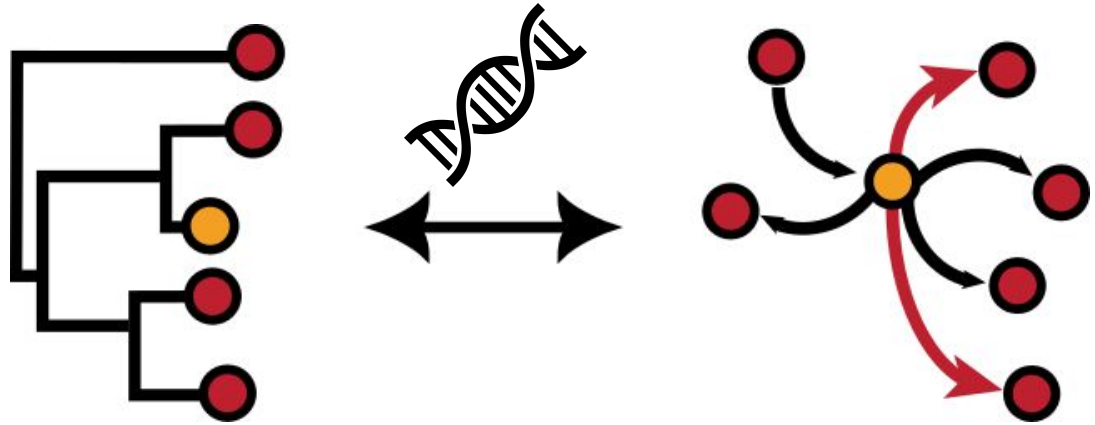
EPAH6052: Pathogen Genomic Epidemiology

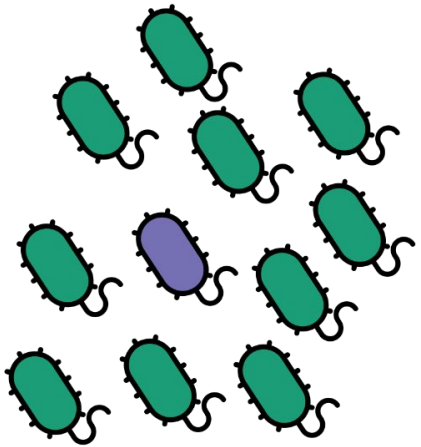


Finlay Maguire
Dalhousie University
Shared Hospital Laboratory
Public Health Alliance for Genomic Epidemiology

Outcomes

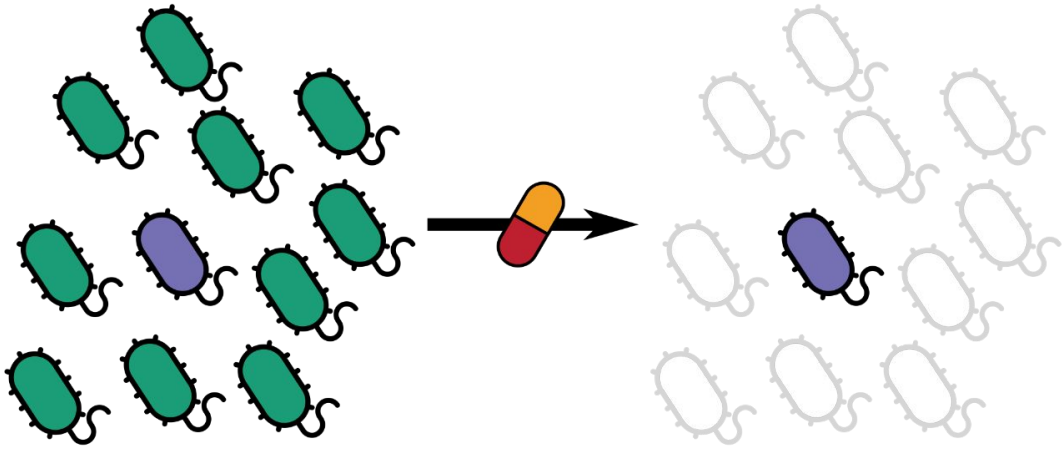
- Explain how evolution and infectious disease epidemiology are connected
- Identify what additional information genomic data can provide
- Give example of how genomics can be used for diagnostics
- Explain how you can infer an evolutionary tree (phylogeny)
- Provide examples of processes which determine the shape of a phylogeny
- Articulate the role of Bayesian models in genomic epidemiology
- List at least 3 ways in which a phylogeny can be used in epidemiology

