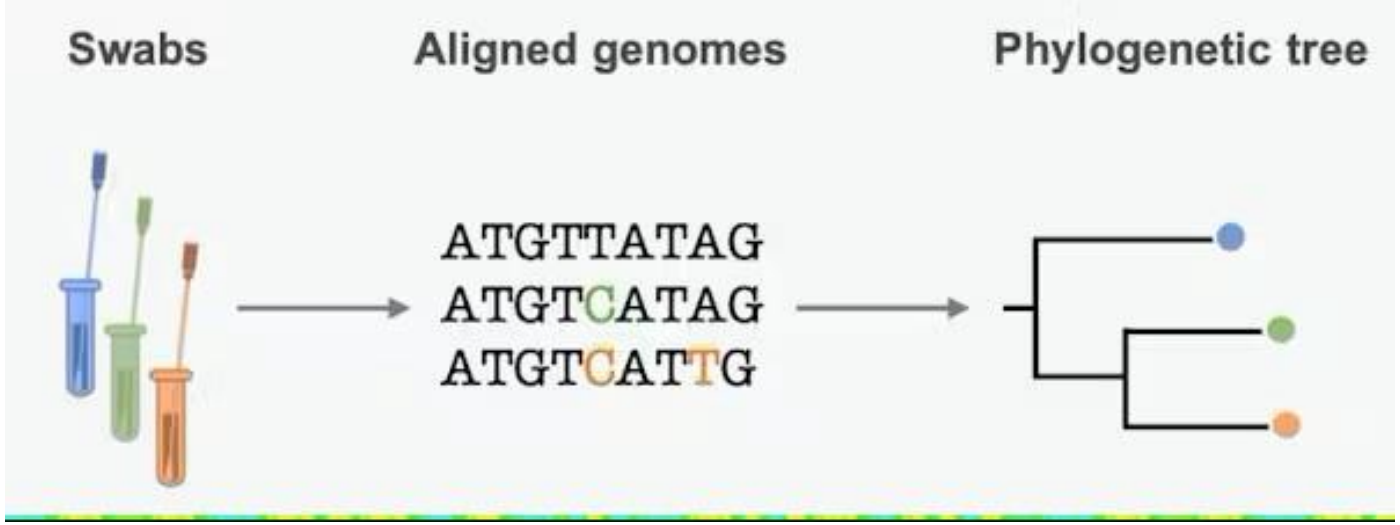
Session 2

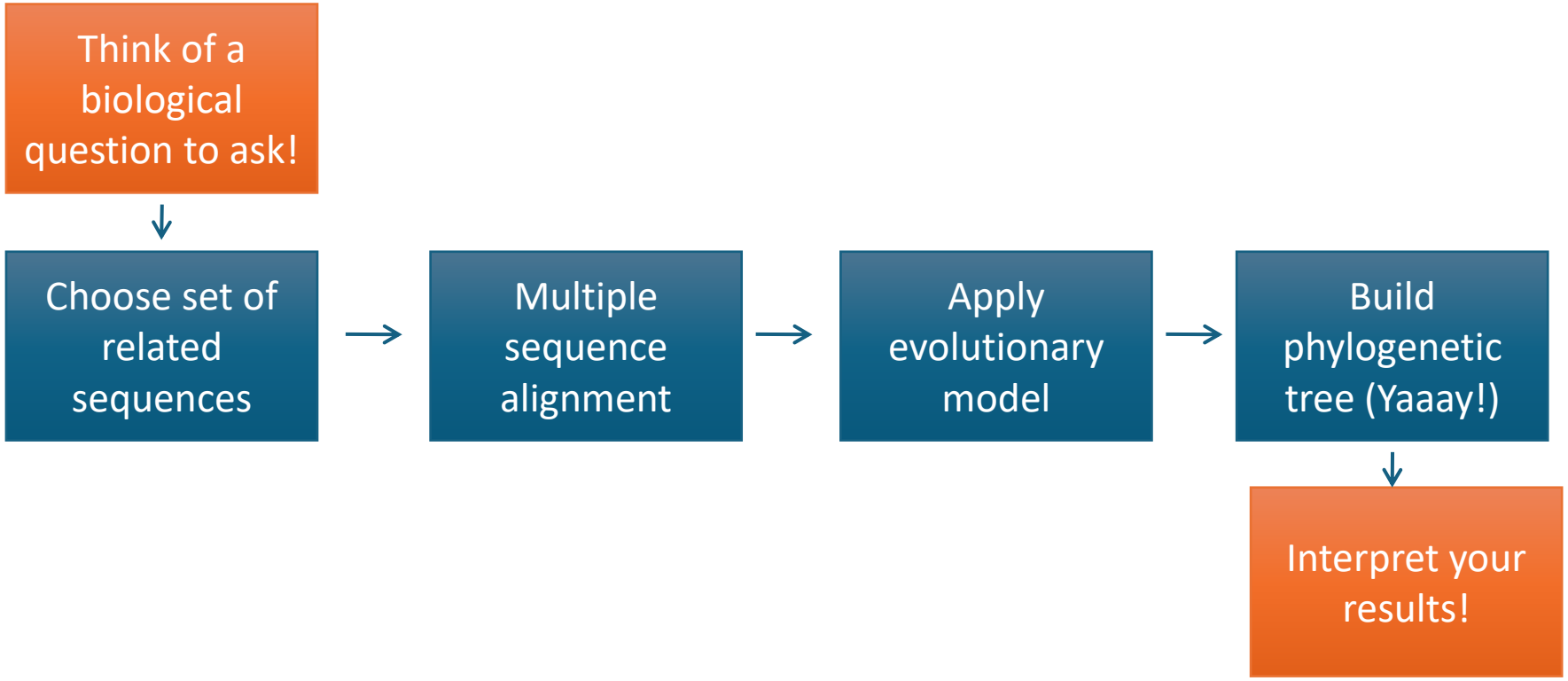
Genomic epidemiology basics

Genetic data is a lens into the history of an outbreak



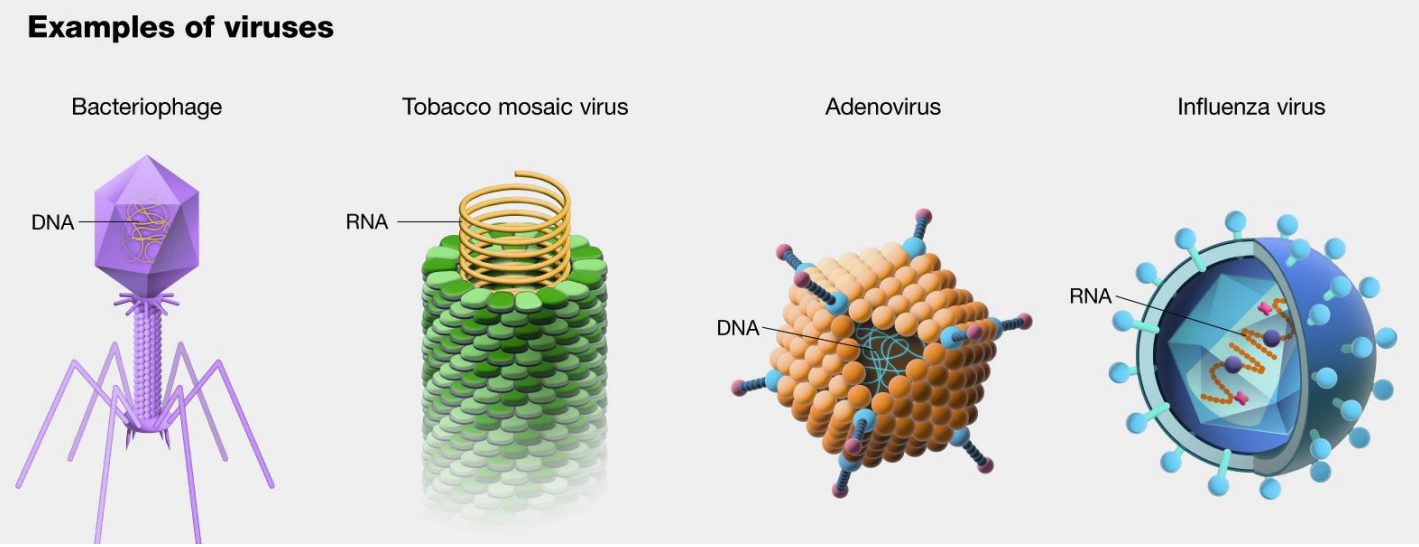
- As pathogens replicate, they make errors, like a game of telephone
- By sequencing multiple individuals in an outbreak, we can use these errors to reconstruct a likely history of the disease
- We can also track when new challenges develop, e.g., antibiotic resistance

The generic phylogenetic analysis workflow



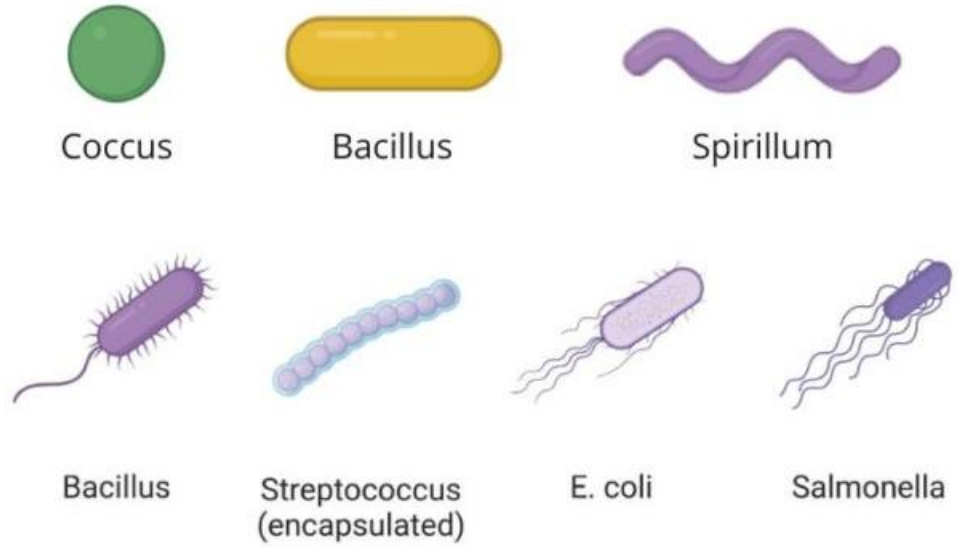
The types of pathogens and what they mean for genomics

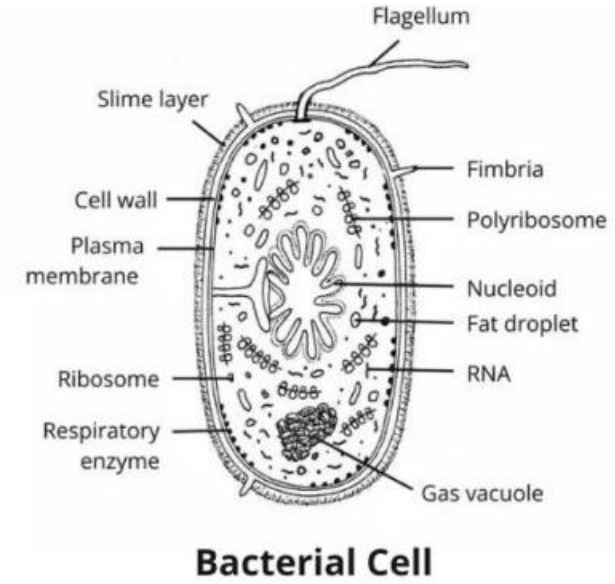
Viruses



- The simplest type of pathogen, a bundle of genetic material wrapped in a protein coat that can hijack our cells
- Genome sizes vary between 2kb to 1Mb long
- Some viruses have segmented genomes, allowing them to shuffle gene segments with other viruses infecting the same cell (e.g., flu)

Bacteria

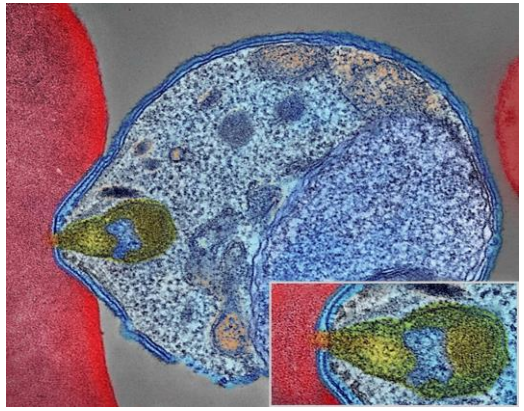


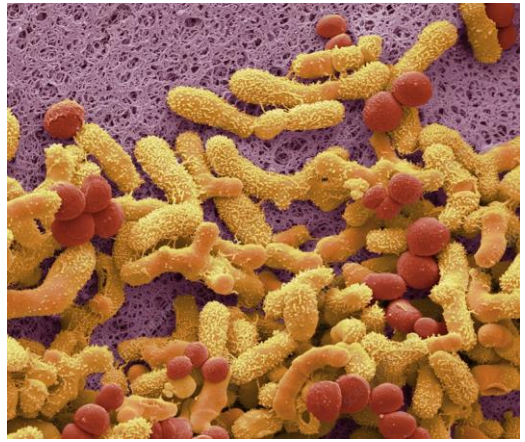


- 500 kb to 12Mb long genomes
- Often have extra-chromosomal material, like plasmids, which can allow them to trade genes (like antibiotic resistance) between one another and complicate our analyses

