independently of speciation



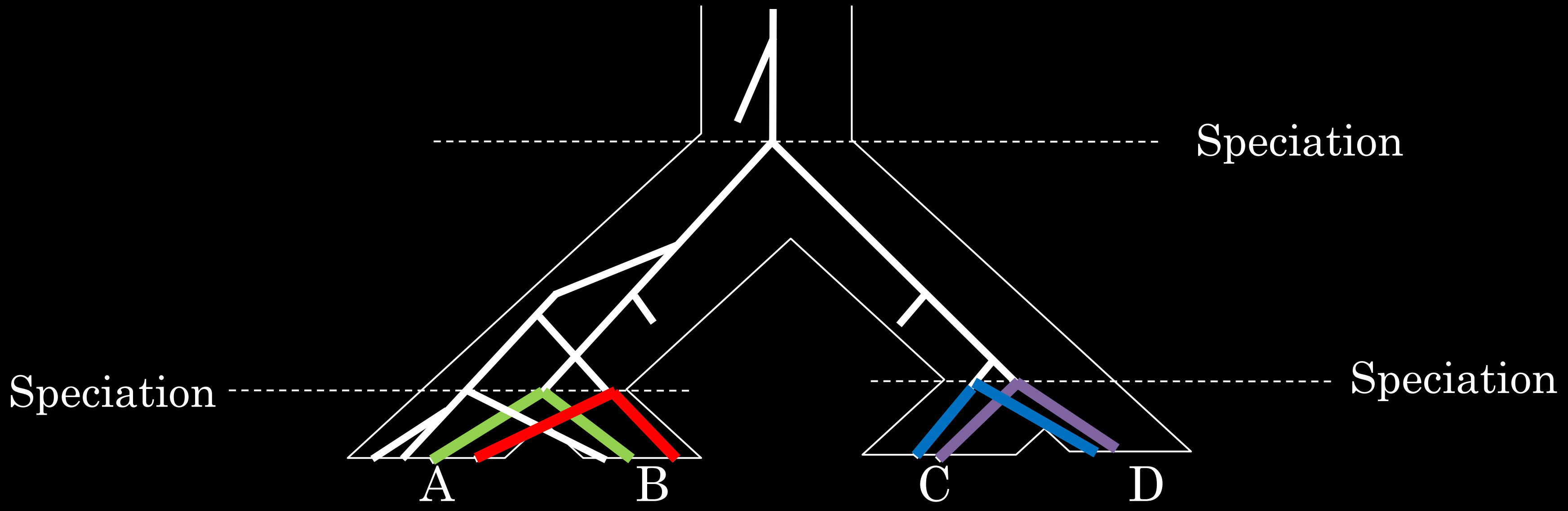
Homology

- Any set of sequences (or, more generally, traits) that are descended from a common ancestor



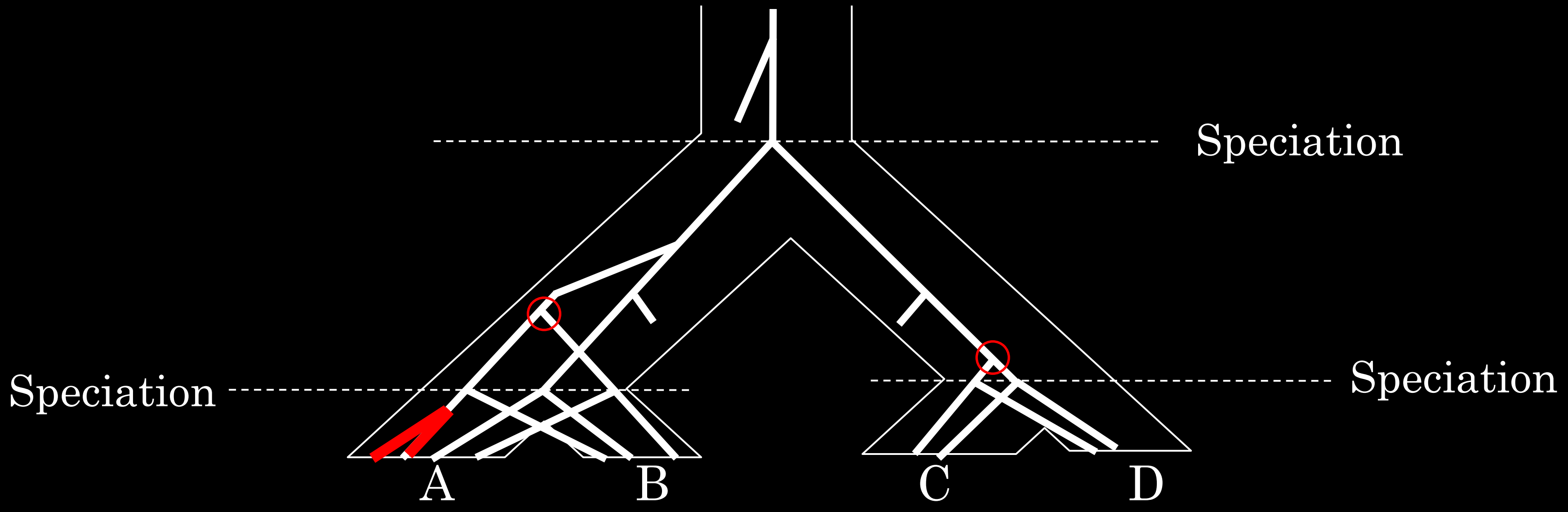
Orthology

- Genes whose last common ancestor occurred at a speciation event



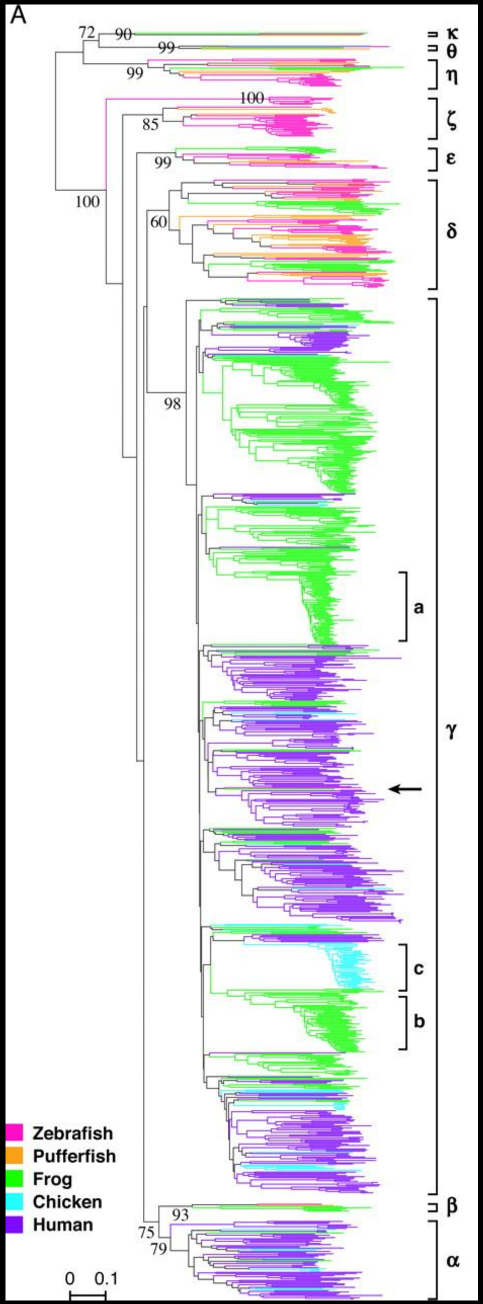
Paralogy

- Two genes whose last common ancestor arose via a gene duplication event



Why do we care about the difference?