Agents of infectious diseases undergo evolution

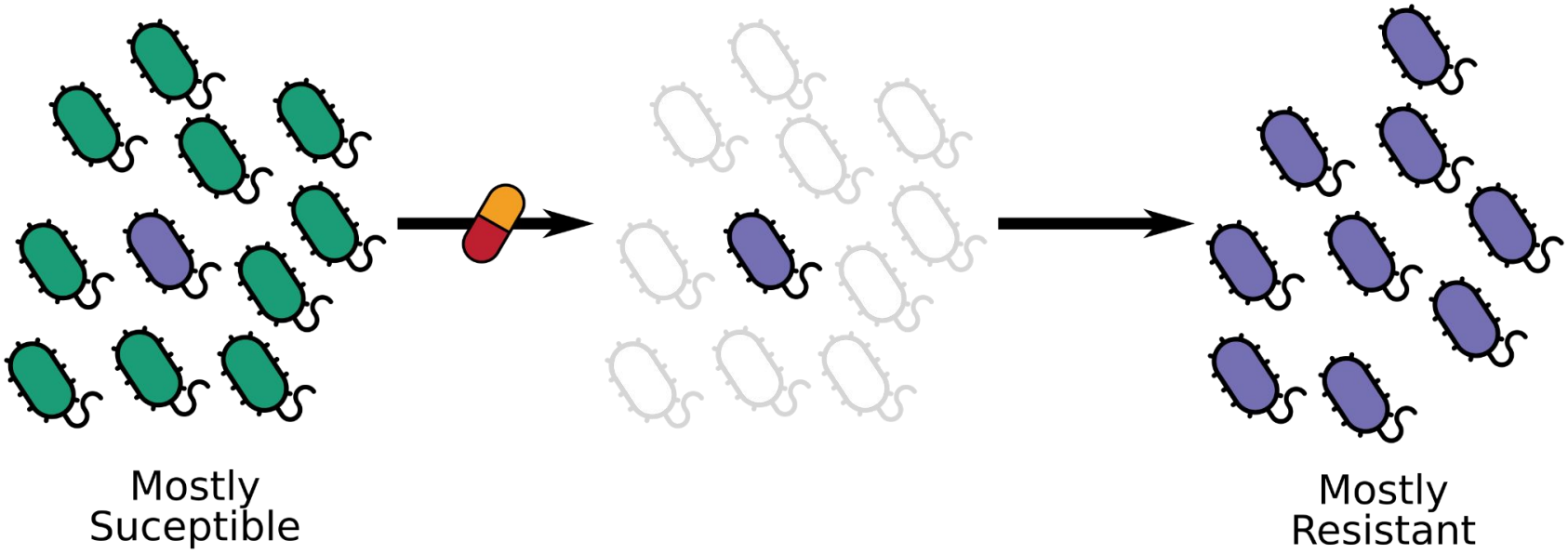


Mostly
Suceptible

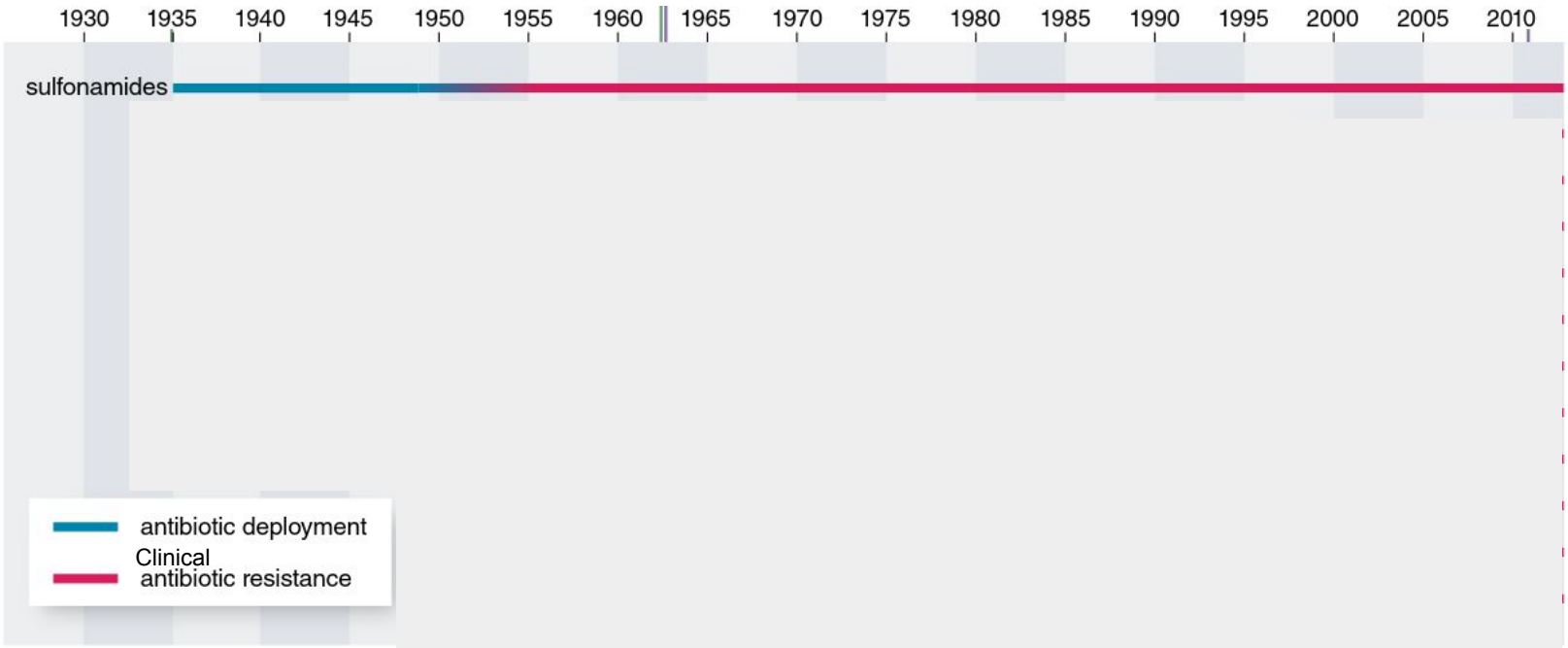
Agents of infectious diseases undergo evolution