Fungi and other eukaryotes

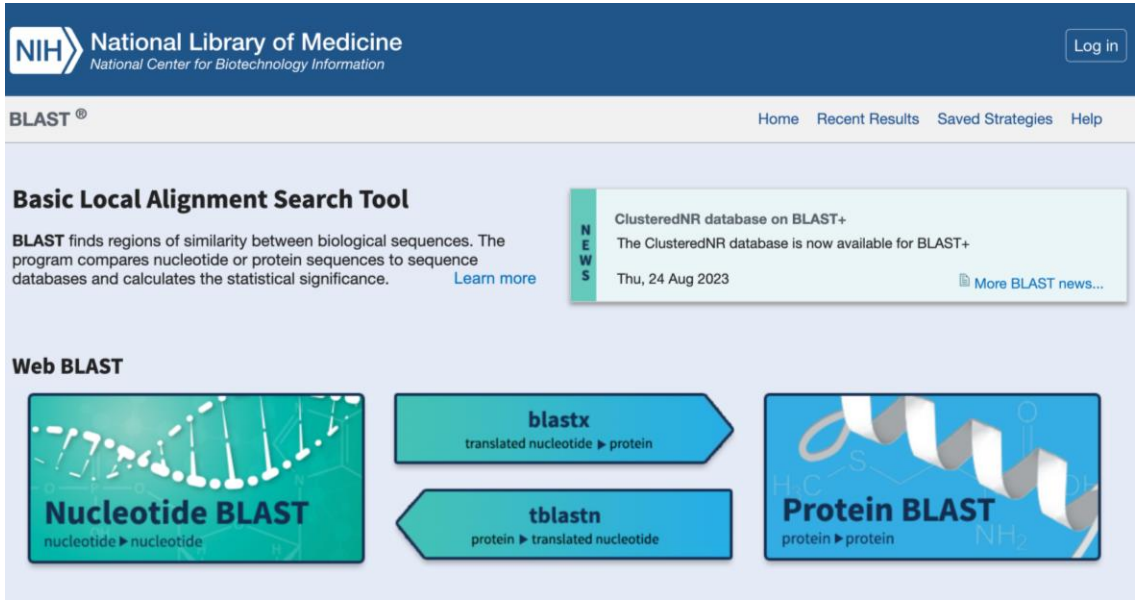
- Include unicellular pathogens like protozoa, fungi like yeast, and multicellular parasites like tapeworm.
- From 10Mb to 100 Gb long genomes
- Recombination and multiple chromosomes common





Pathogen identification

Pathogen detection using sequencing

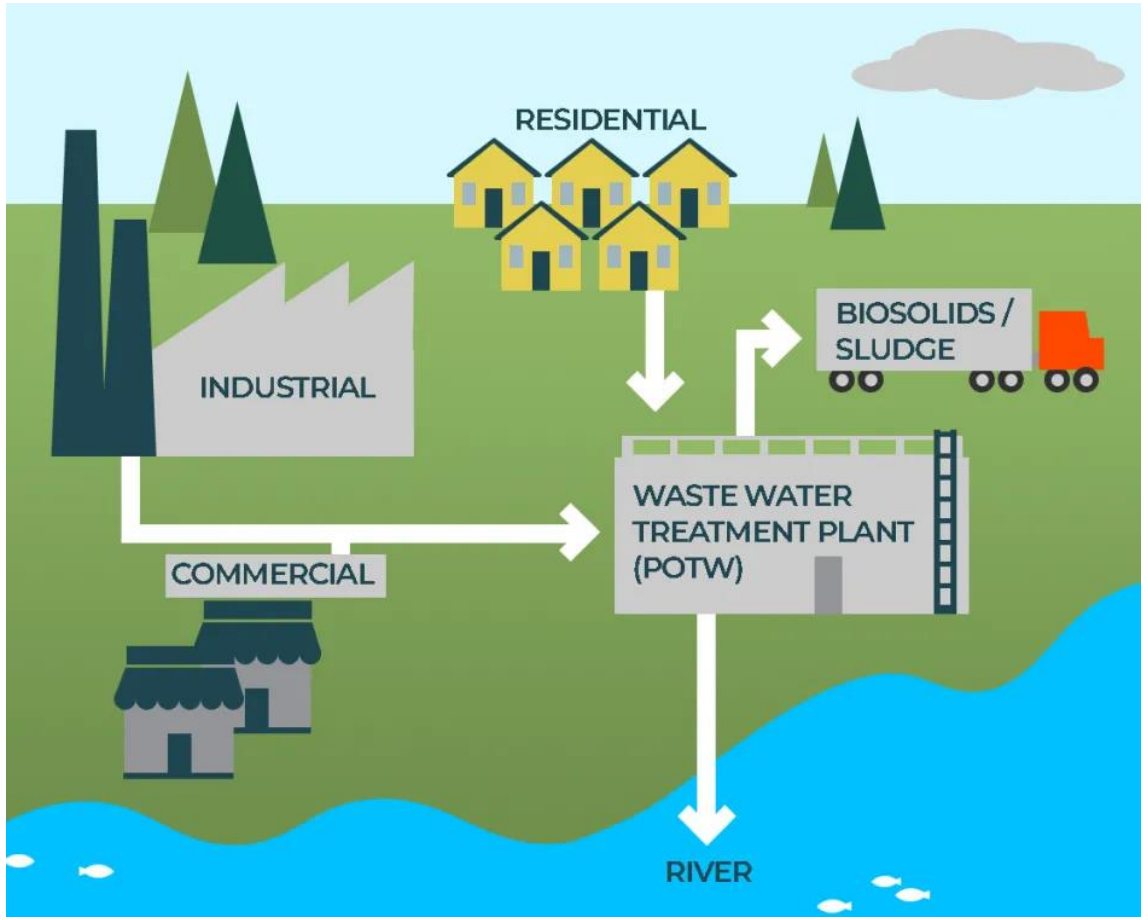


- Sometimes you have access to the sequence, but you are unable to culture a microbe from an isolate
- Instead, you can use molecular techniques to identify the pathogen
- Tools like BLAST allow you to search for similar sequences in a database and identify likely organisms that match

Scaling up, sometimes you're curious about what pathogens circulate in a community

- Everybody poops, so we can sometimes use wastewater to identify transmission of pathogens

- Recently, this has become popular as SARS-CoV-2 increased the need for indirect surveillance of transmission



Antimicrobial resistance

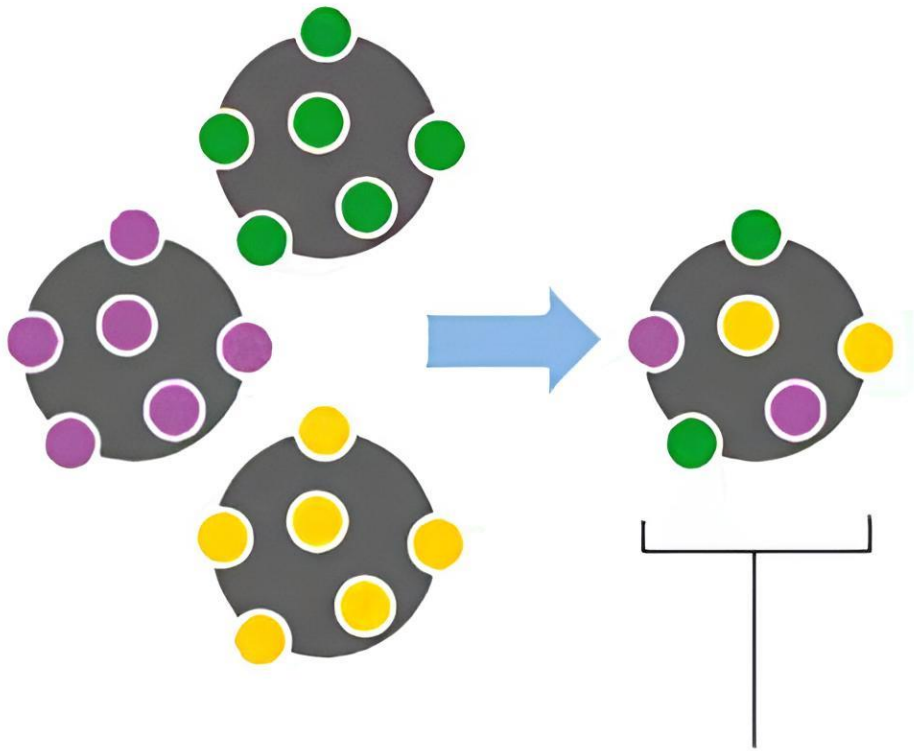
Antimicrobial Resistance (AMR)

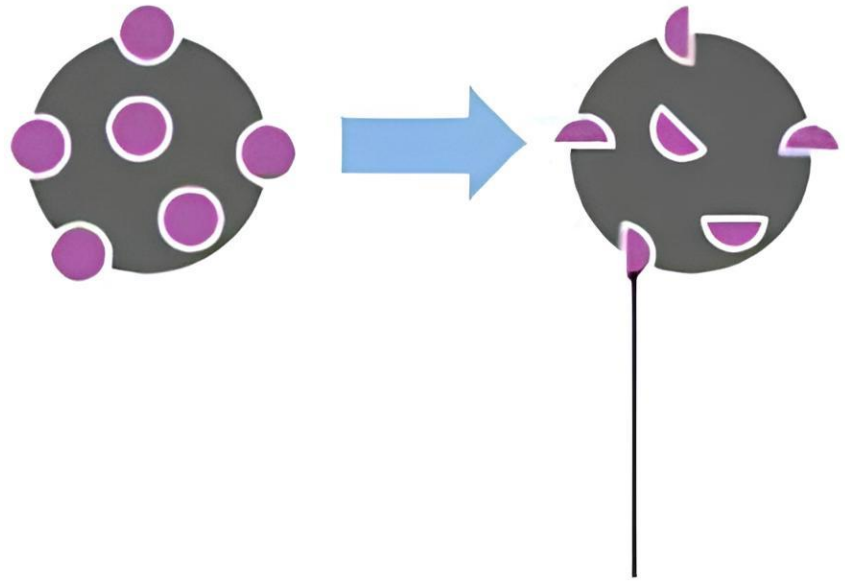
- AMR is one of the top global public health threats
- Bacterial AMR was estimated to be directly responsible for 1.27 million global deaths in 2019, and contributed to an additional 4.95 million deaths globally
- AMR doesn’t just impact folks who have infections; it makes c-sections, chemotherapy, and surgery much riskier
- AMR impacts folks everywhere, but it most acutely impacts low and middle-income countries