- The “ortholog conjecture”
- Orthologs are (kinda) the “same gene” in different species, so they are more likely to have evolved in similar ways and retained the same or similar function
- Paralogs are descendants in a single lineage – they are often lost, but since someone is still doing the original job, they may mutate to specialize over time, or “explore” new functions
  - Some antibiotic-resistance genes are paralogs of much more “innocent” genes

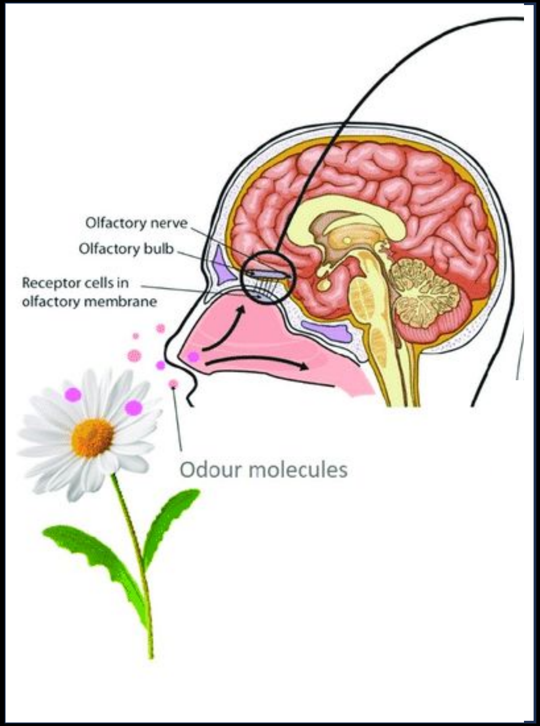


Niimura and Nei (2005) *PNAS*
Vilela et al (2019) *Foods*

Olfactory receptors

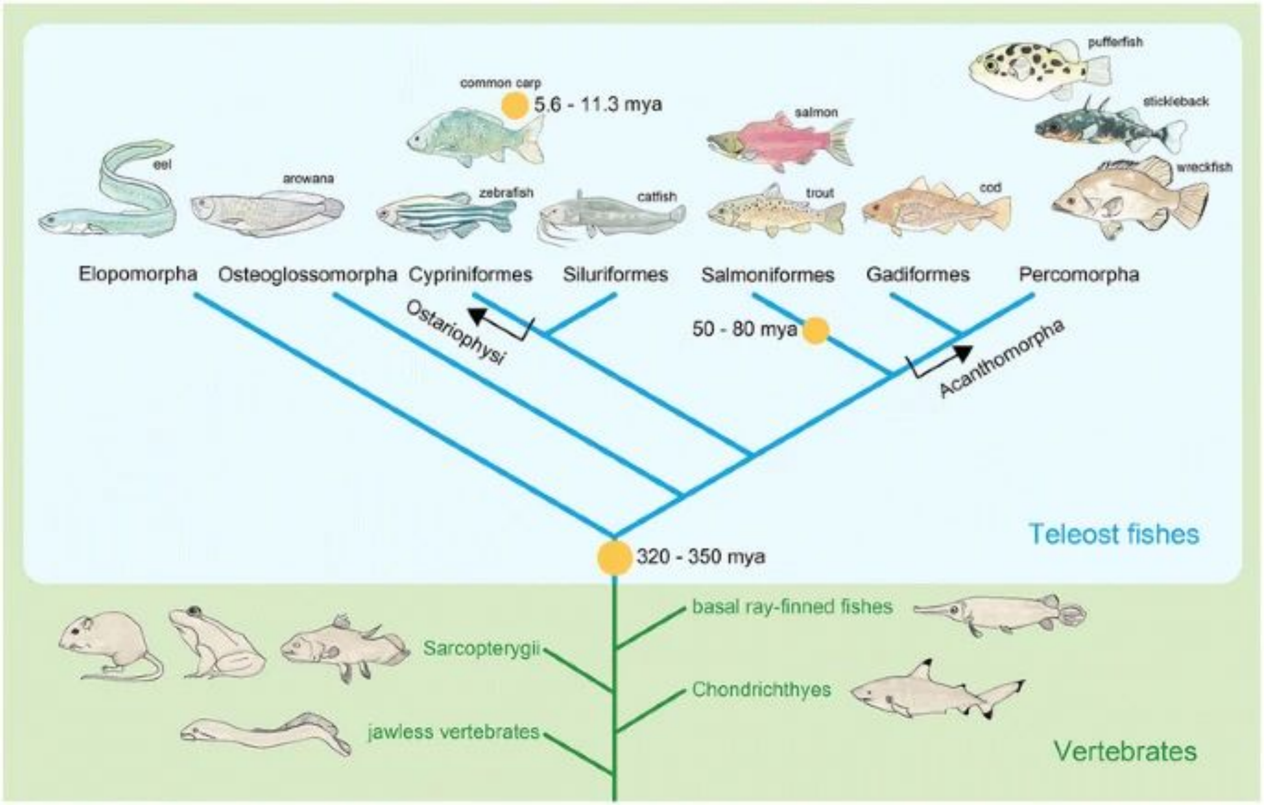
Many different target molecules, but all ORs:

- Sit in the cell membrane
- Recognize a specific set of odor molecules
- Trigger neural signalling



These sequences are all homologous, but trying to disentangle paralogy and orthology is a mess

Whole-genome duplication – mega paralogy!