Agents of infectious diseases undergo evolution

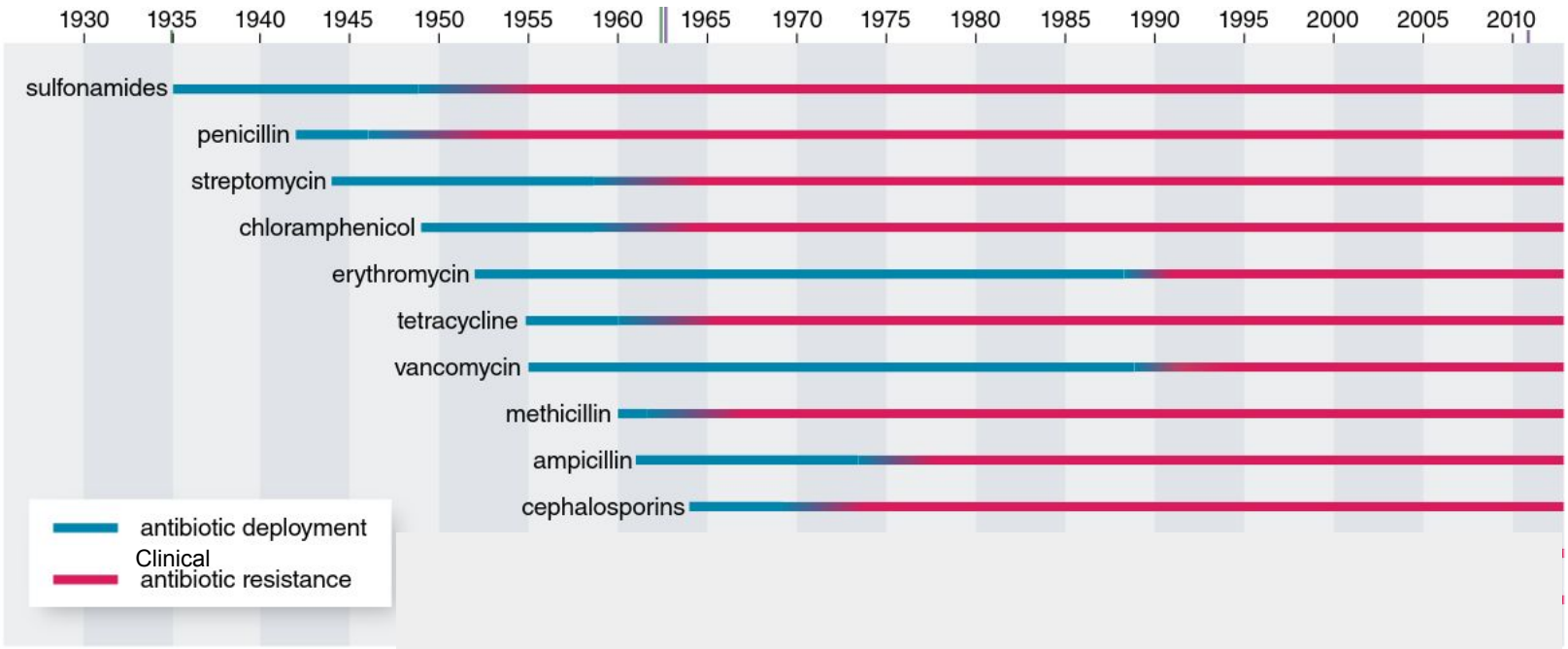


Pathogen evolution impacts treatment



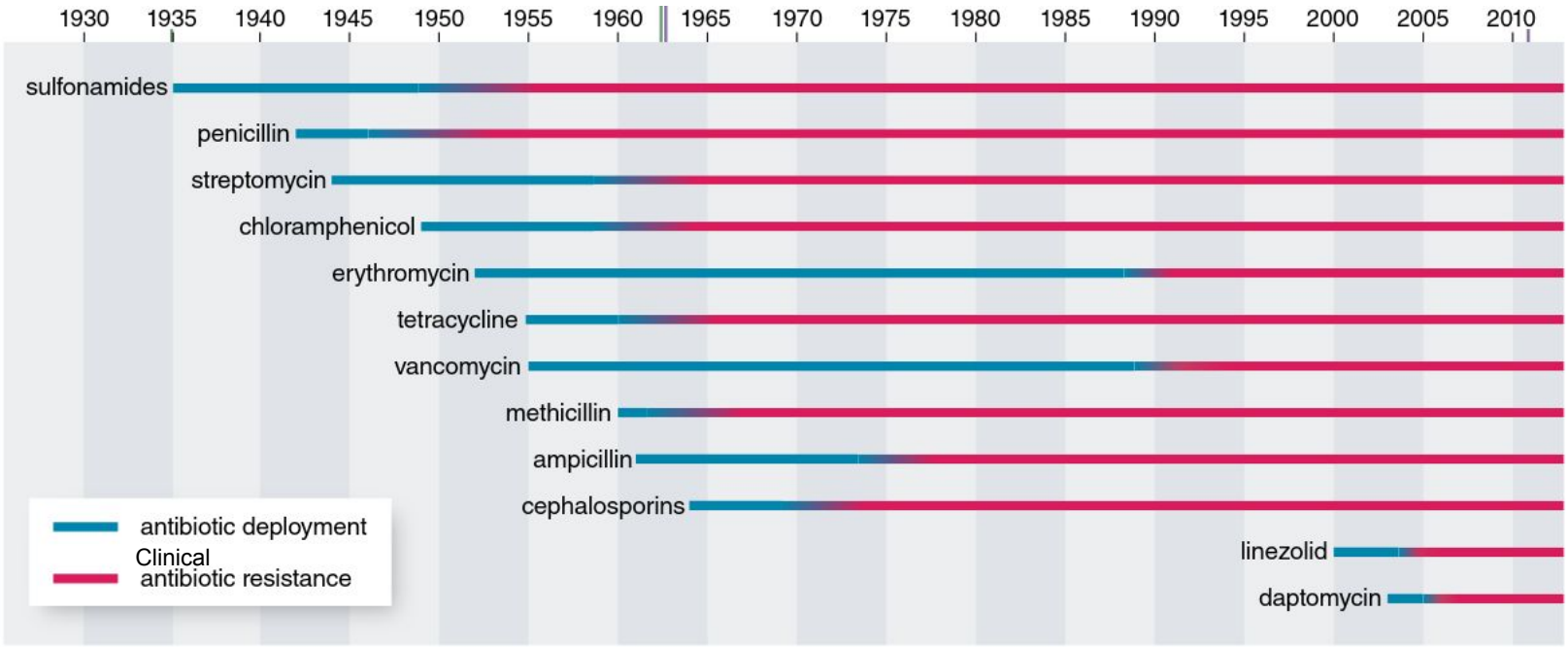
10.1511/2014.106.42

Pathogen evolution impacts treatment



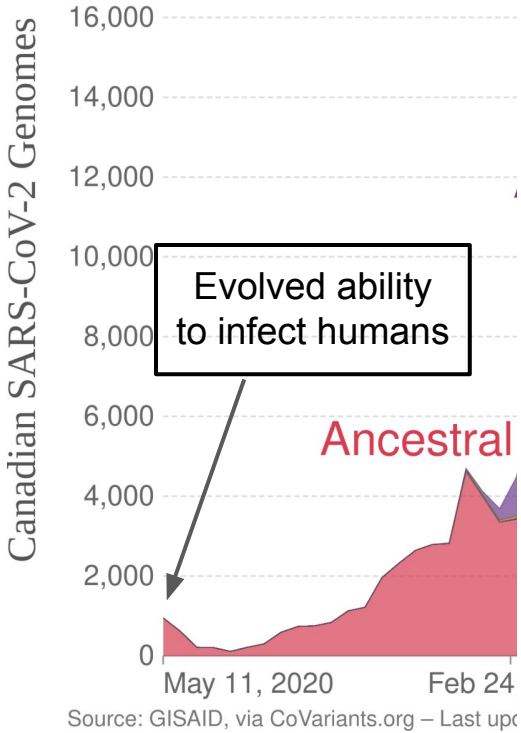