Genomic Drift and Shift

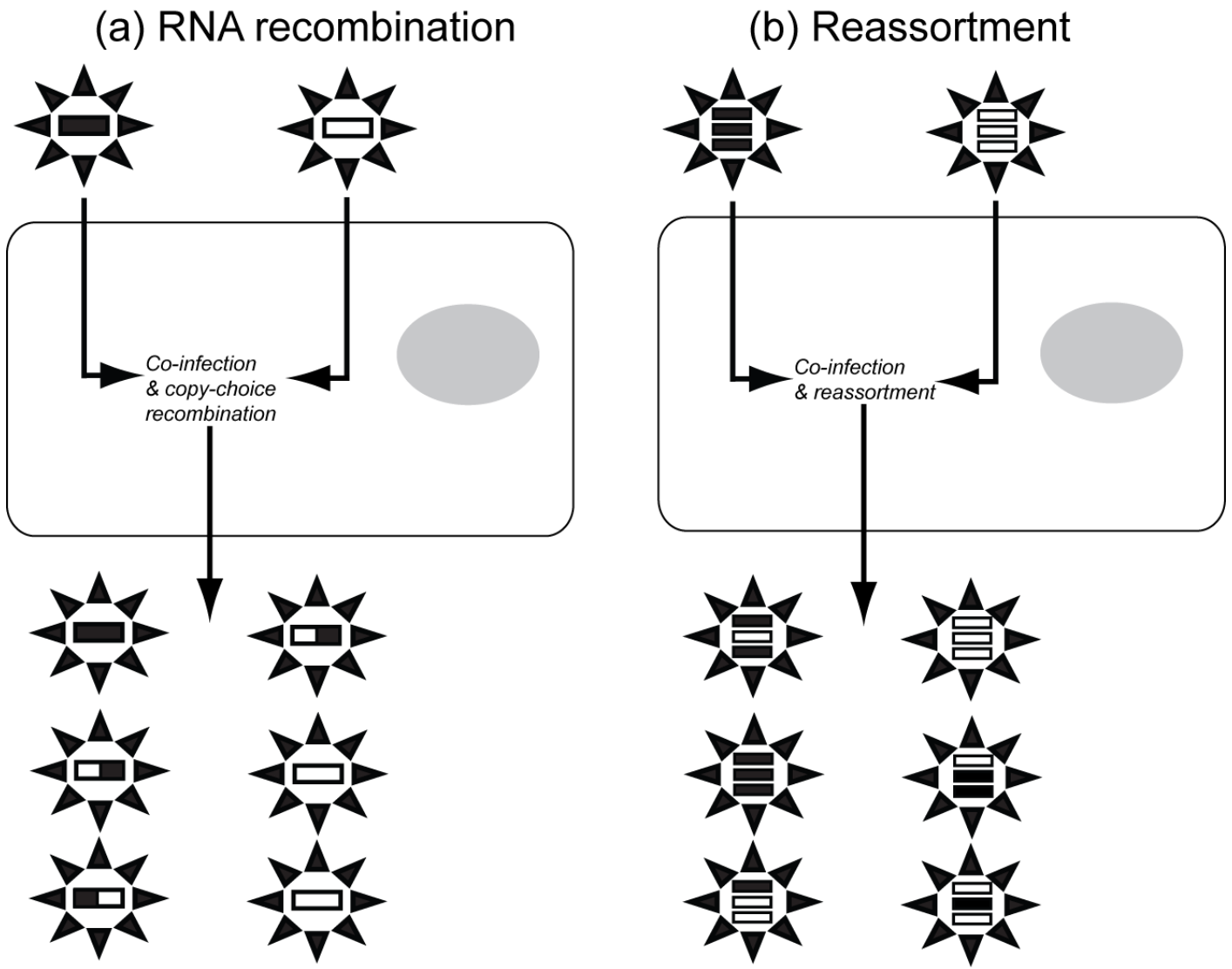
Differences Between

Antigenic shift & Antigenic drift

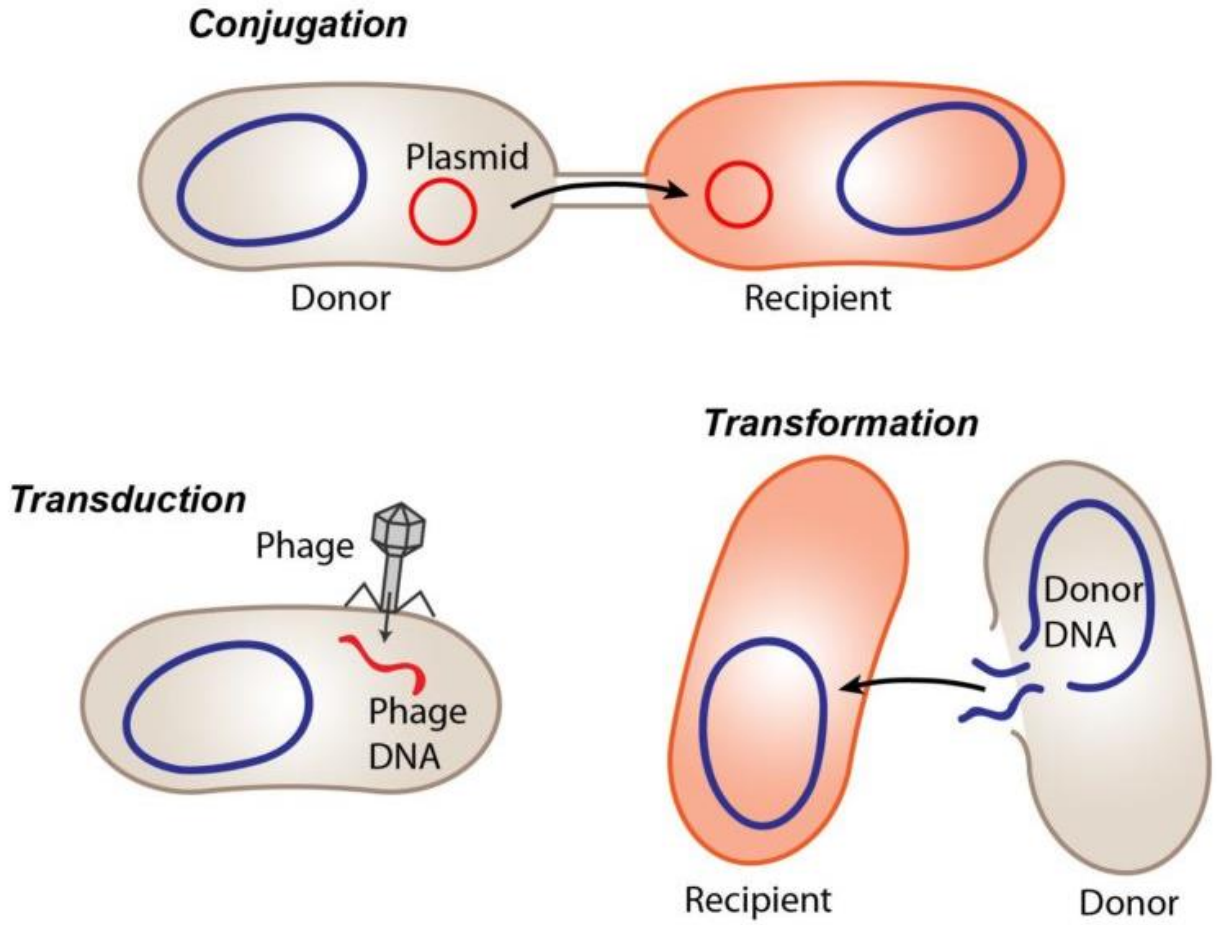




Viruses


(a) RNA recombination
Co-infection
& copy-choice
recombination
(b) Reassortment
Co-infection
& reassortment

Bacteria


Conjugation
Plasmid
Donor
Recipient
Transduction
Phage
Phage
DNA
Transformation
Donor
DNA
Recipient
Donor

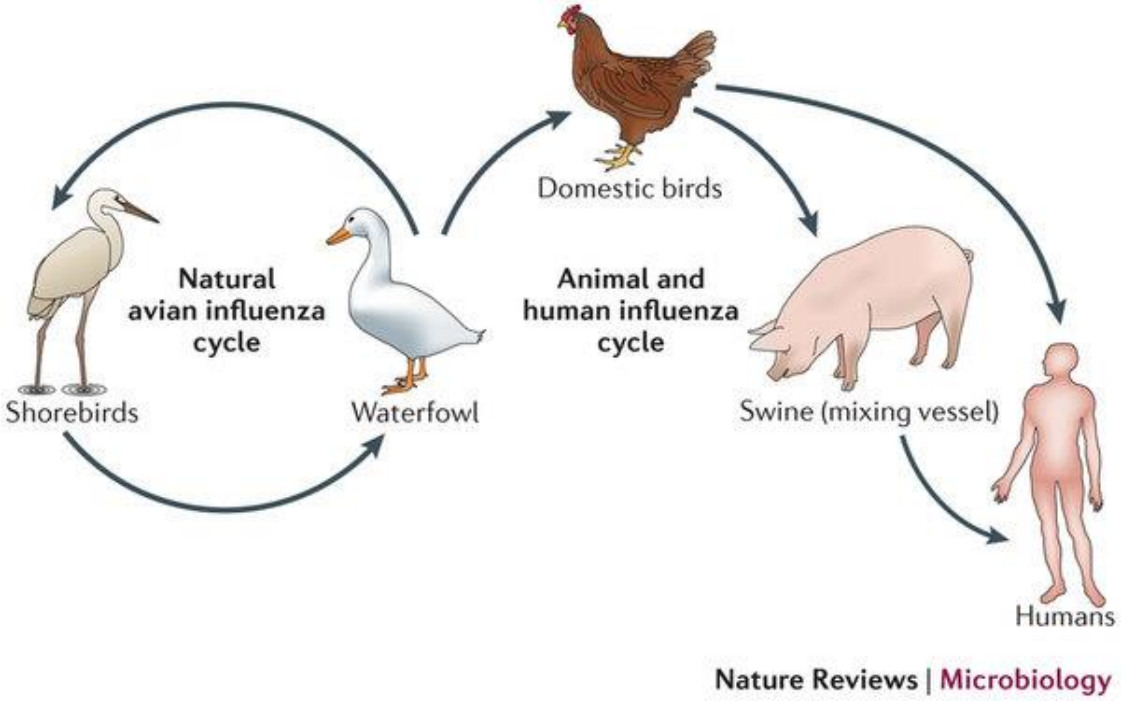
Antigenic shift and pandemics

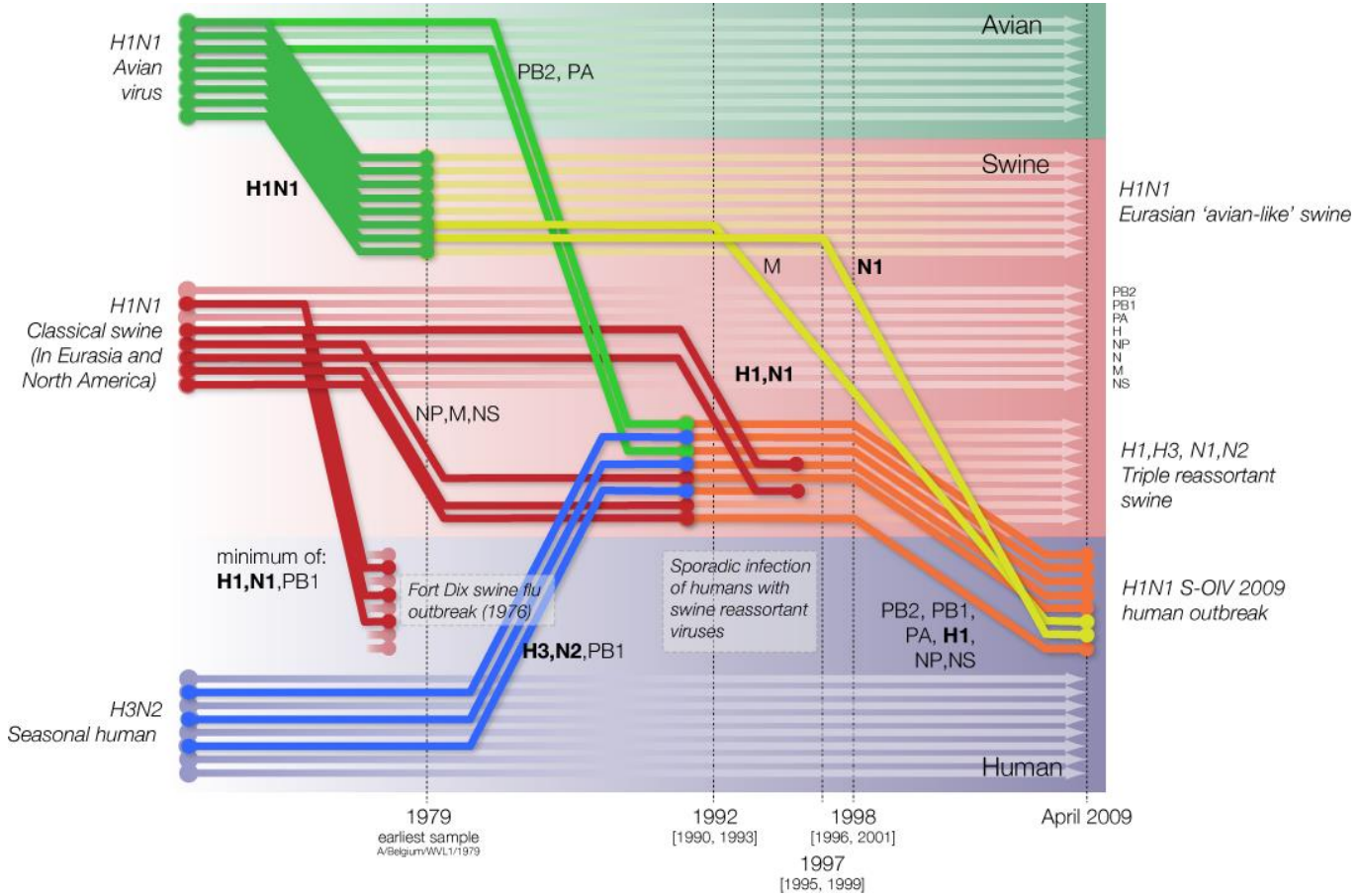
- Influenza is a perfect example of the danger in antigenic shifts
- The 1918 H1N1 influenza pandemic (aka the Spanish Flu) killed 50 million people worldwide
  - Lowered life expectancy in the US by 12 years
  - Higher death rates among healthy adults 15-34 years old
- This outbreak, and many others caused by antigenic shift events

Since 1918

- Three other pandemics:
  - 1957 H2N2 pandemic
  - 1968 H3N2 pandemic
  - 2009 H1N1 pandemic
- Each of these has caused large increases in mortality, but none to the same degree
- As a result of influenza's ubiquity and reassortment frequency, many nations have stood up surveillance programs specific to influenza

Genomic Shift: H1N1pdm (Swine Flu)





Why genomic epidemiology?