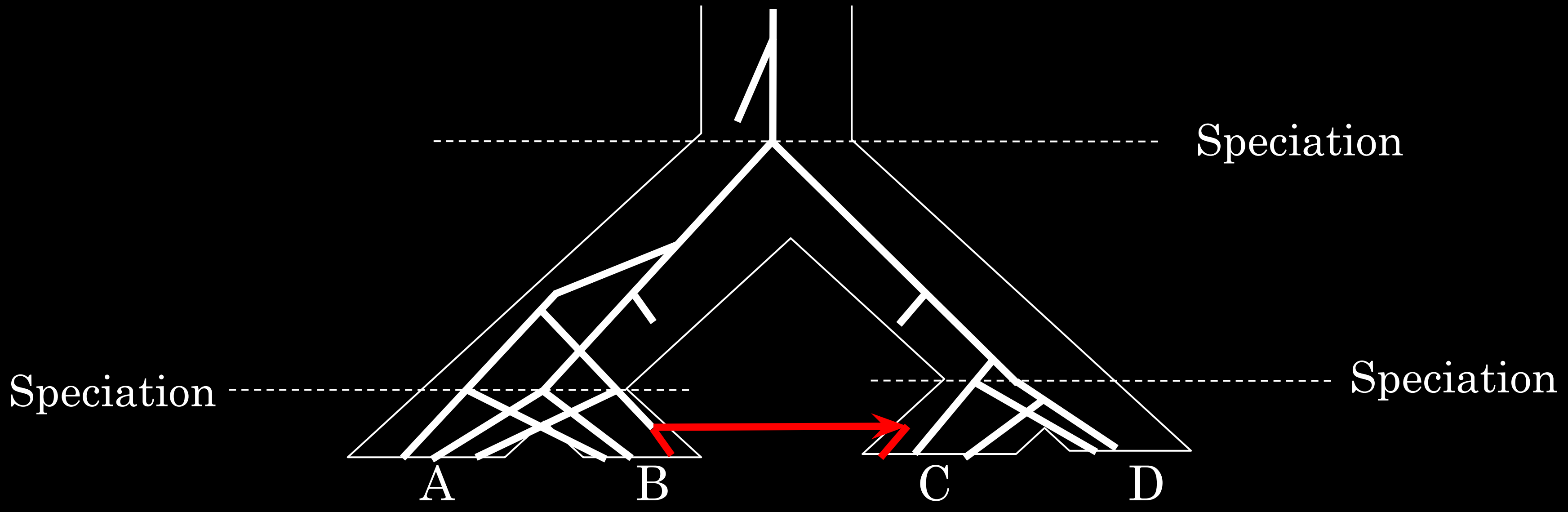
Descendants have 2x the genetic material of the ancestor

Many of the newly duplicated genes are lost or mutate and gain new / specialized functions

Glasauer and Neuhauss (2004) *Molecular Genetics and Genomics*

Xenolog: lateral gene transfer

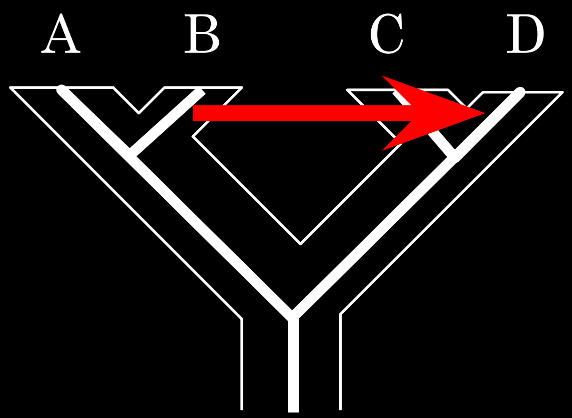
- Homologous sequences found in different species because of horizontal/lateral gene transfer



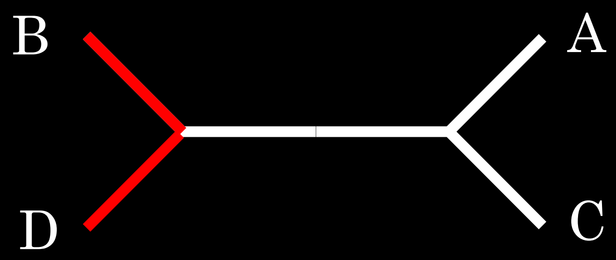
Uh-oh that ain’t a tree anymore!

Lateral gene transfer

- The transmission of genetic material between genomes in a manner other than parent to offspring
- A gene from organism ‘B’ is acquired by organism ‘D’ - thus B and D appear to be very closely related