10.1511/2014.106.42

Pathogen evolution impacts treatment

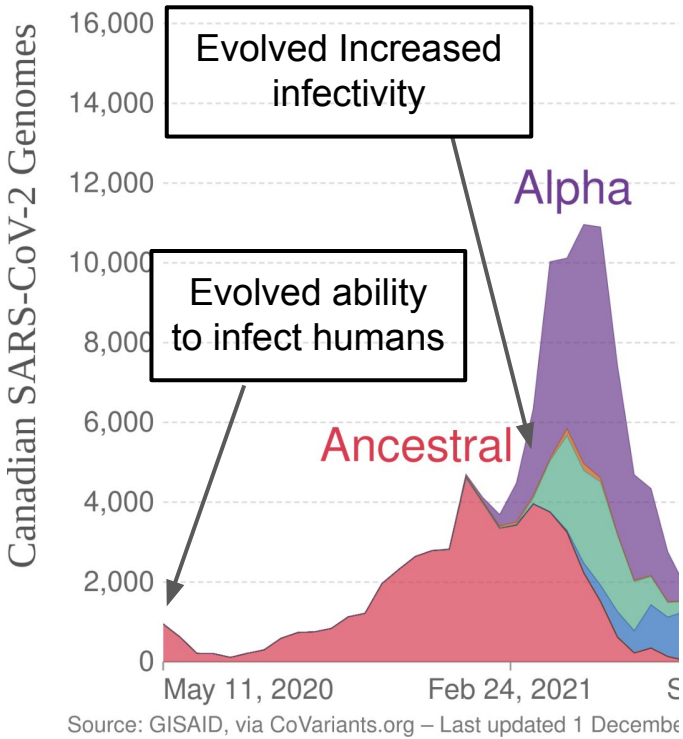


10.1511/2014.106.42

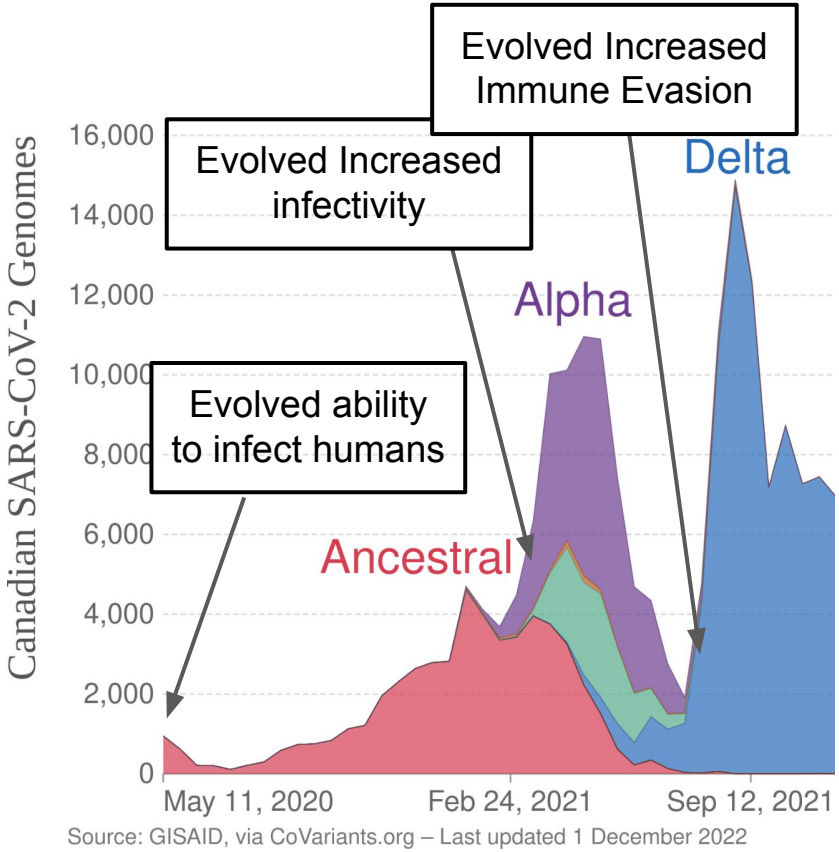
Evolution drives cases: SARS-CoV-2 waves



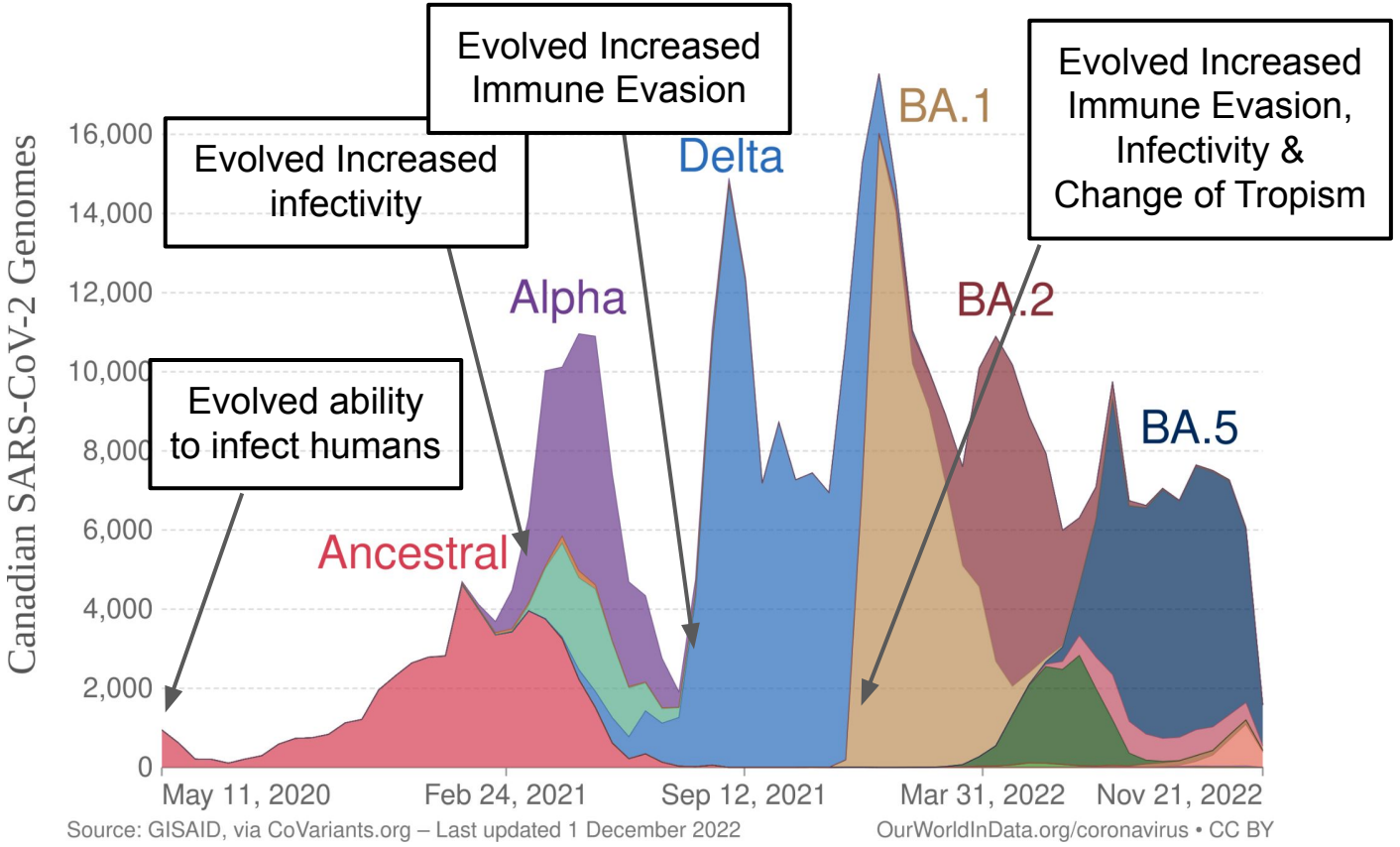
Evolution drives cases: SARS-CoV-2 waves