Epidemic Process

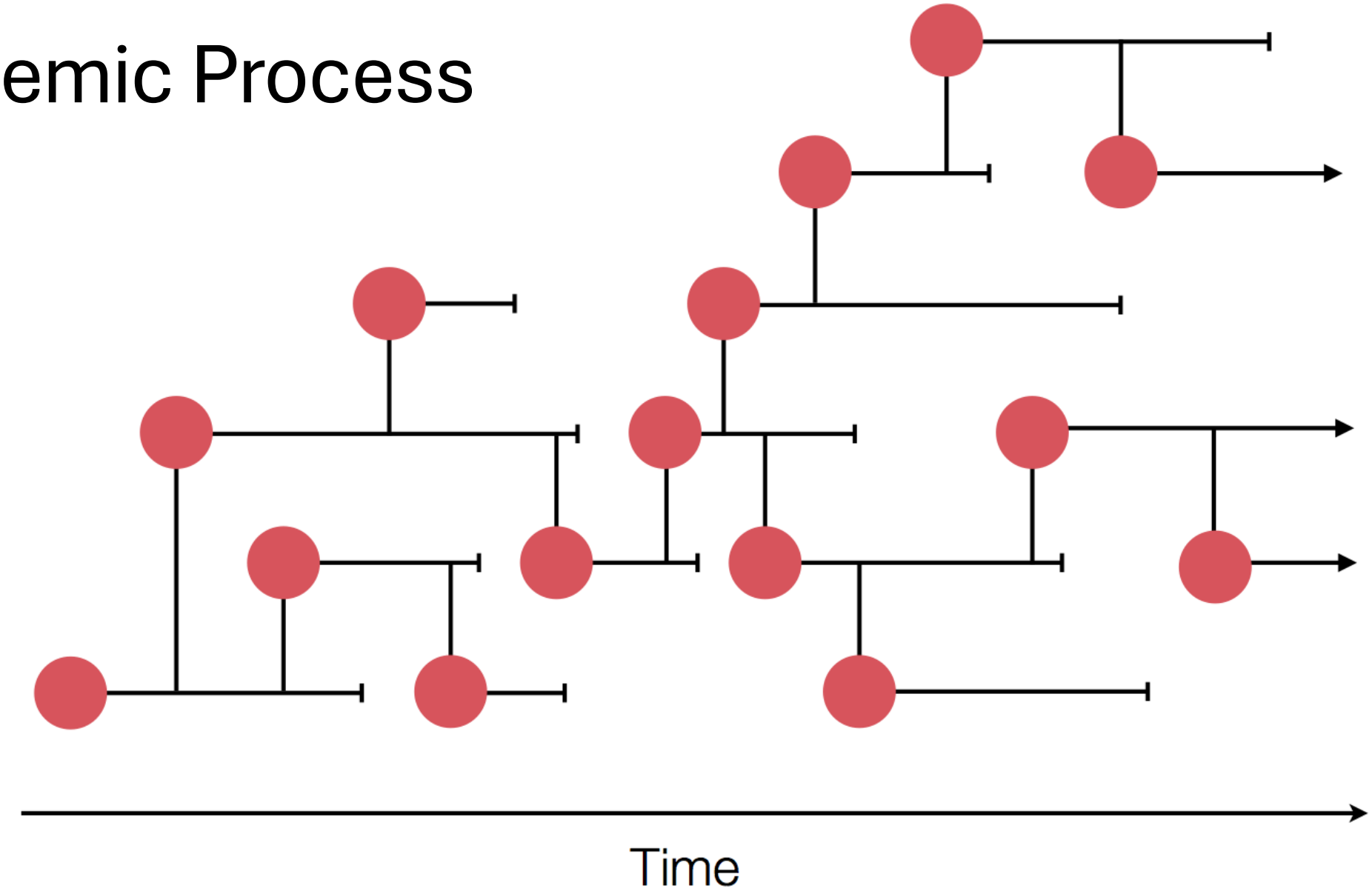
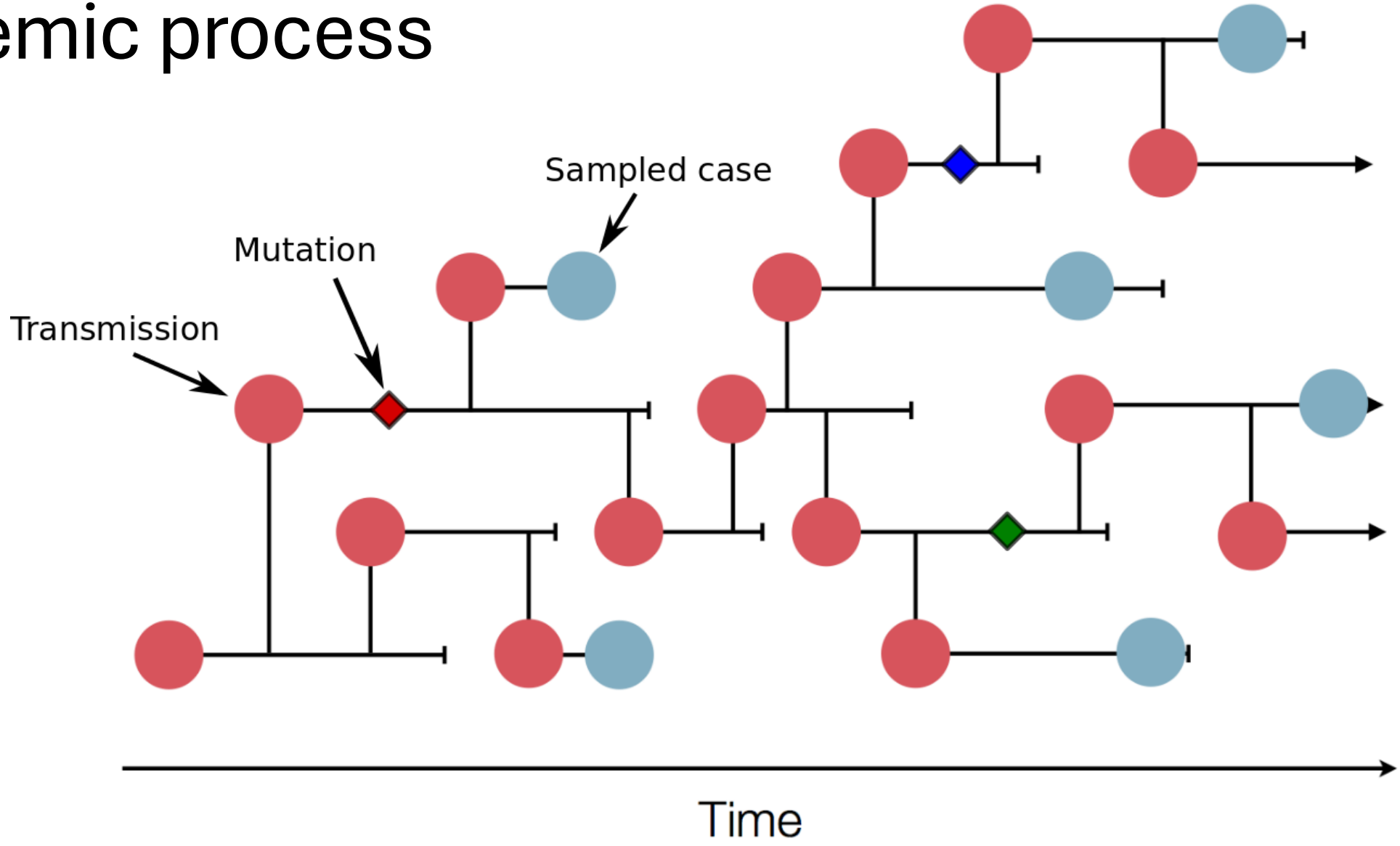
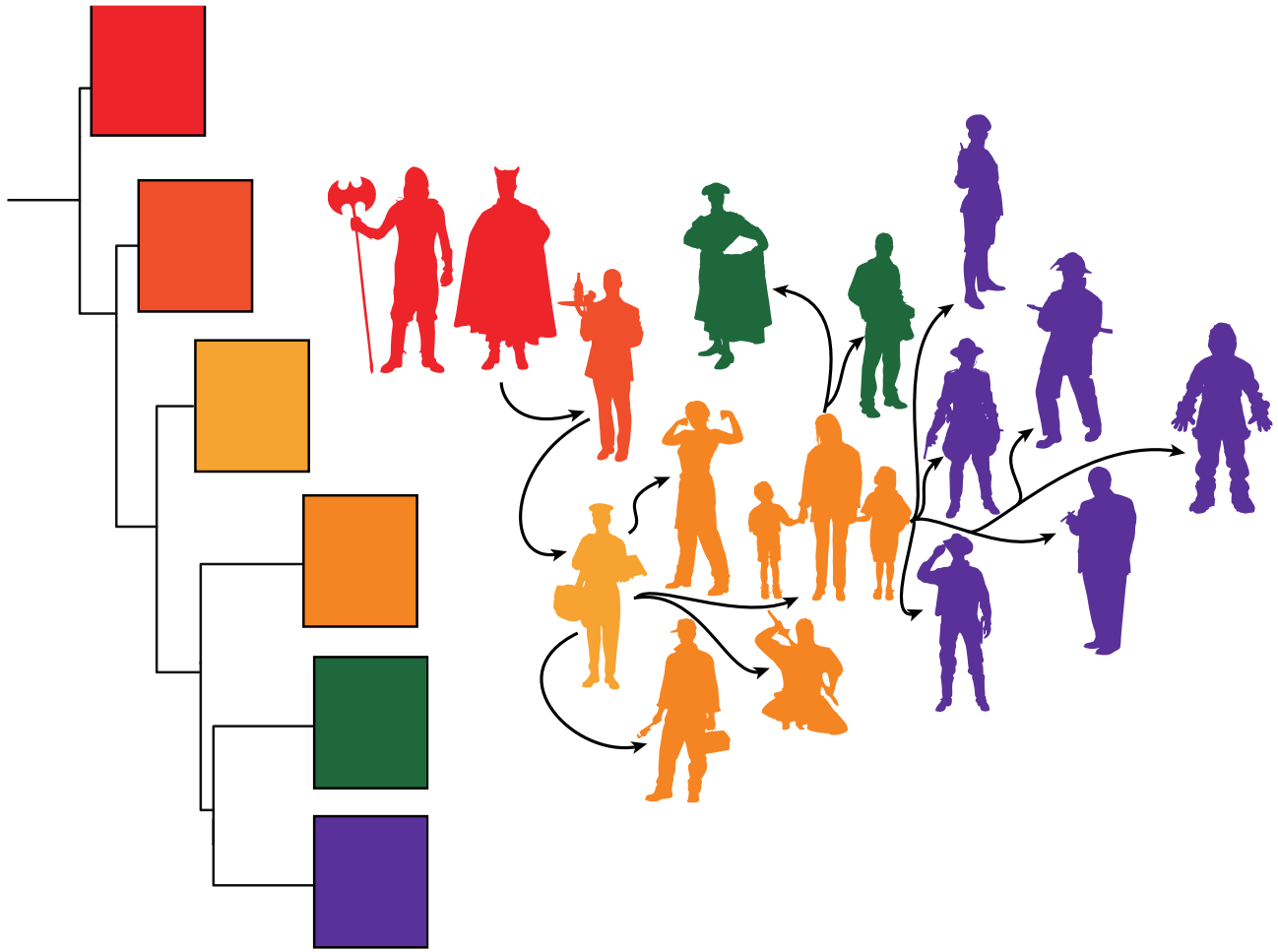


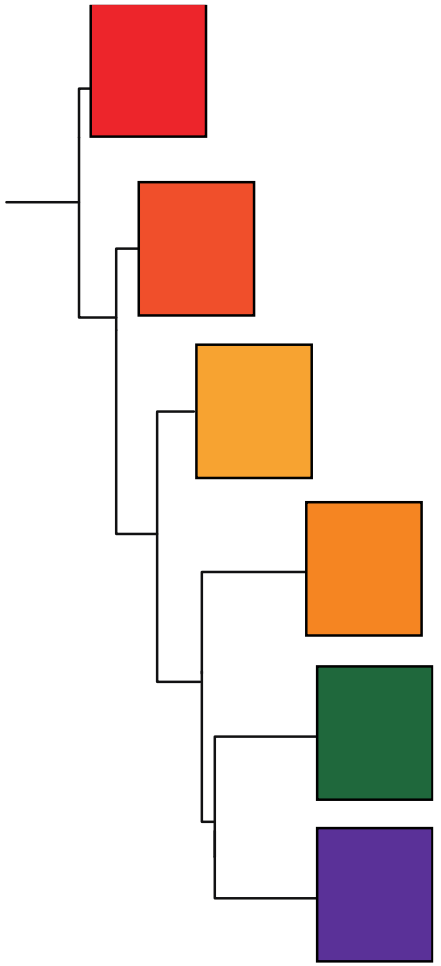
Image from Trevor Bedford: https://bedford.io/projects/phylodynamics-lecture/intro.html#/6

Viruses accumulate mutations during the epidemic process





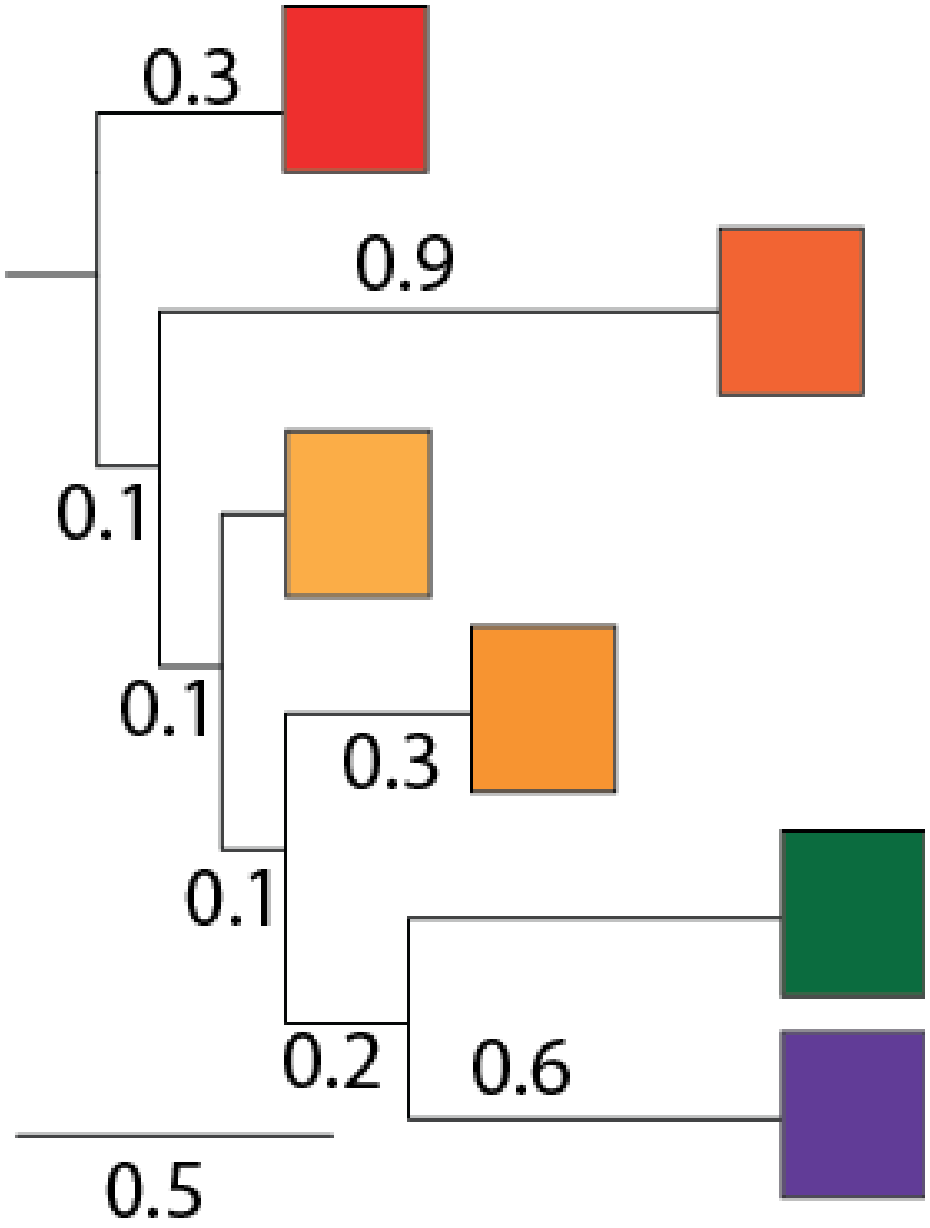
Phylodynamics



The rapid evolution of many pathogens, particularly RNA viruses, means that their evolution and ecology happen at the same time! By combining these concepts, we get better conclusions about our data.