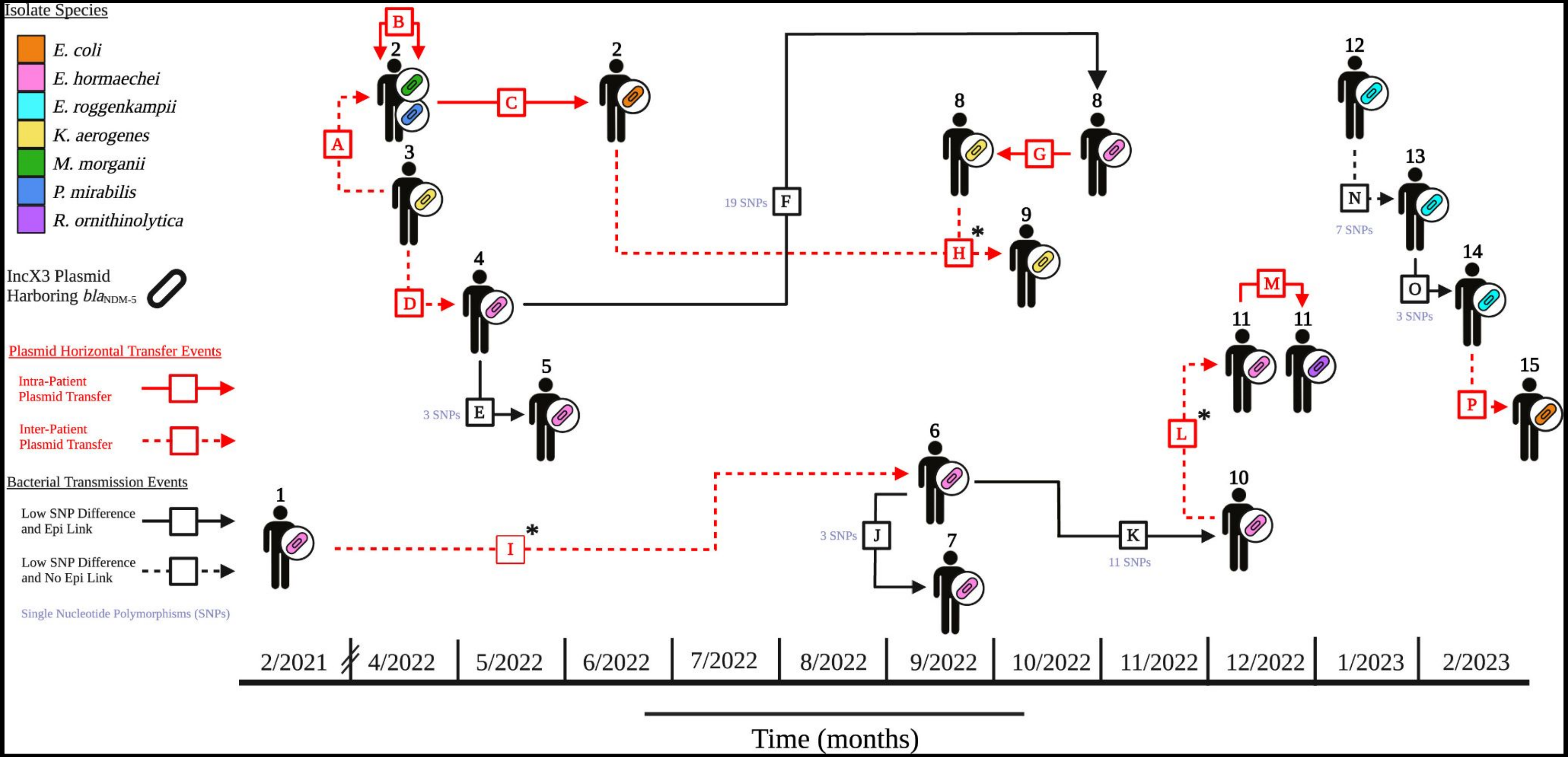


Organism phylogeny with LGT event

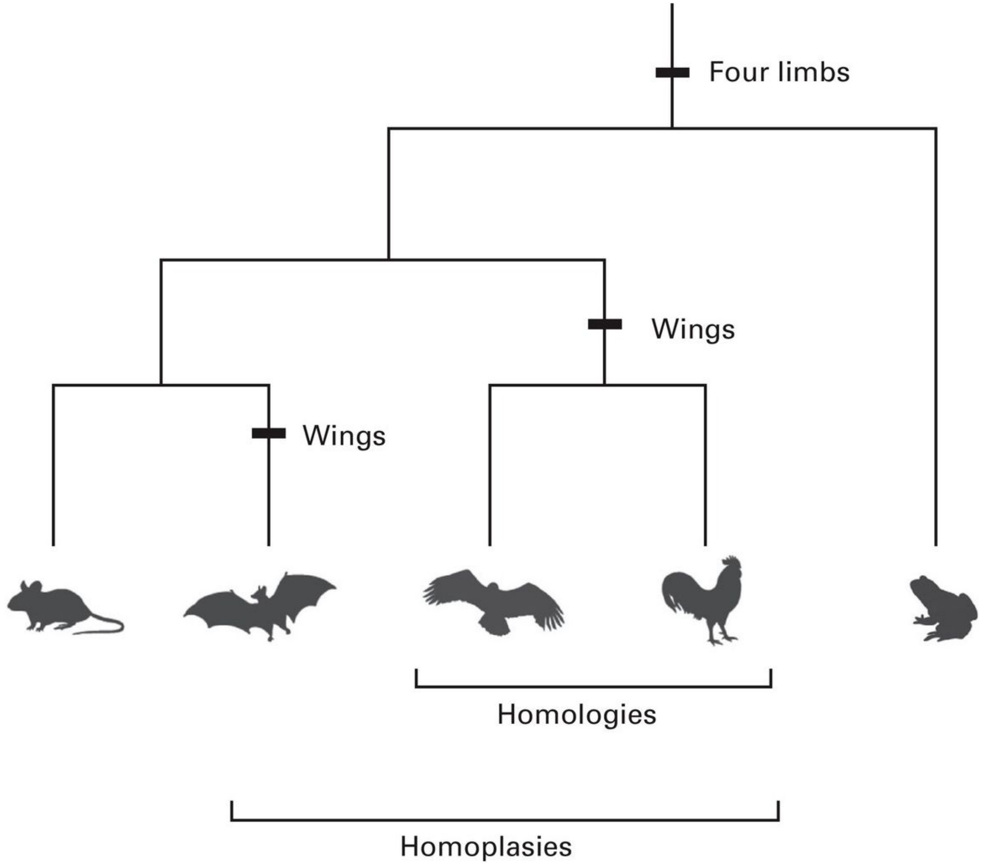


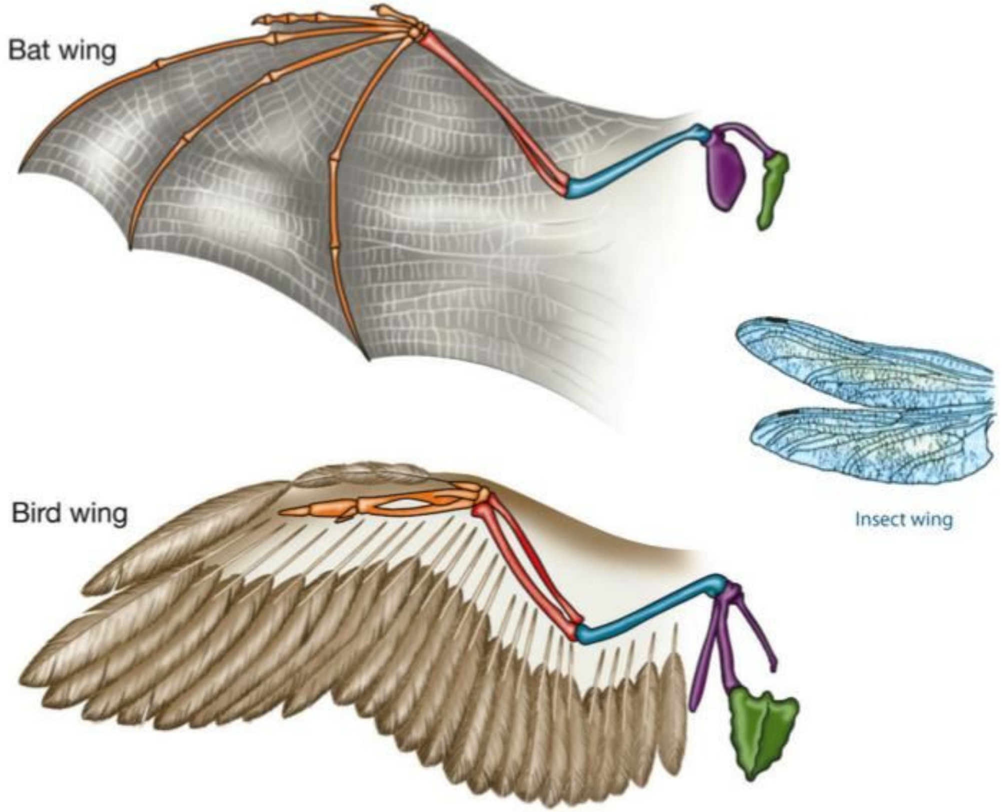
Inferred gene tree

LGT has major clinical implications

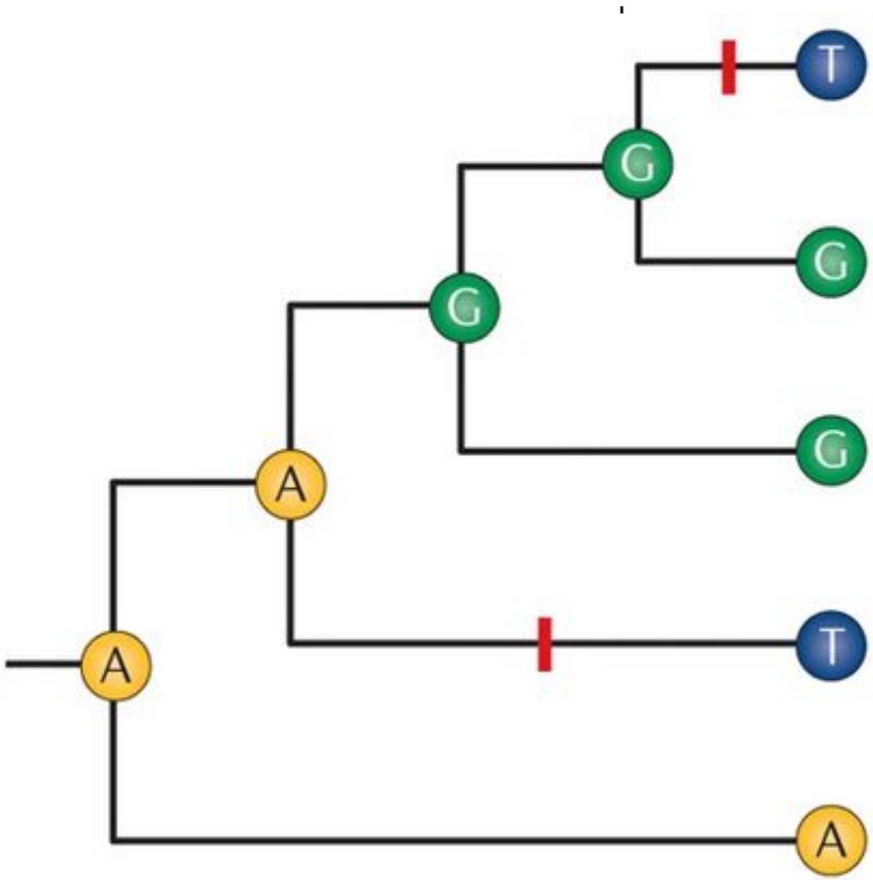


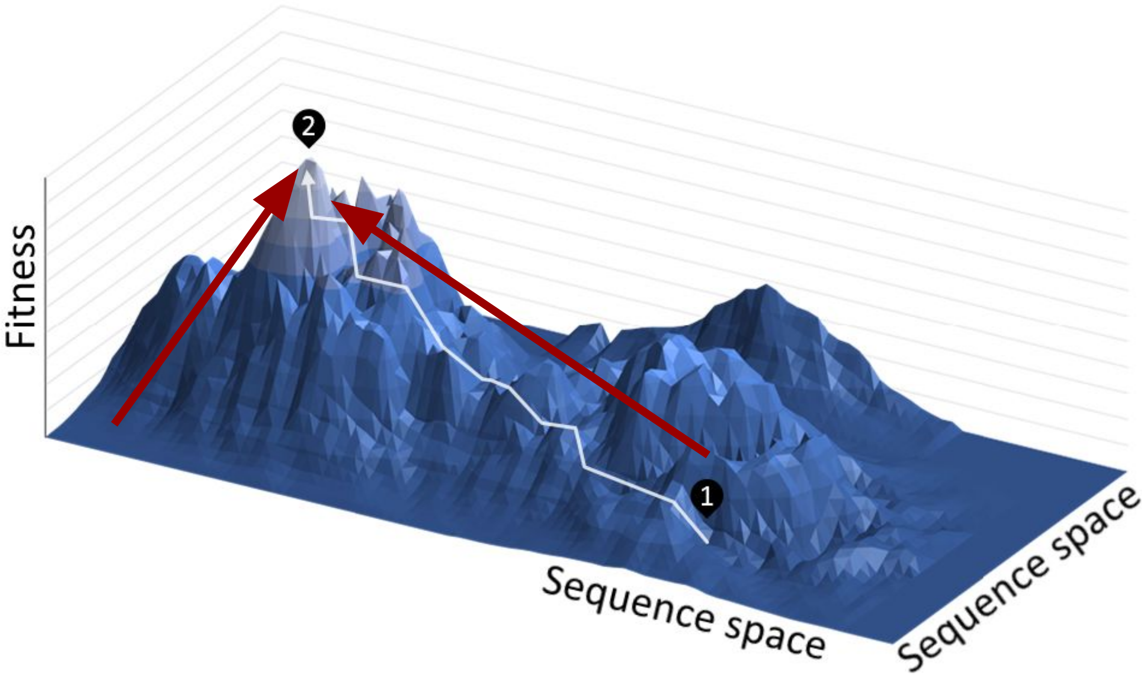
Convergent evolution of wings