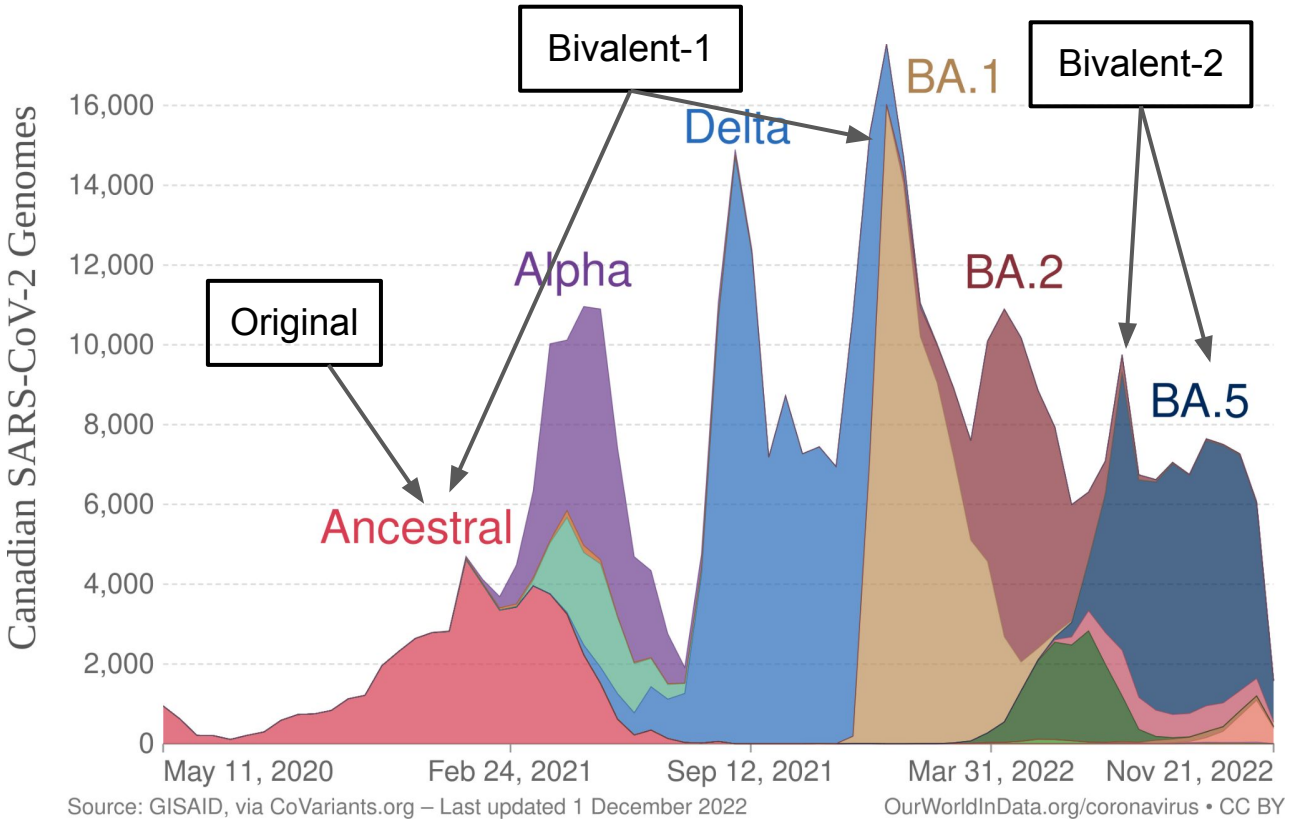
Evolution drives cases: SARS-CoV-2 waves



Evolution drives cases: SARS-CoV-2 waves

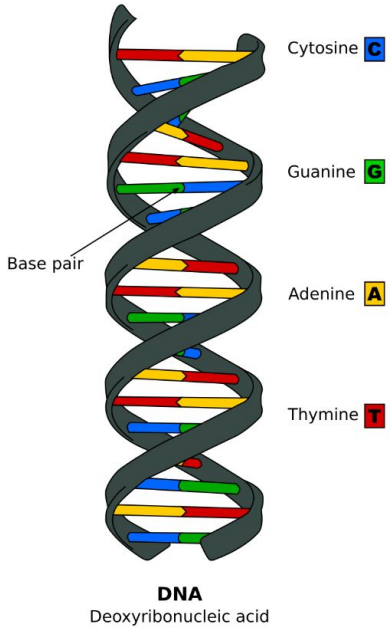


Evolution drives vaccine design/effectiveness

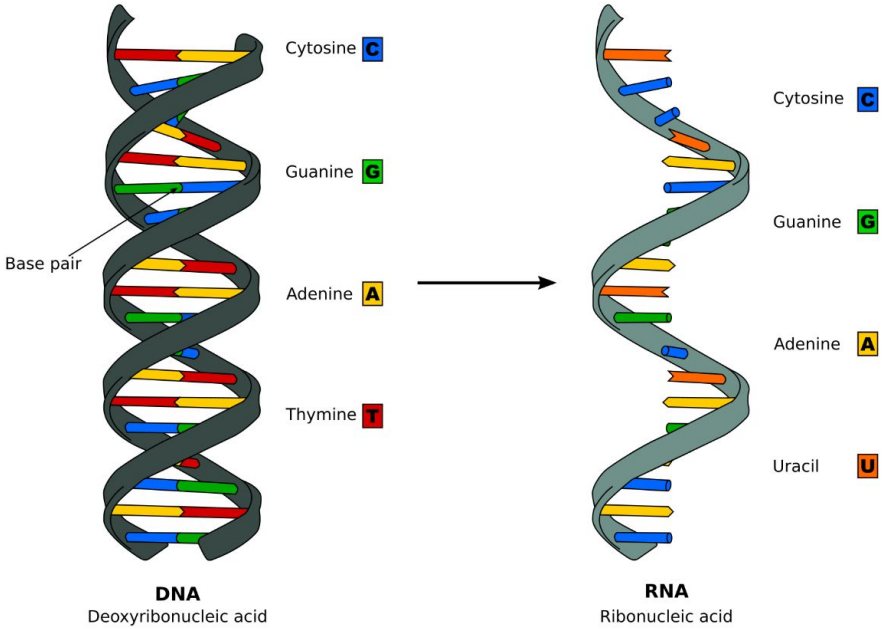


How can we monitor evolution?