- All observations (i.e. viruses isolated during an outbreak) are related by common descent
- Evolution is bifurcating (Tree-like)
- small genetic distances = closely related

Dimensions



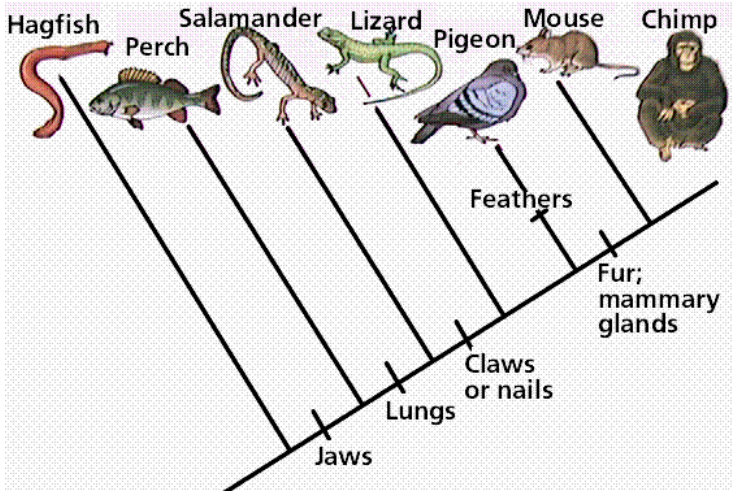
Vertical dimension has no meaning

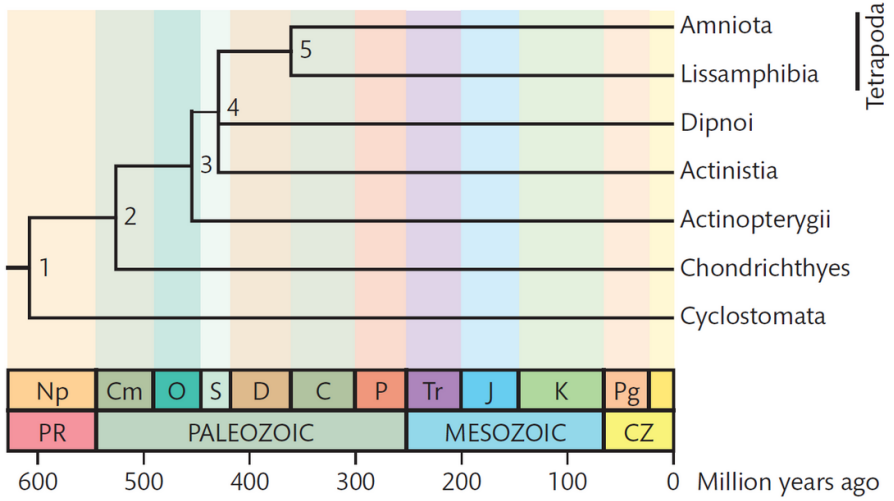
Horizontal dimension gives the amount of genetic change

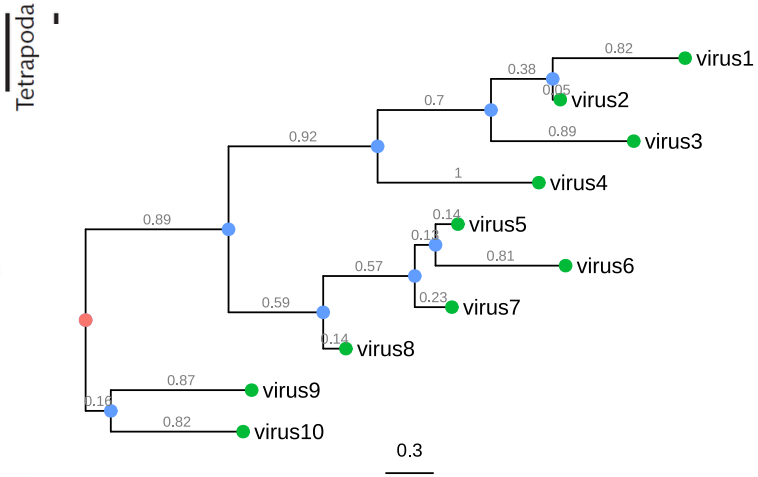
Horizontal lines show evolutionary changes in lineages over time

Scale bar shows the length of branch equivalent to 0.5 nucleotide substitutions/site (i.e., how distantly related they are)

Types of phylogenetic trees







Cladogram:

- No scale axis
- Sometimes includes phenotypic changes
- Used only for large phylogenies and relationships

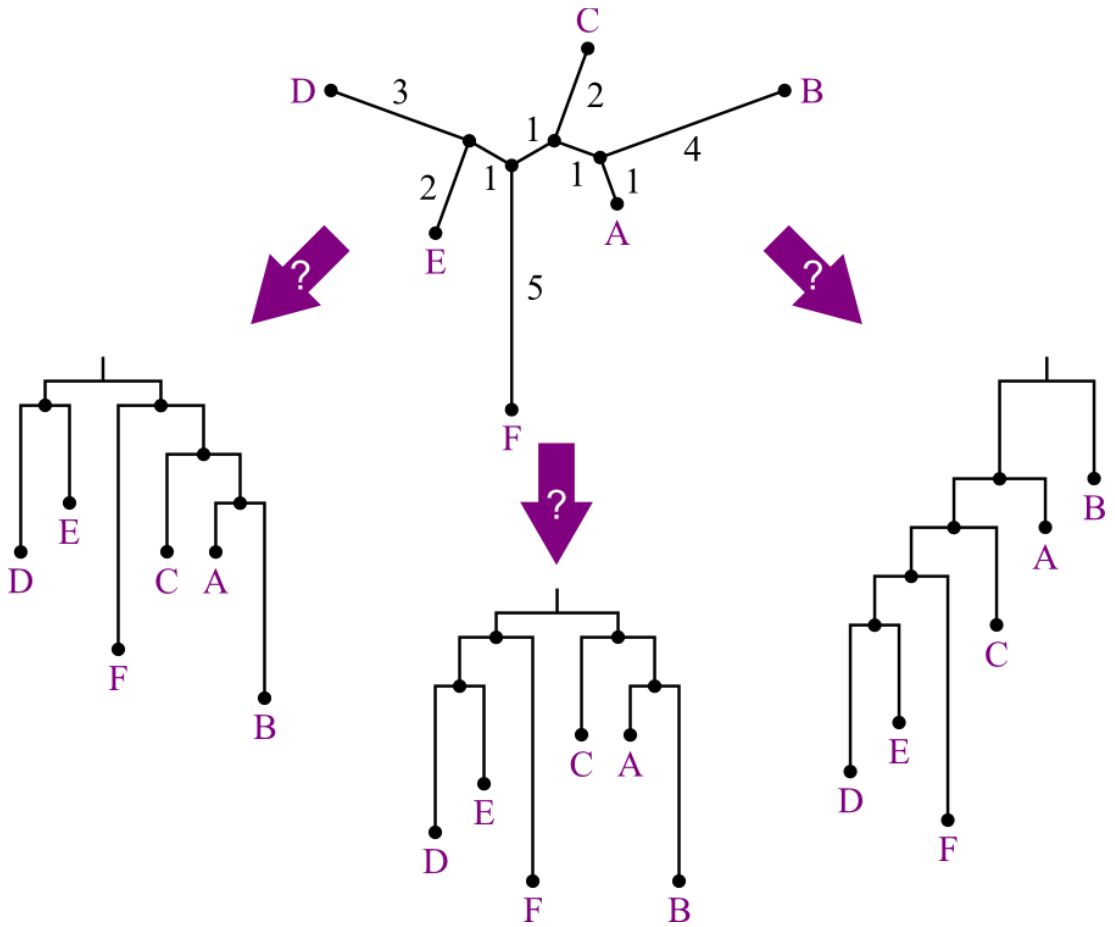
Time-scaled phylogenetic tree:

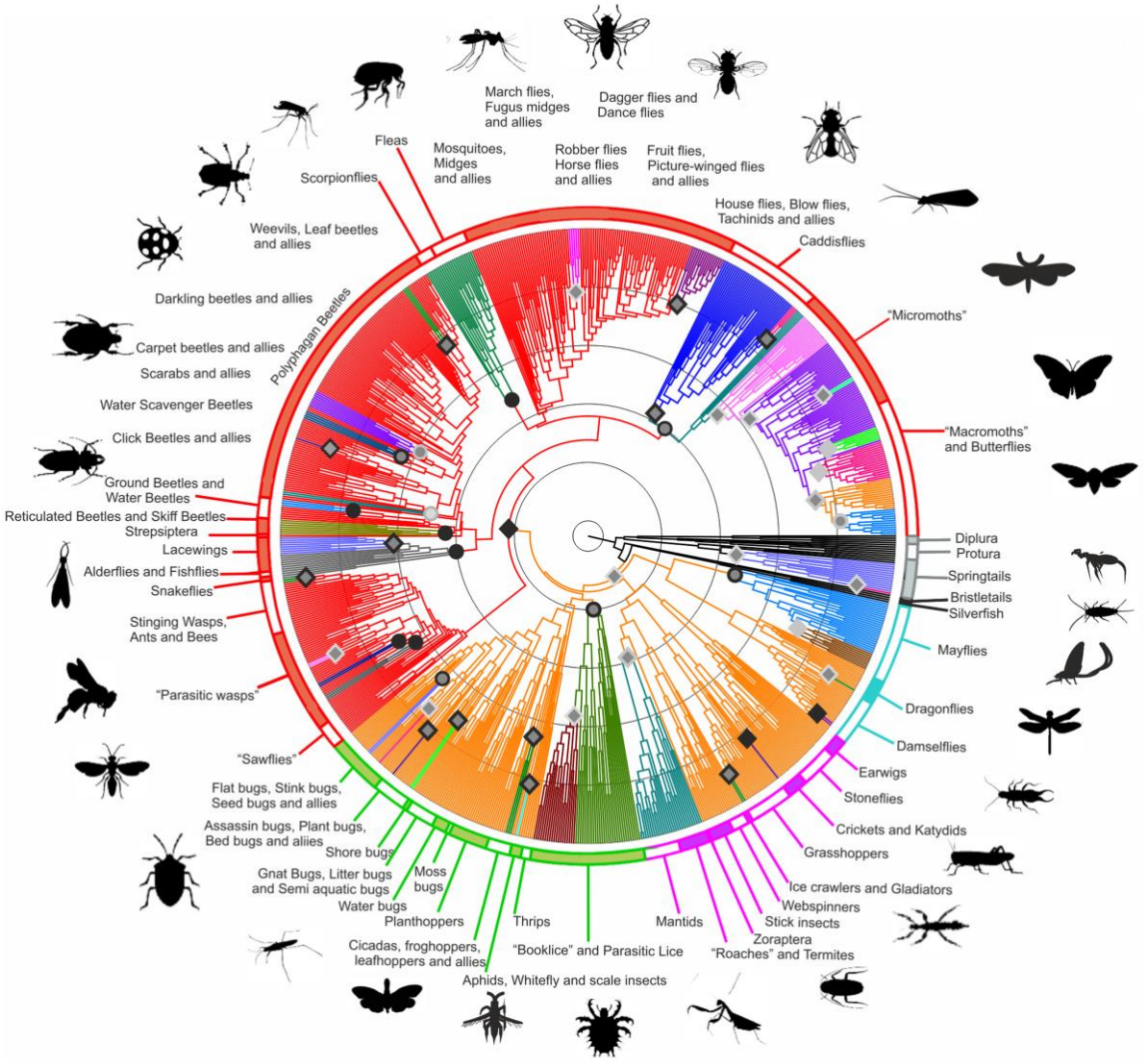
- Scale provided in years of evolution
- Built using a substitution tree + an evolutionary model
- Can differ from substitution tree

Substitution phylogenetic tree:

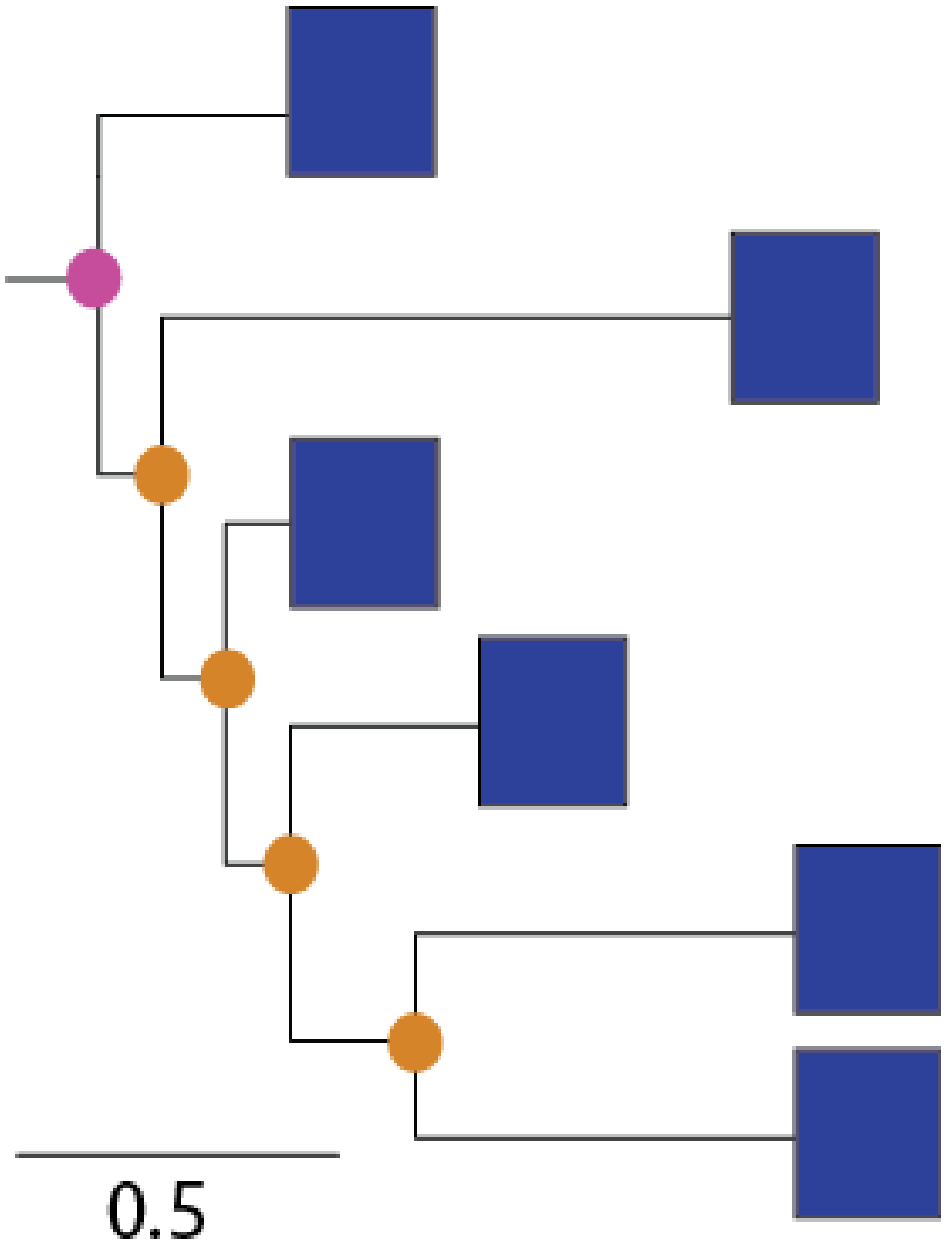
- Scale provided in number of substitutions or proportion of genome changed
- The most common tree you will see