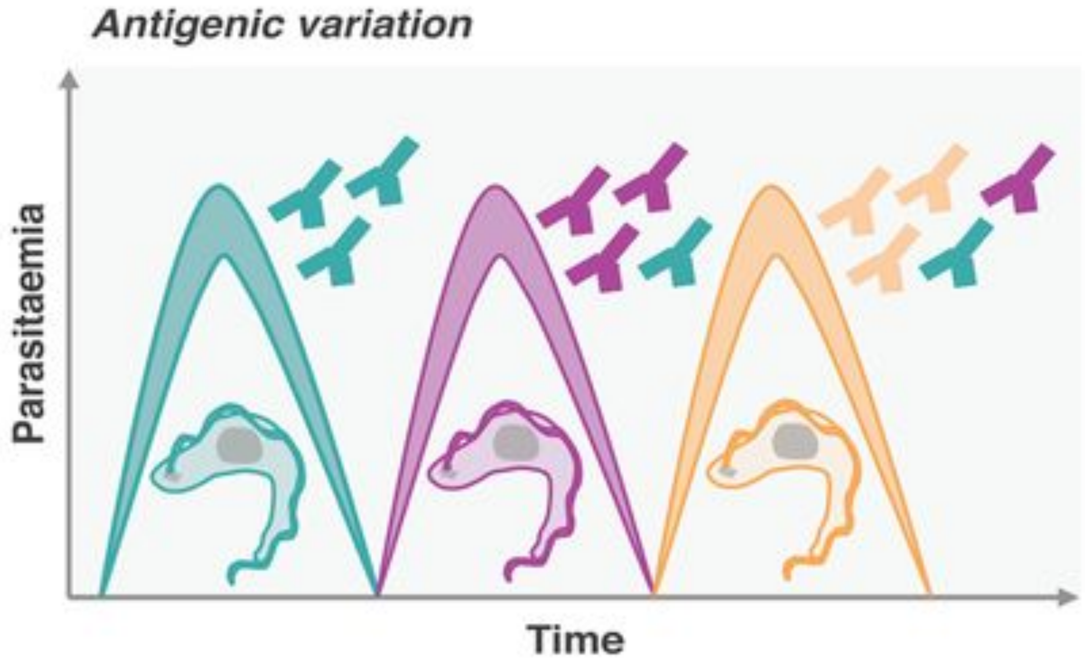


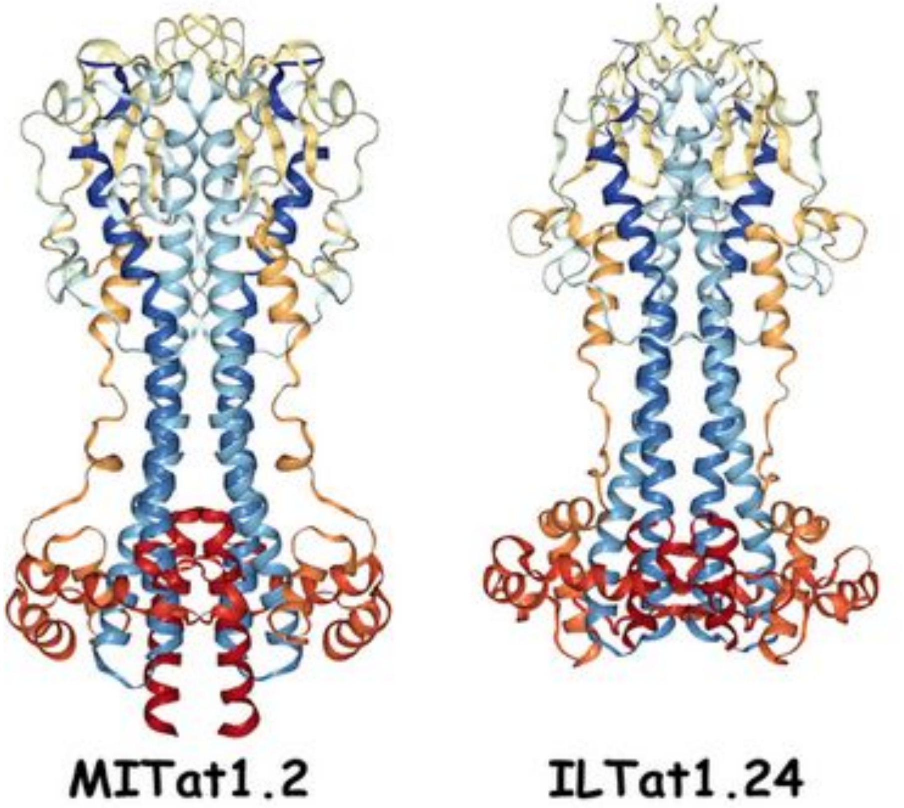
Convergent (or homoplasic) mutations common





Convergent structures without convergent sequences





16% sequence identity!

https://www.researchgate.net/figure/SG-Structure-A-The-solved-structures-of-the-NTDs-of-three-type-A-VSGs-are-presented_fig2_328589540

Summary

Mutations change the gene sequence, and often the protein sequence as well

Selection weeds out less-effective versions of the protein (and thus the mutation)

By tracing the history of mutation and selection, we can recover the evolutionary relationships among biological sequences