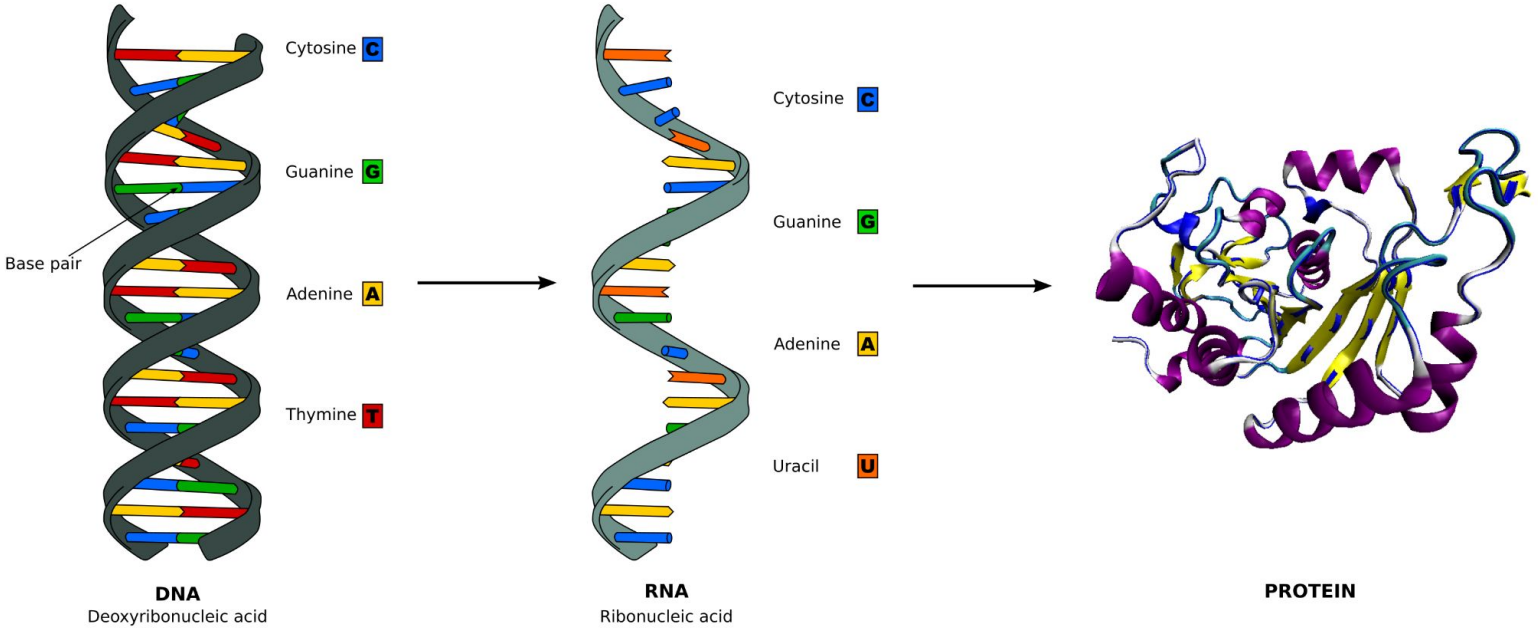
Genomes are the substrate of evolution



Genomes are the substrate of evolution

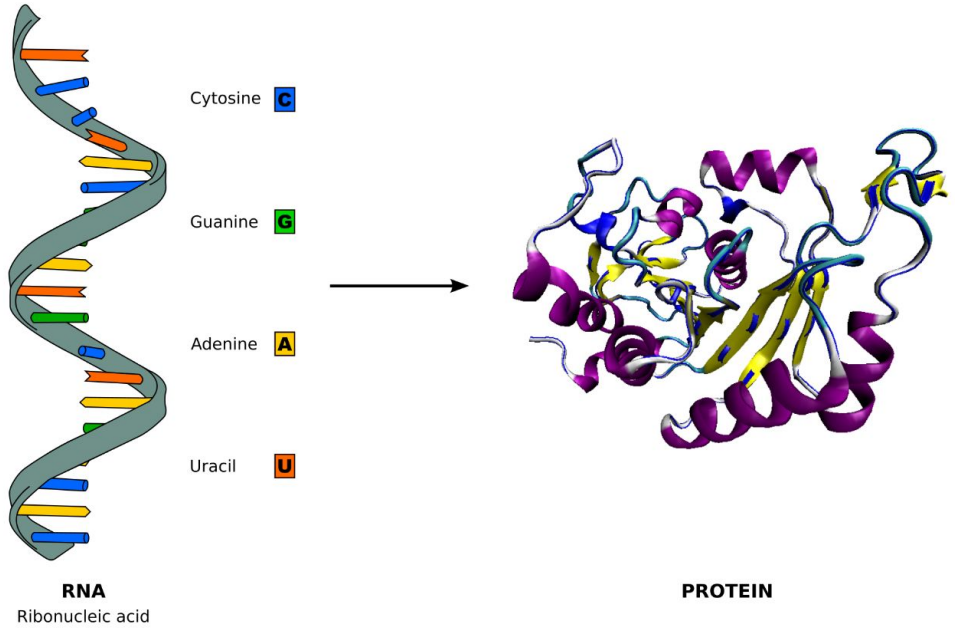


Genomes are the substrate of evolution



- DNA encodes RNA which encodes proteins

Genomes are the substrate of evolution



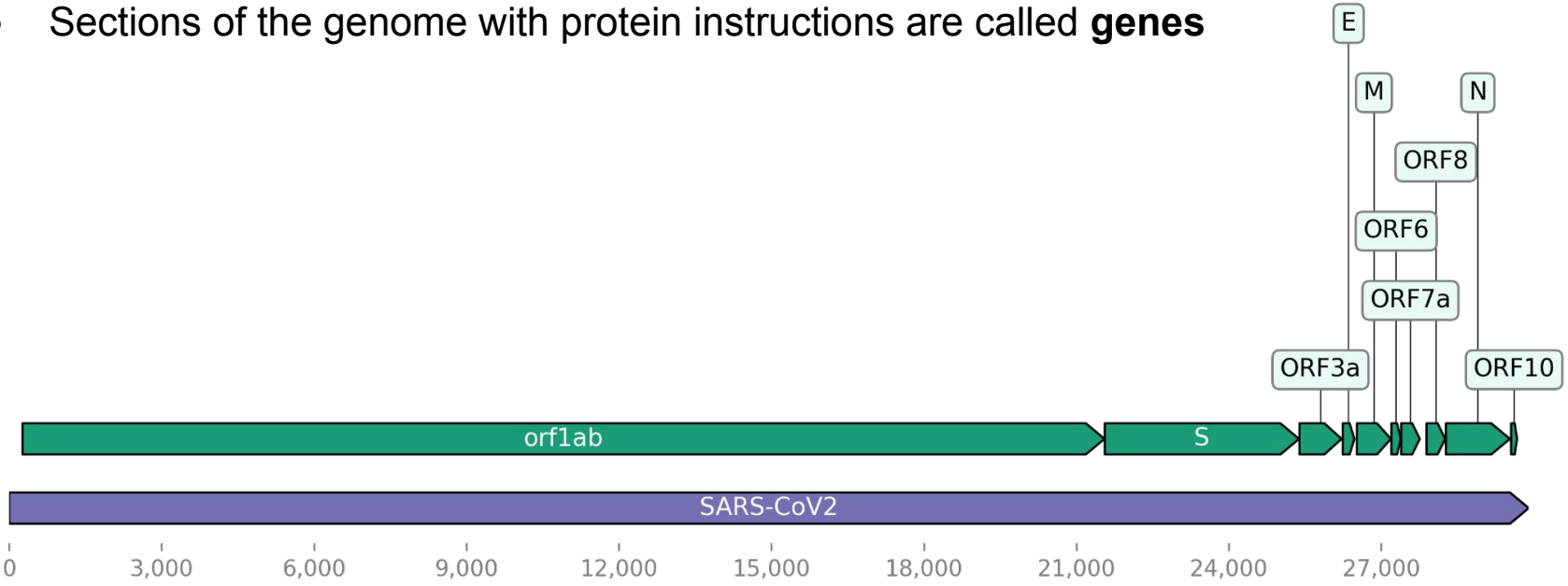
- DNA encodes RNA which encodes proteins
- Viruses like SARS-CoV-2 skip DNA

Genomes are the substrate of evolution

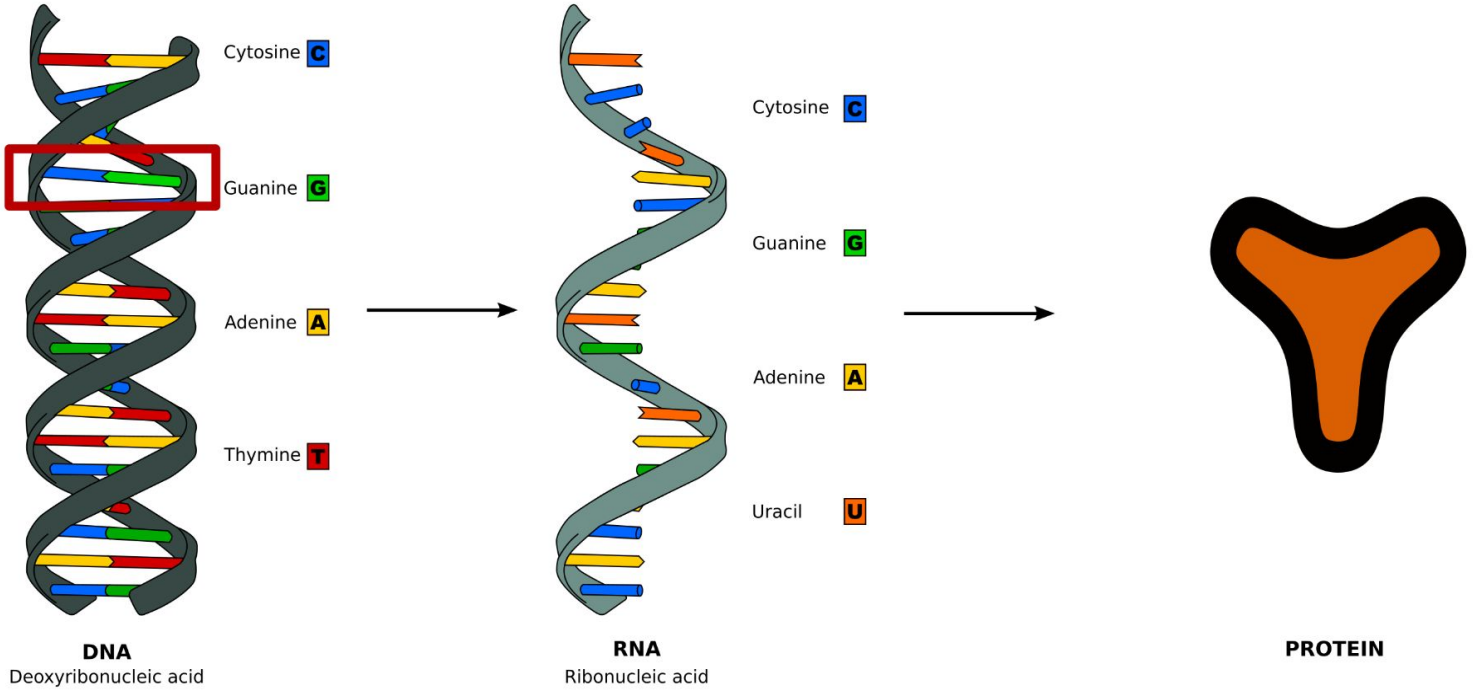
- Genomes are the complete collection of genetic instructions