Ways to visualize your trees





Structure

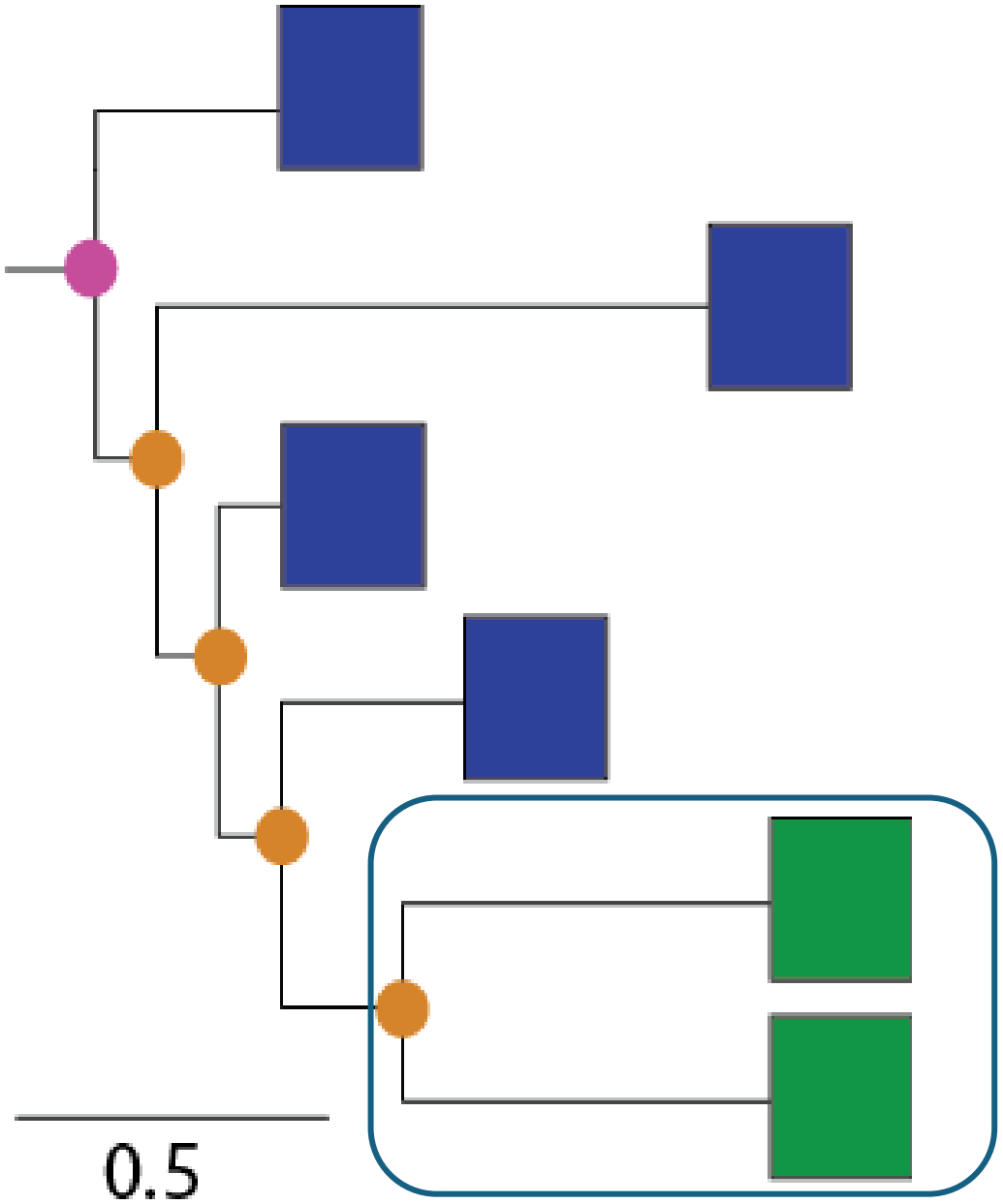


Tip/Taxon/Virus/Leaf
-Sequenced Viruses
-we also know date and location of isolation, host and clinical features

Internal Node
Our theoretical common Ancestor

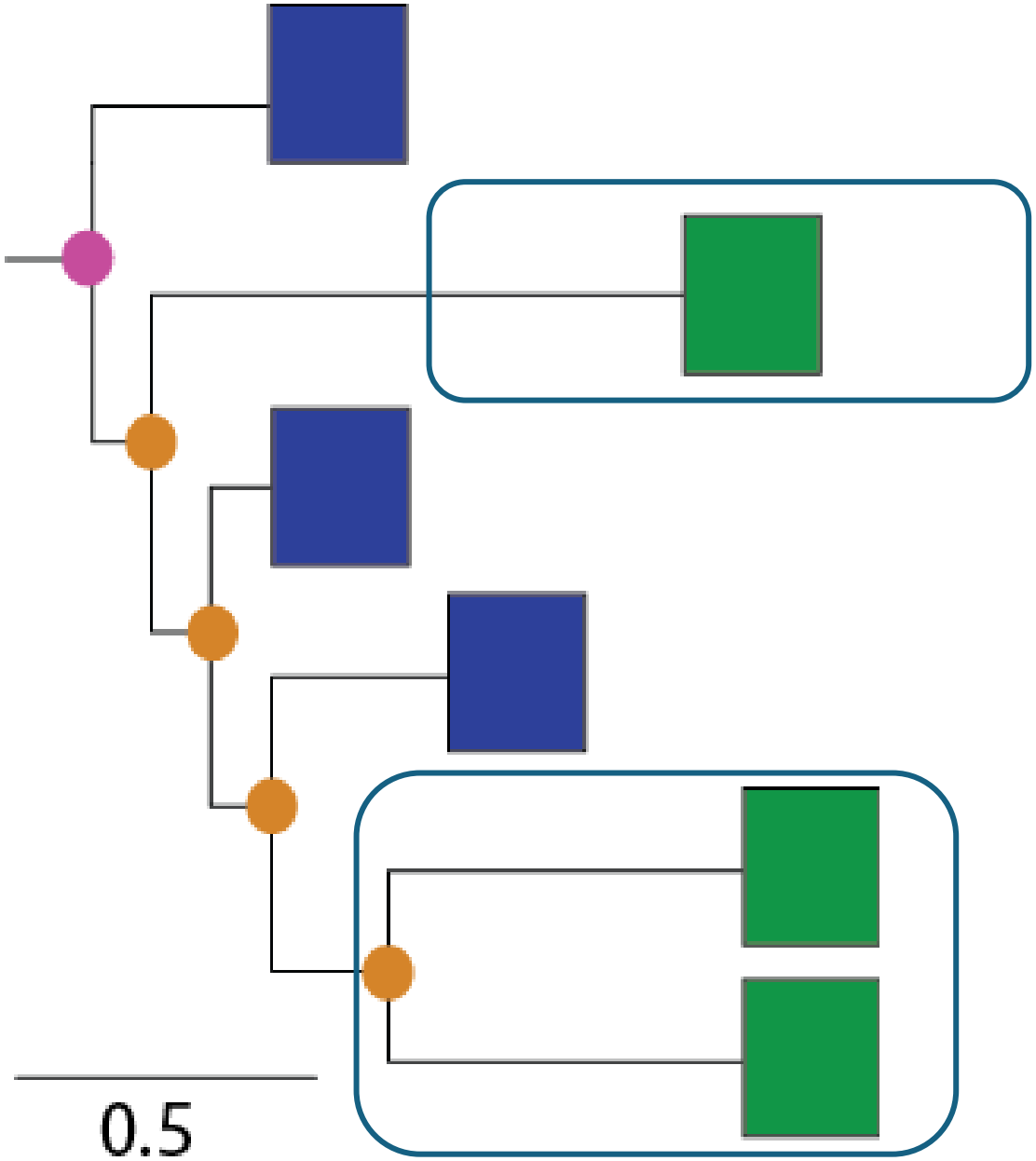
Root – provides an order for a sequence of events leading to observation (sequenced virus)

Terminology



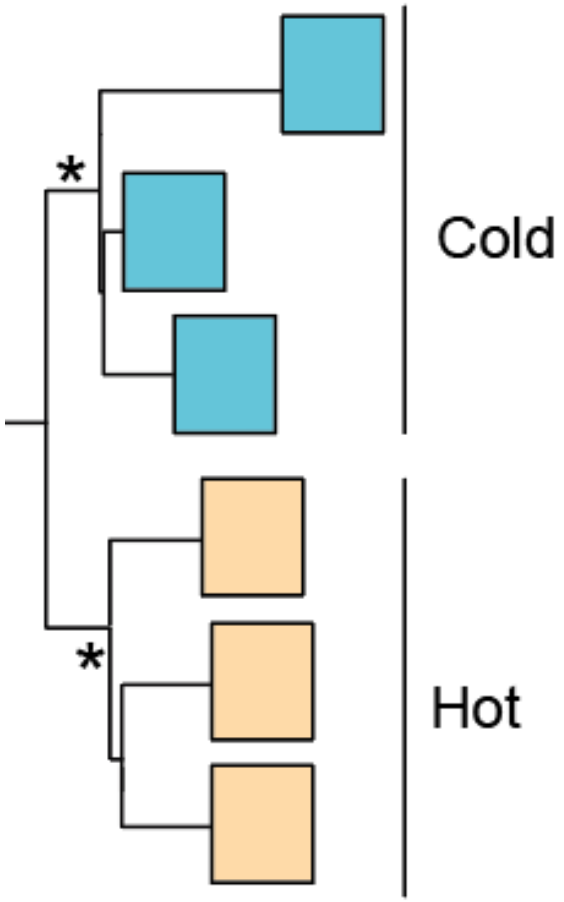
Monophyletic group (Clade) = all members share common ancestry

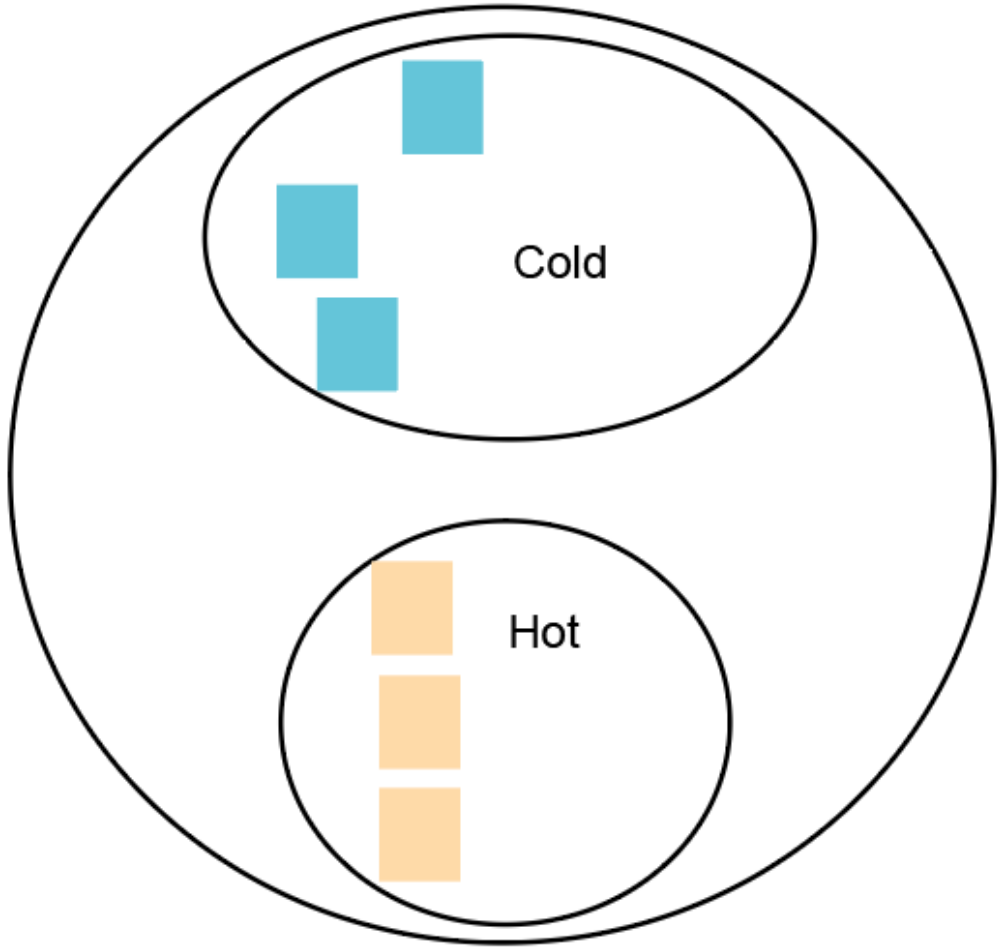
Terminology



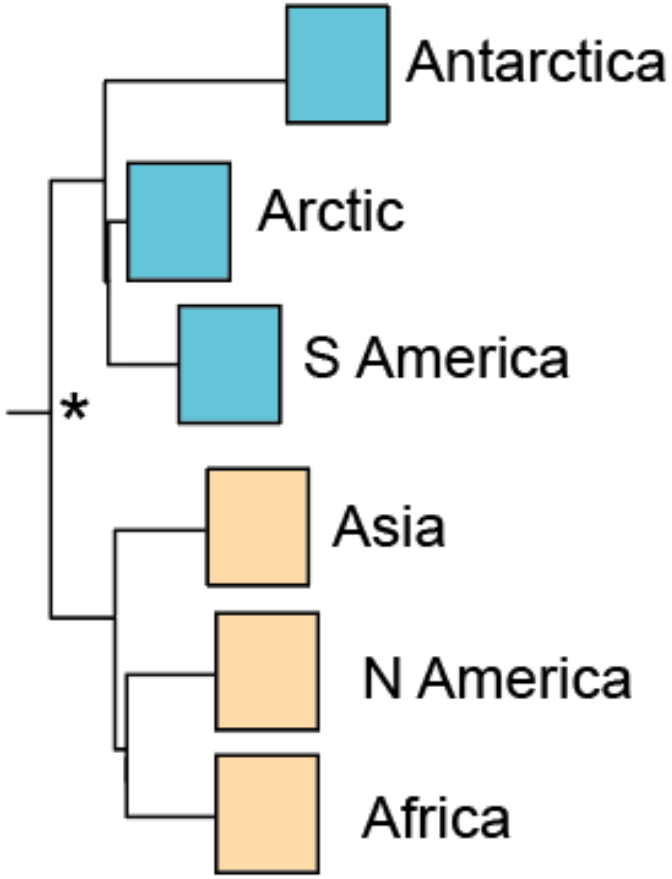
Paraphyletic = a group of organisms descended from a common ancestor, but not including all the descendant groups