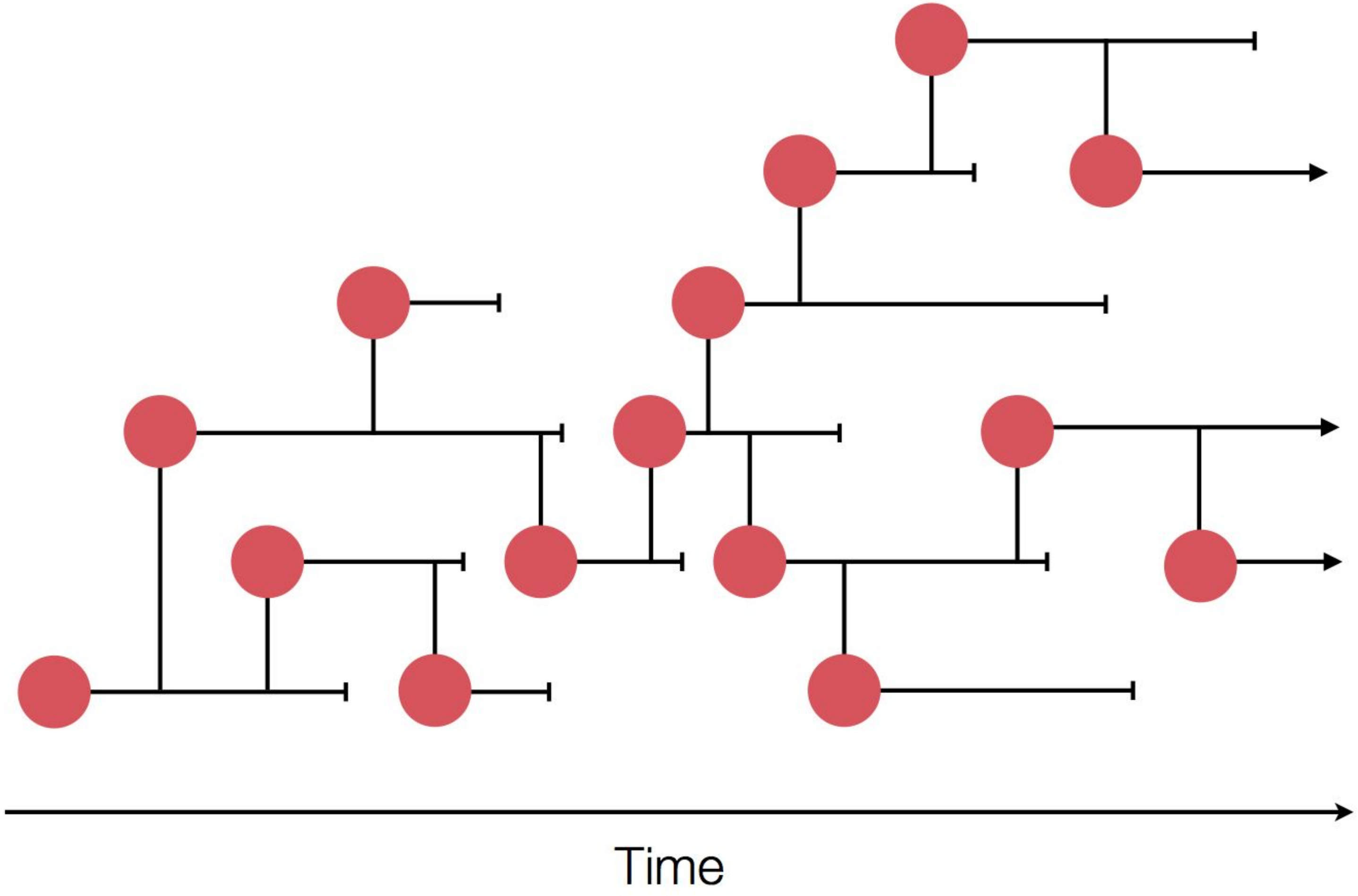
Summary

All genes that are related by descent from a common ancestor are homologous

Orthologs are sometimes thought of as ‘the same gene in different species’

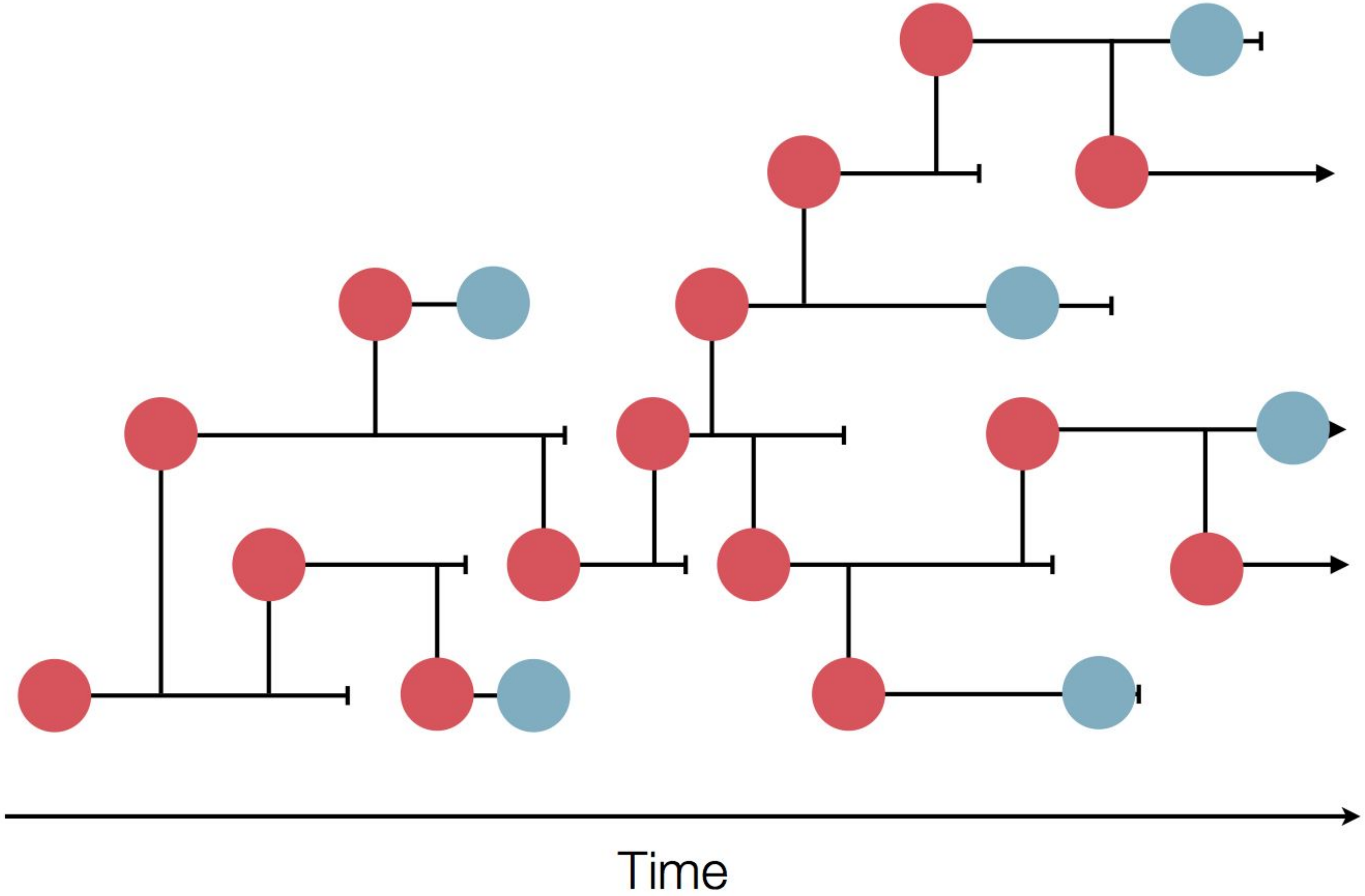
Paralogs arise by duplication, and can lead to proteins with new functions