Genomes are the substrate of evolution

- Genomes are the complete collection of genetic instructions
- Sections of the genome with protein instructions are called **genes**

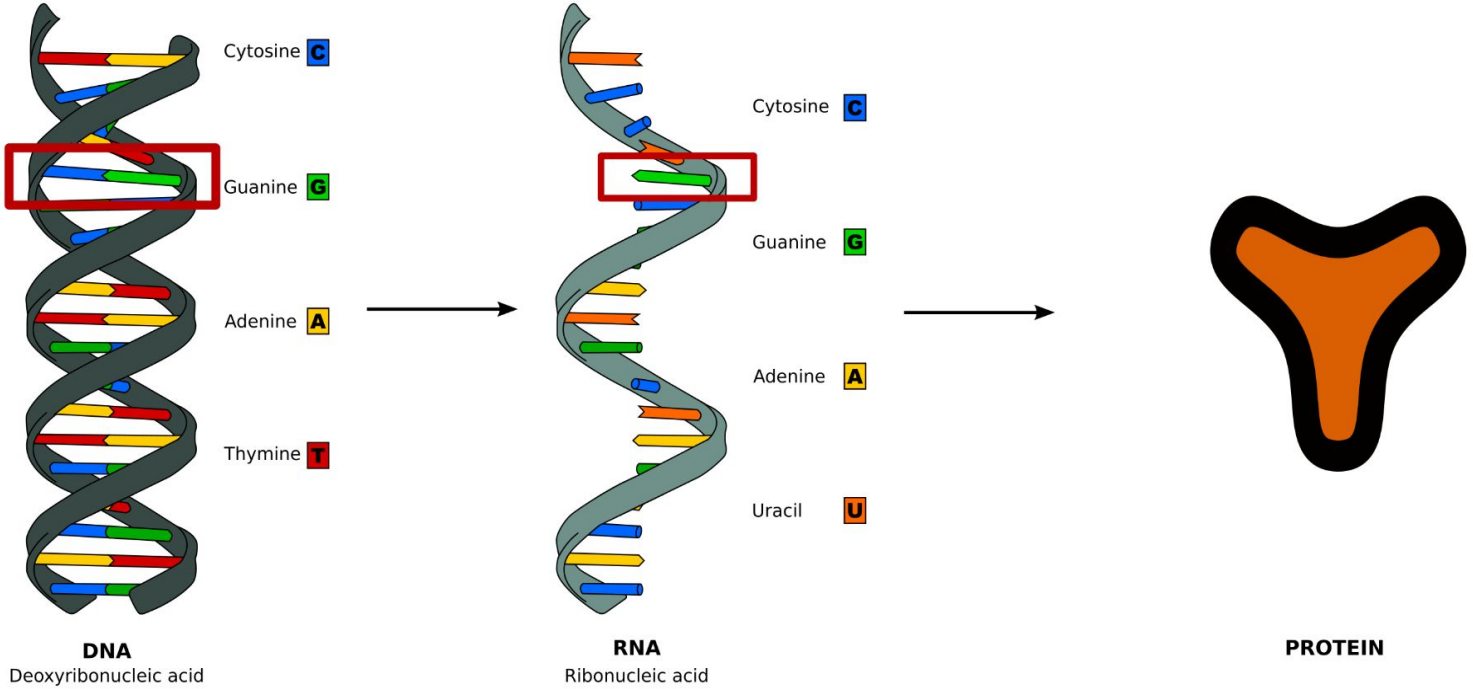


Mutations provide the variation upon which evolution acts



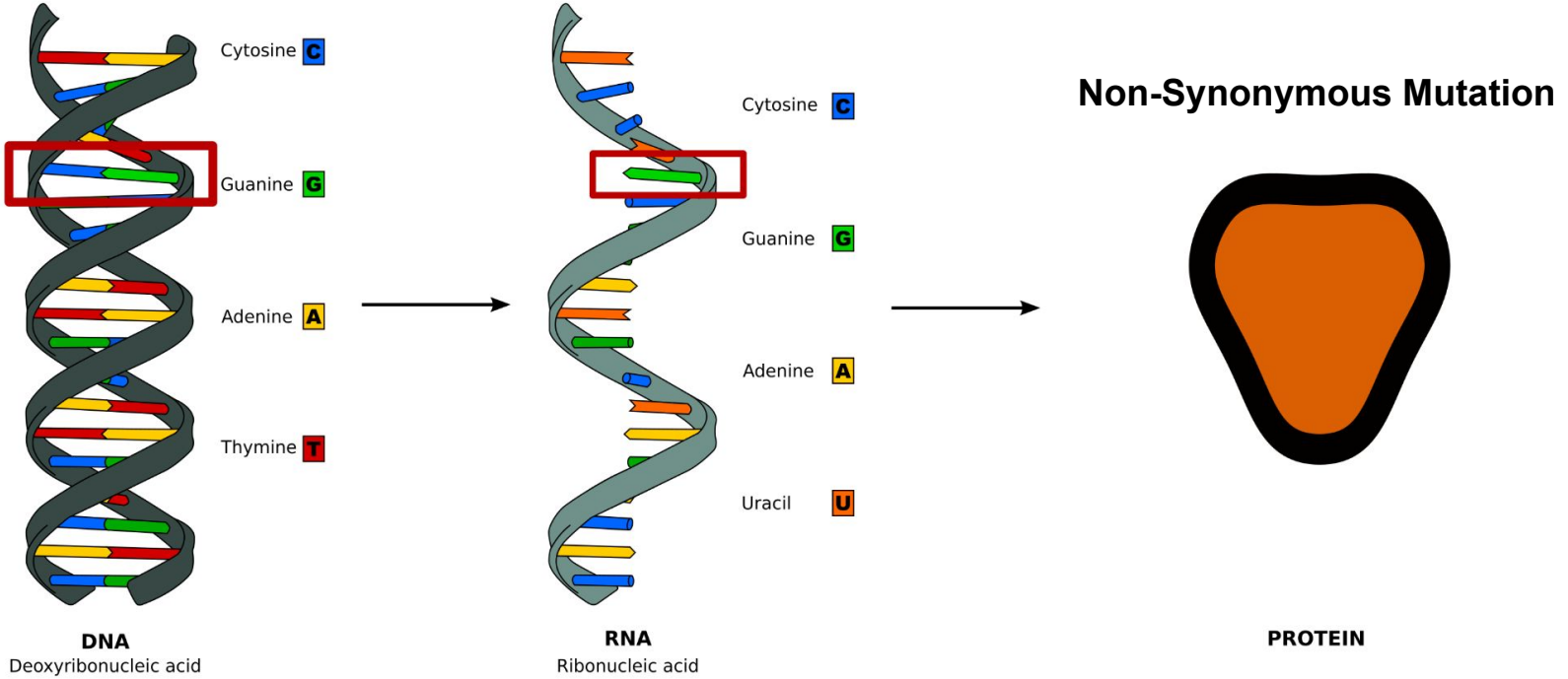
- Genome copying is error-prone

Mutations provide the variation upon which evolution acts



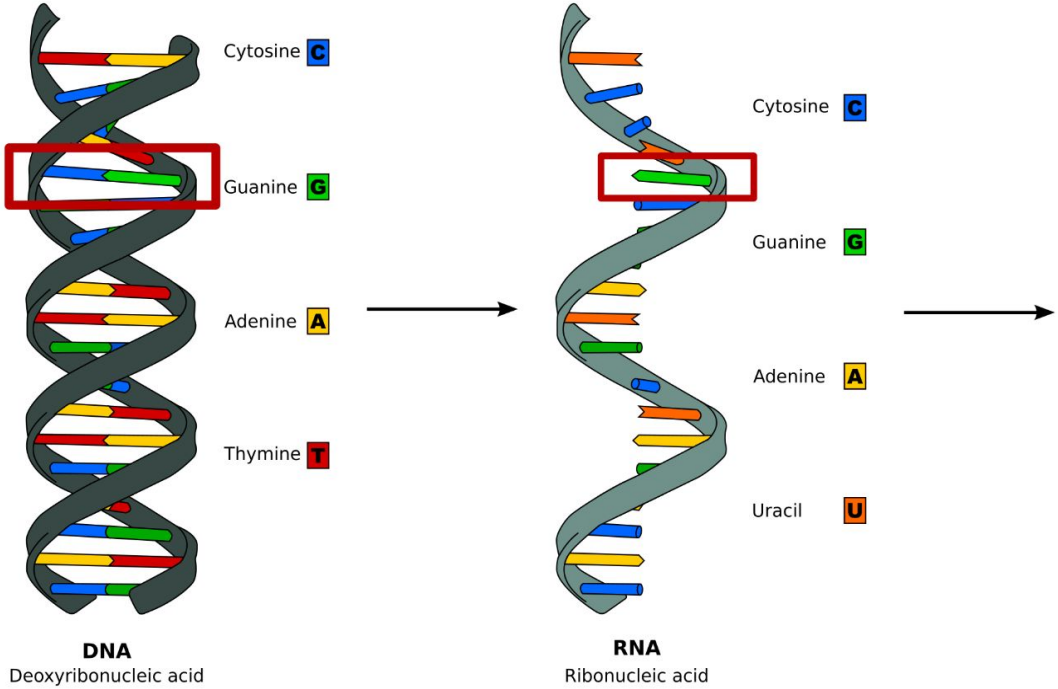
- Genome copying is error-prone
- Errors are called mutations

Mutations provide the variation upon which evolution acts



- Genome copying is error-prone
- Errors are called mutations
- Mutations can change protein sequence

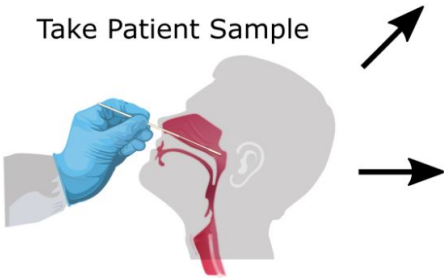
Mutations provide the variation upon which evolution acts



- Genome copying is error-prone
- Errors are called mutations
- Mutations can change protein sequence - *but don't always*

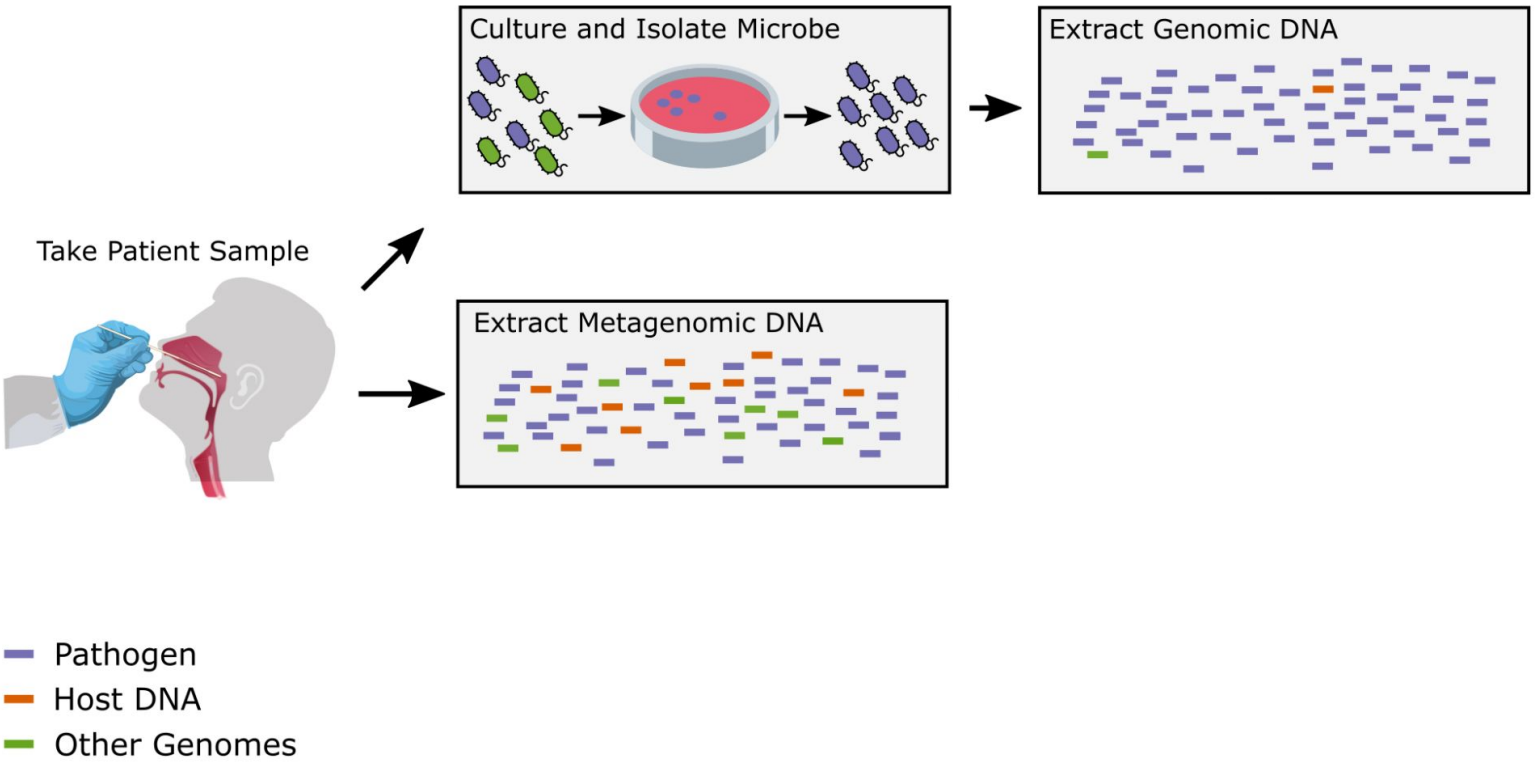
So, how do we get genomes?