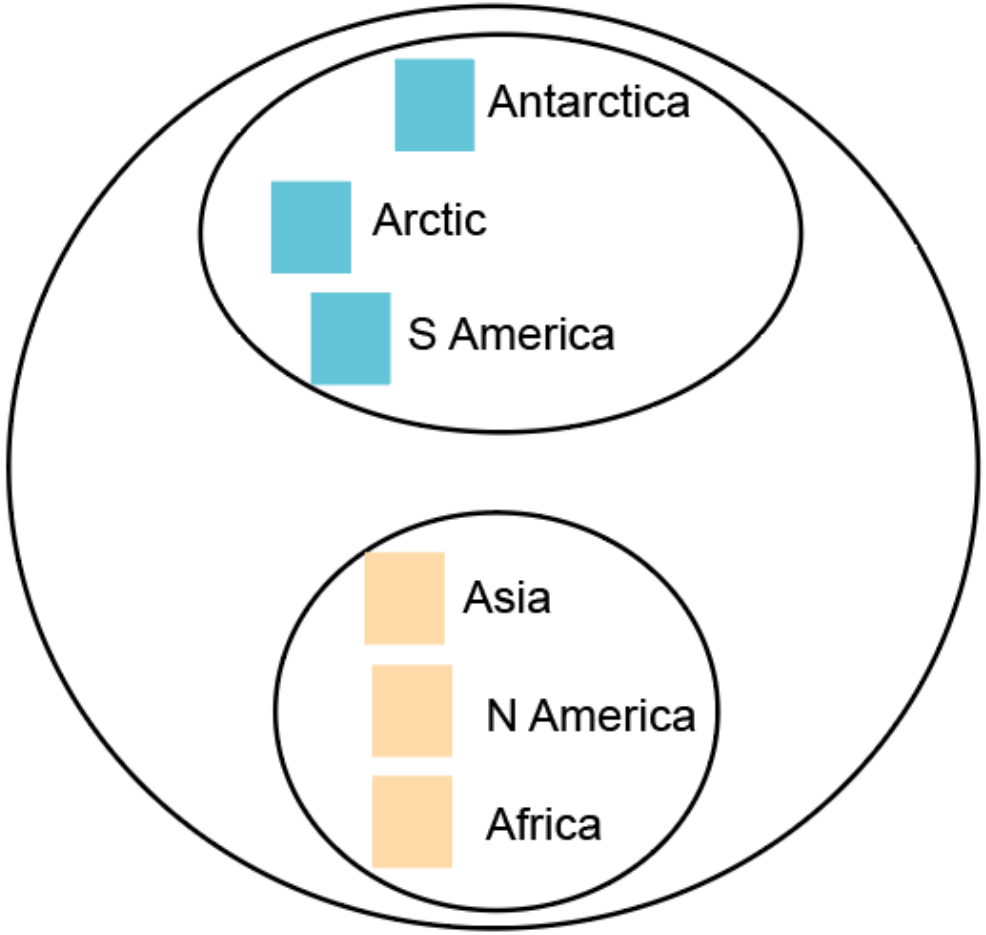
Reading a tree: Relationships



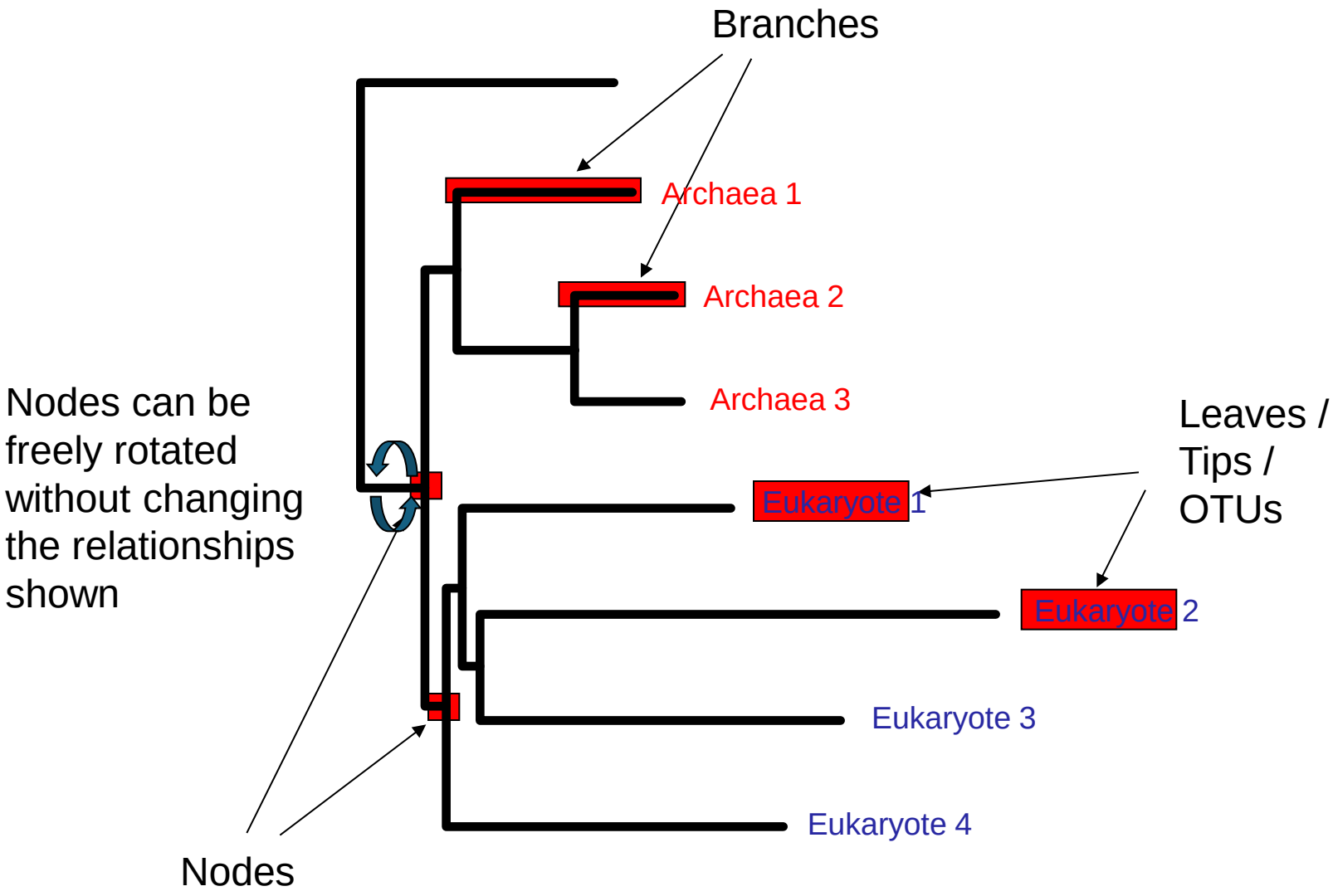


Reading a tree: Relationships



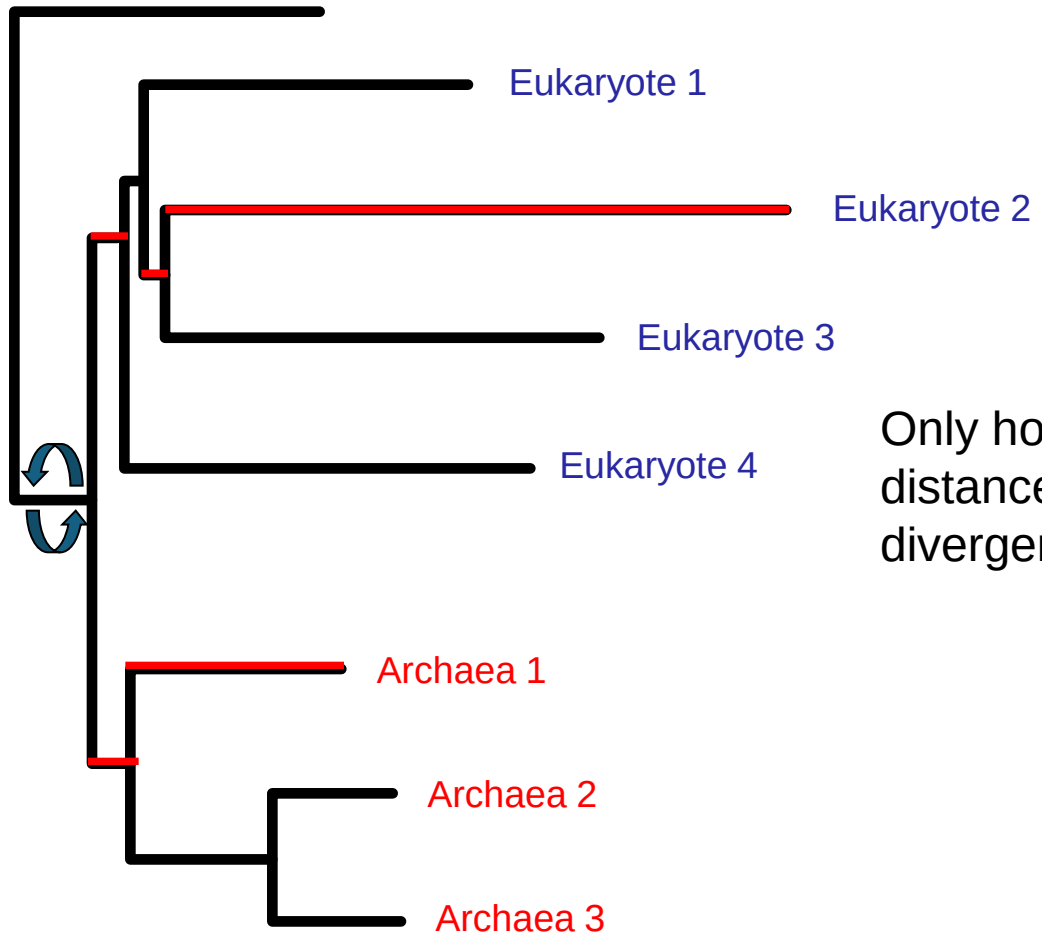


Some Tree Terms and Facts



Some Tree Terms and Facts

Nodes can be freely rotated without changing the relationships shown

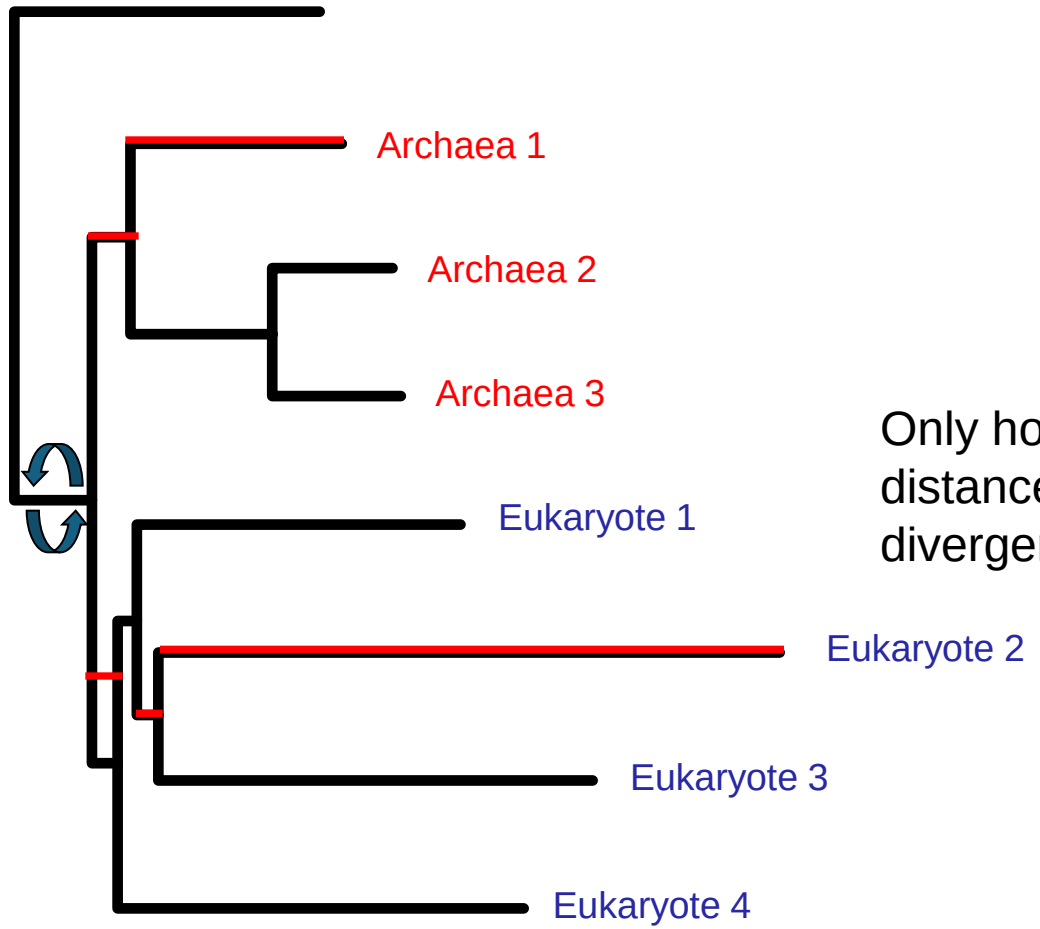


Only horizontal distances indicate divergence

Total distance =

Some Tree Terms and Facts

Nodes can be freely rotated without changing the relationships shown



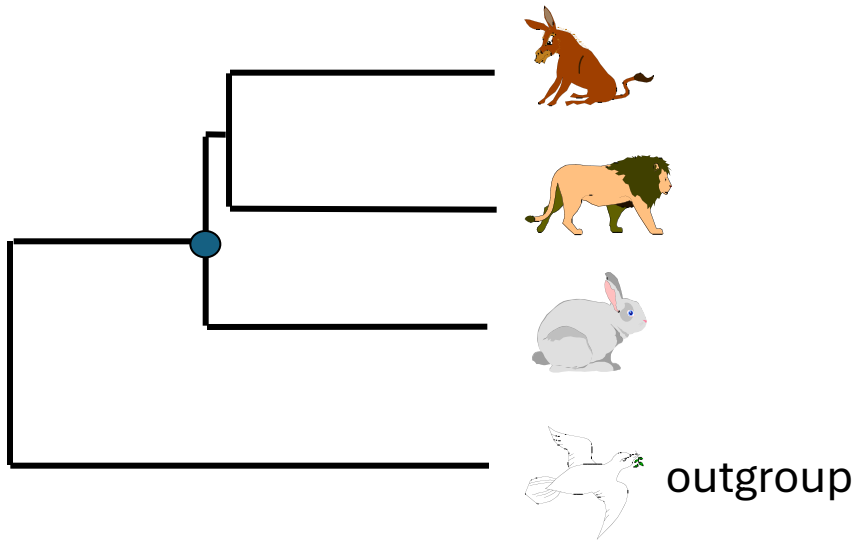
Only horizontal distances indicate divergence

Total distance =

There are two major ways to root trees:

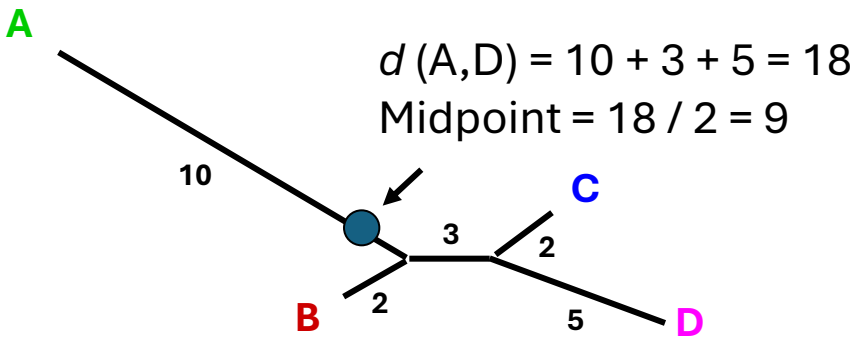
By outgroup:

Uses taxa (the "outgroup") that are known to fall outside of the group of interest (the "ingroup"). Requires some prior knowledge about the relationships among the taxa. The outgroup can either be species (*e.g.*, birds to root a mammalian tree) or previous gene duplicates (*e.g.*, α-globins to root β-globins).



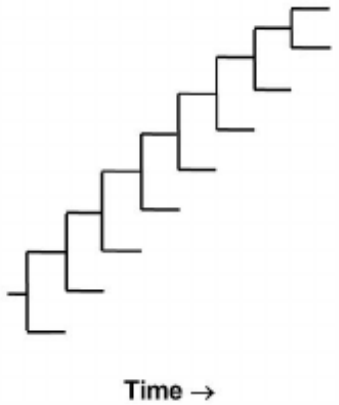
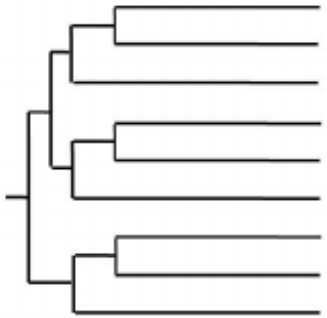
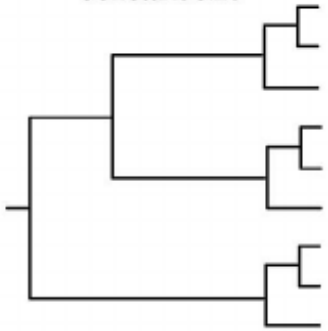
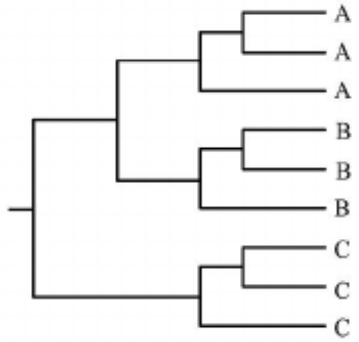
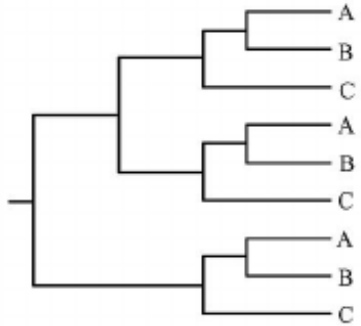
By midpoint or distance:

Roots the tree at the midway point between the two most distant taxa in the tree, as determined by branch lengths. Assumes that the taxa are evolving in a clock-like manner. This assumption is built into some of the distance-based tree building methods.

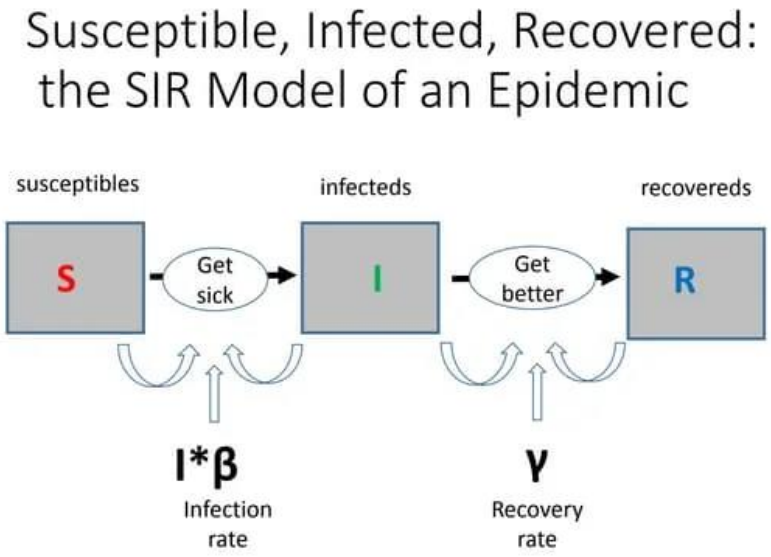


Evolution provides another tool for investigating epidemiological trends

Tree shape matters!

	Continual Immune Selection	Weak or Absent Immune Selection	