Sampling & partially reconstructing underlying epidemic process



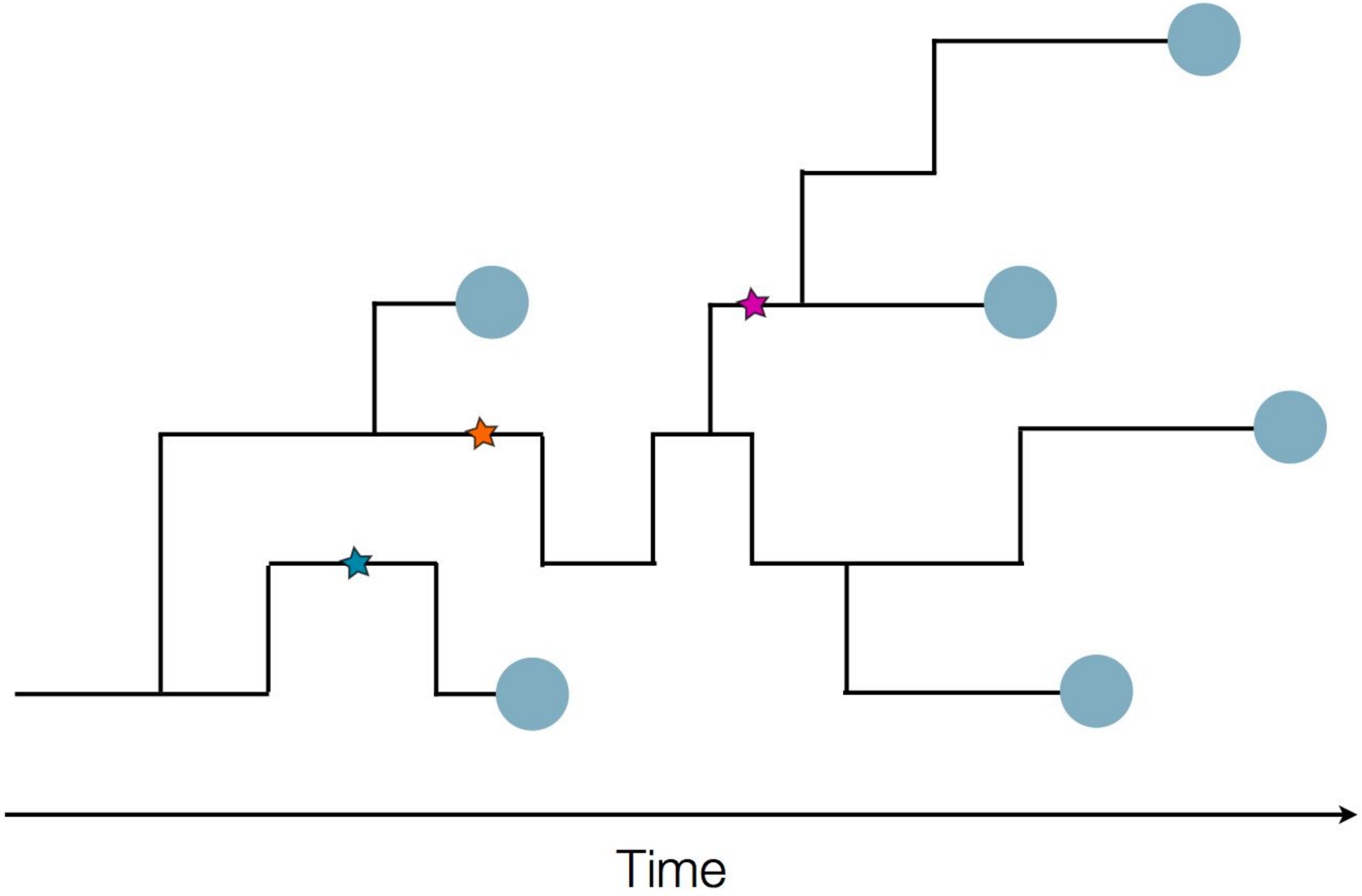
https://github.com/trvrb/phylodynamics-lecture

Sampling & partially reconstructing underlying epidemic process



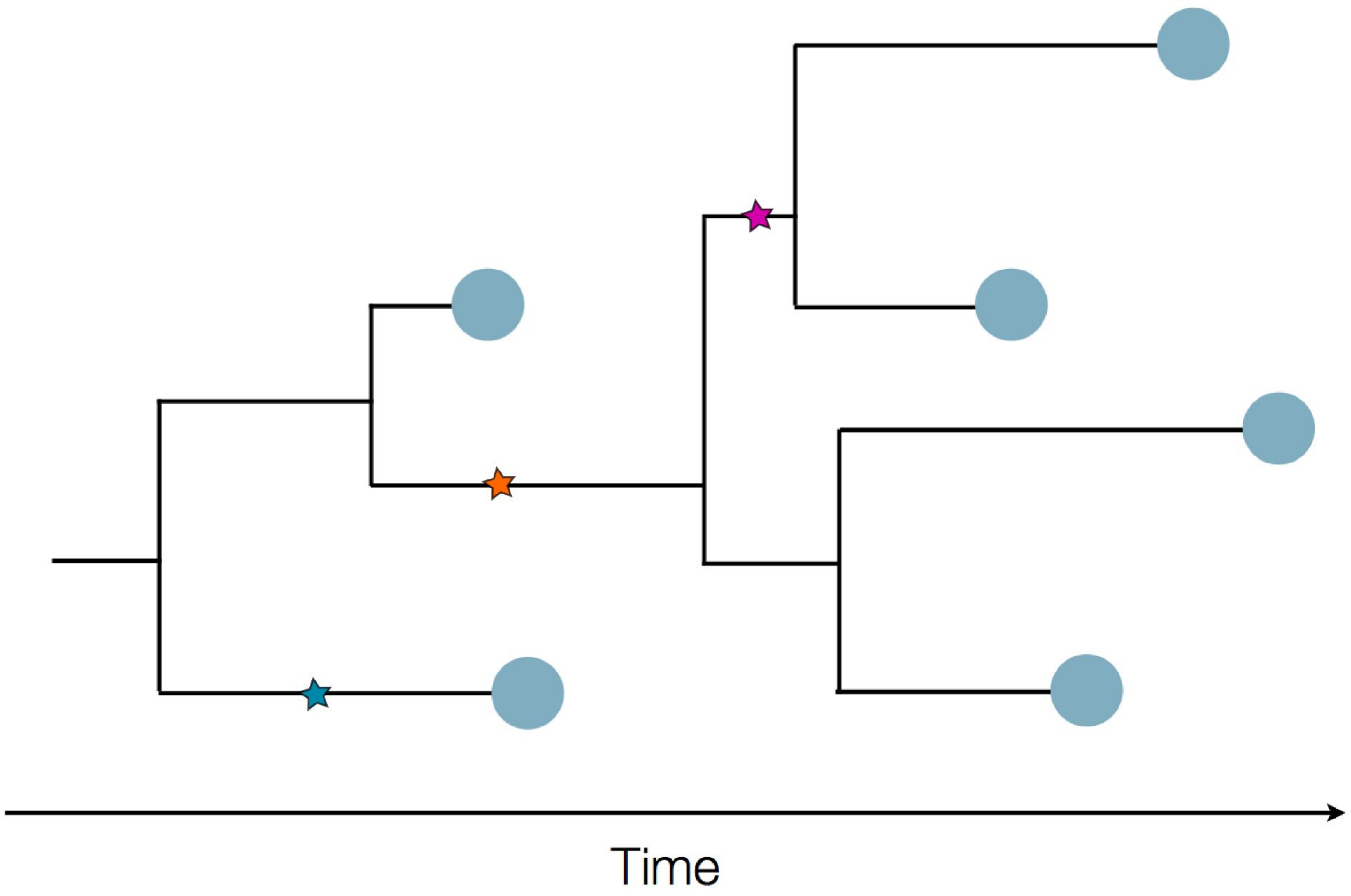
https://github.com/trvrb/phylodynamics-lecture

Sampling & partially reconstructing underlying epidemic process



https://github.com/trvrb/phylodynamics-lecture

Sampling & partially reconstructing underlying epidemic process



https://github.com/trvrb/phylodynamics-lecture