		Tree shape controlled by non-selective population dynamic processes	
		Population size dynamics	Spatial dynamics
Idealized Phylogeny Shapes		**Exponential growth**  **Constant size** 	**Strong spatial structure**  **Weak spatial structure** 
Examples	Human influenza A virus intra-host HIV	inter-host HIV inter-host HCV	Measles, rabies inter-host HIV
Tree Inferences	Detection of antigenic escape mutations	Estimation of population growth rates	Estimation of population migration rates

Evolution as a way to estimate epidemiological patterns

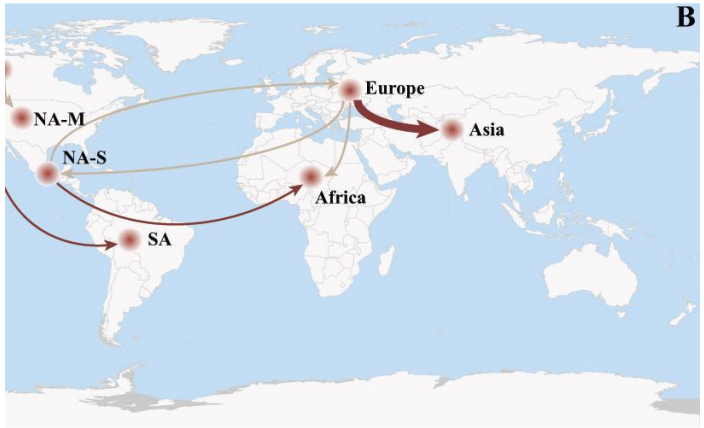


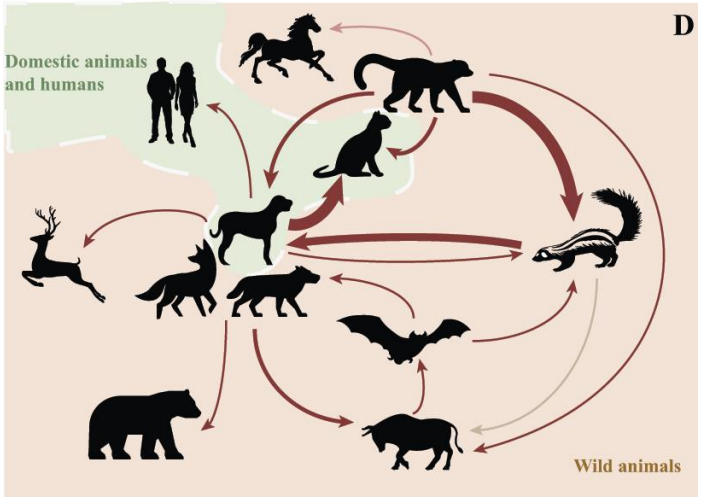
- We can use sequence data to infer the size of our compartments in SER models
- By looking at the number of mutations that have accumulated, we can guess at the number of individuals impacted

Evolution as a tool for identifying the origins and movement of a pathogen

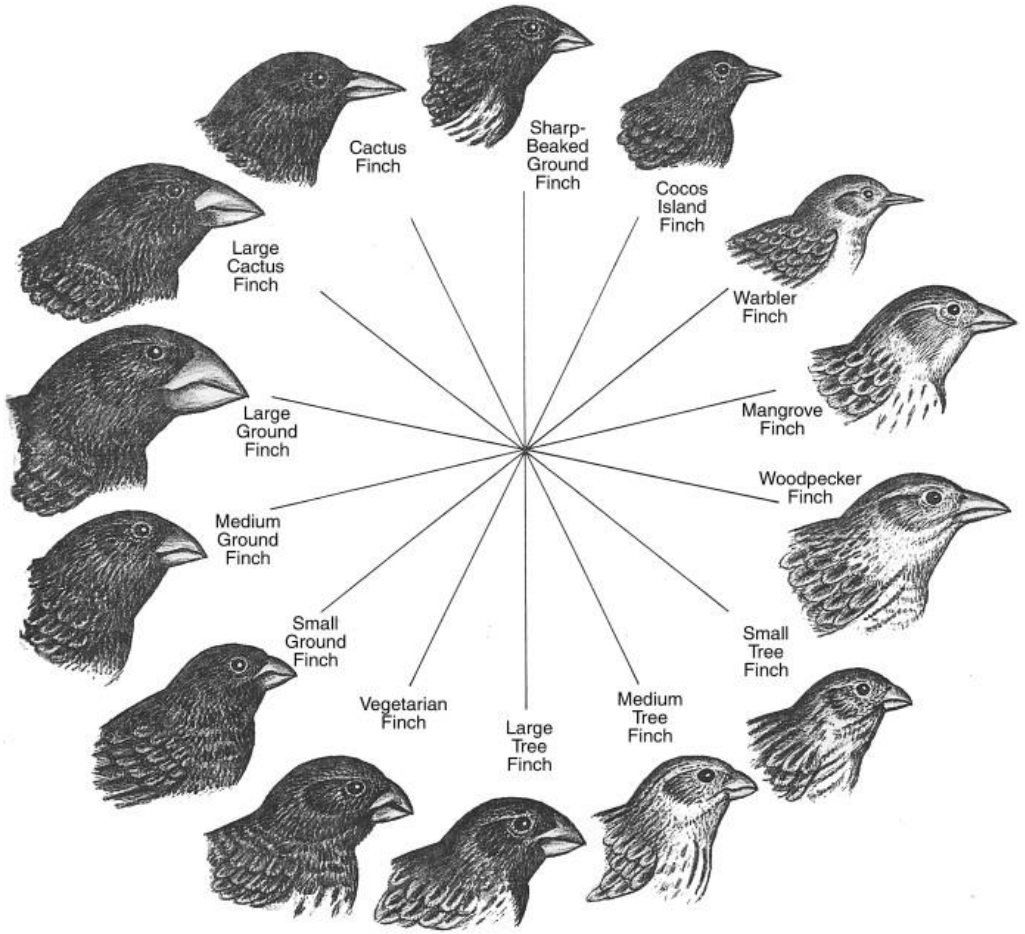
- We can investigate the likely hosts for a given pathogen, as we saw in the previous lecture

- We can investigate things like likely origins of an outbreak (e.g., investigating the origin of SARS-CoV-2)





Evolution as a tool to investigate evolutionary pressures on a pathogen



- When there is a high degree of selection on a pathogen, mutations fix at a faster rate
- This can help us test hypotheses around what segments of a protein are good vaccine targets, among other interesting questions!

Evolution as corroborating evidence for contact tracing

- We do not always have the ability to contact trace
- As mutations fix and move along to new hosts, we can use these to identify who likely infected whom
- Often, we use things like venue affiliation, phylogenetic trees, and some contact tracing to construct the likely transmission tree