Sequencing Pathogen Genomes

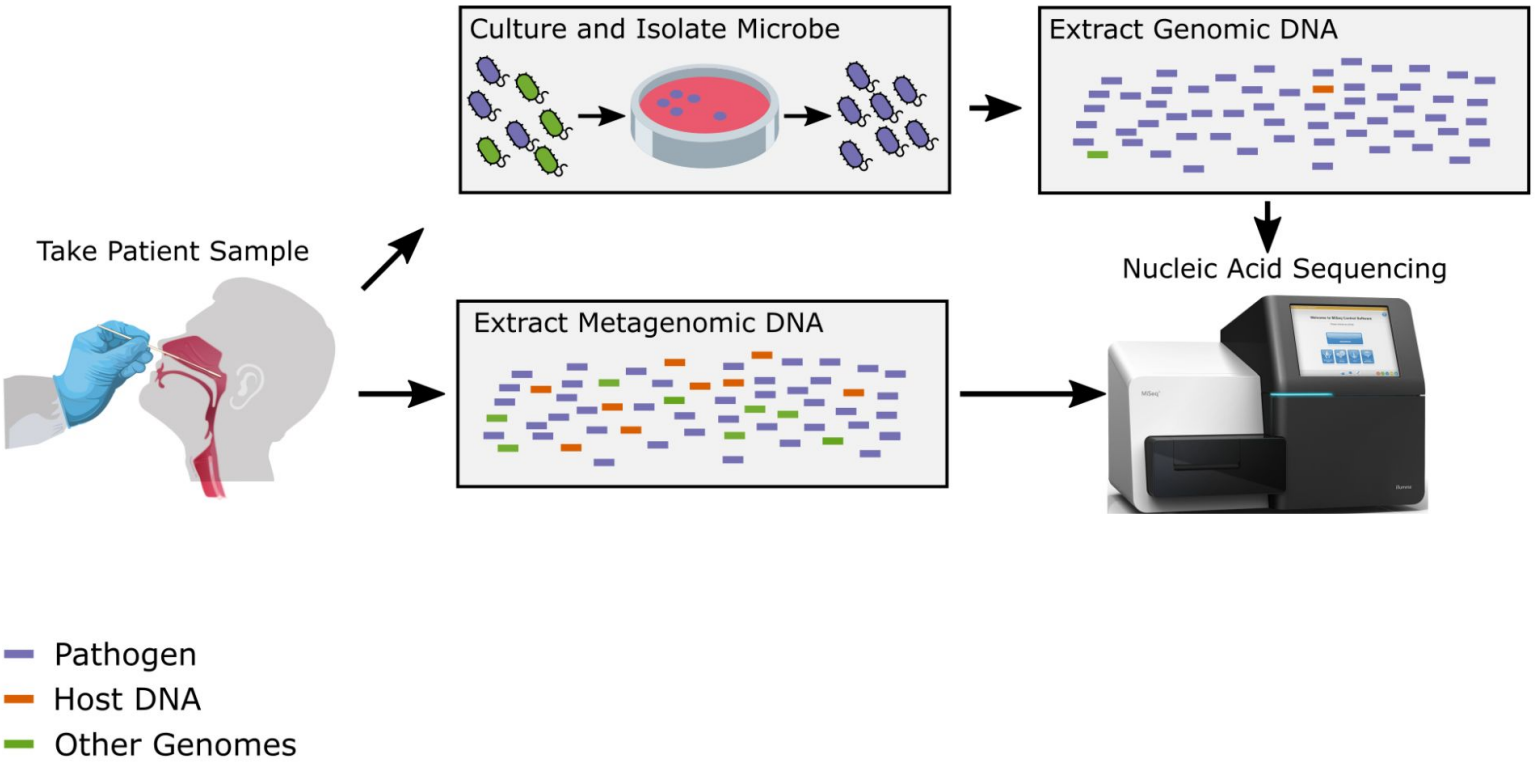


- Pathogen
- Host DNA
- Other Genomes

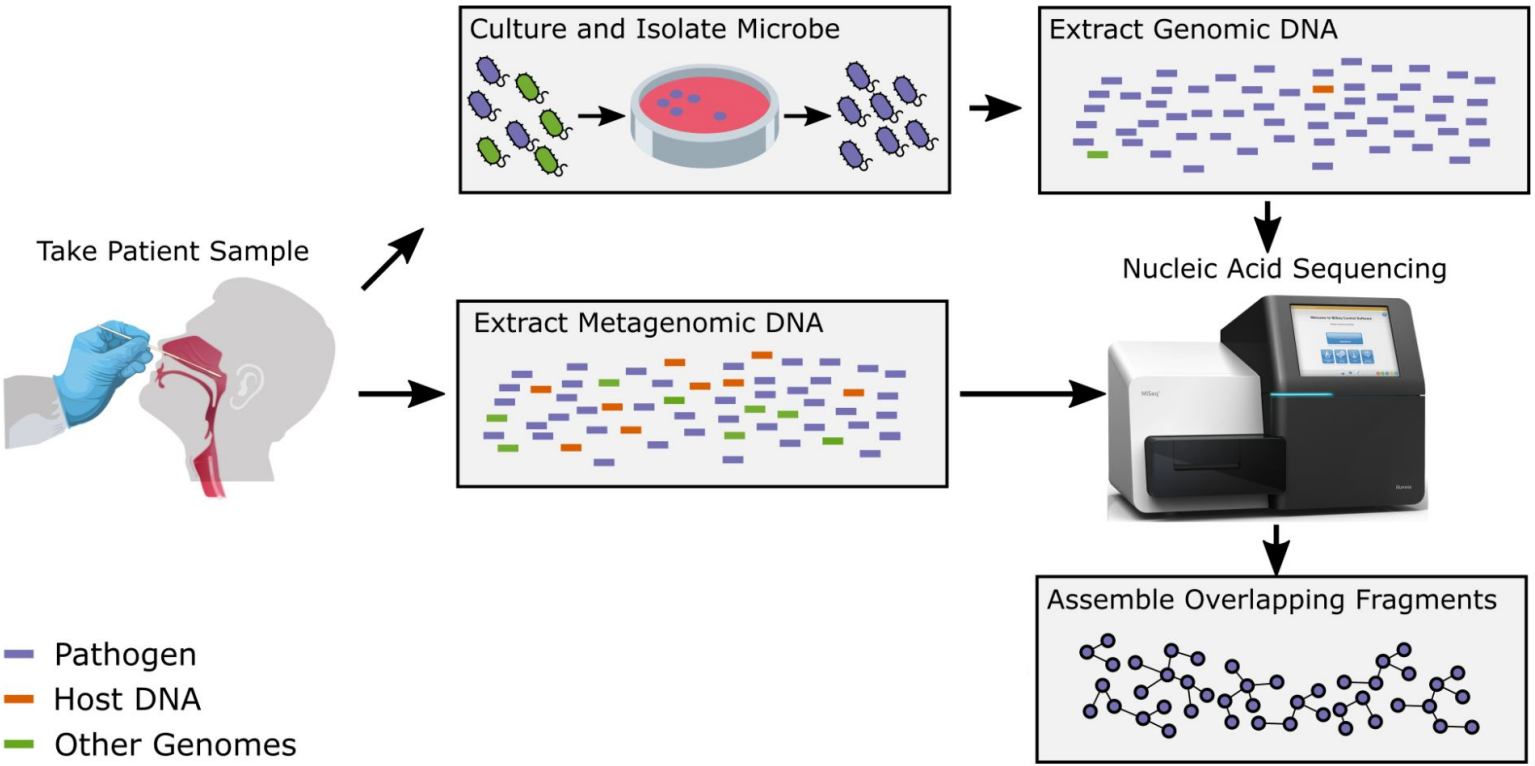
Sequencing Pathogen Genomes



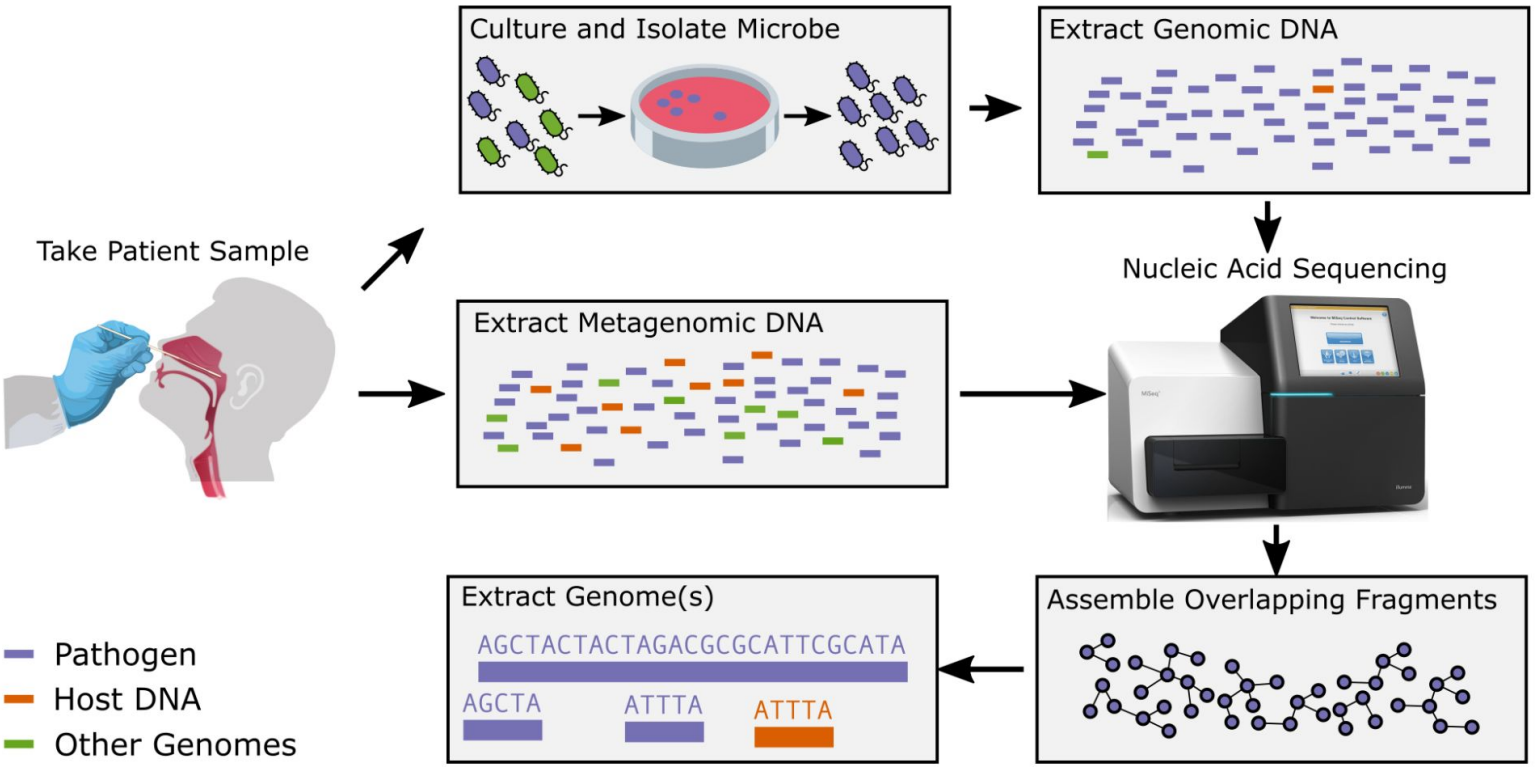
Sequencing Pathogen Genomes



Sequencing Pathogen Genomes

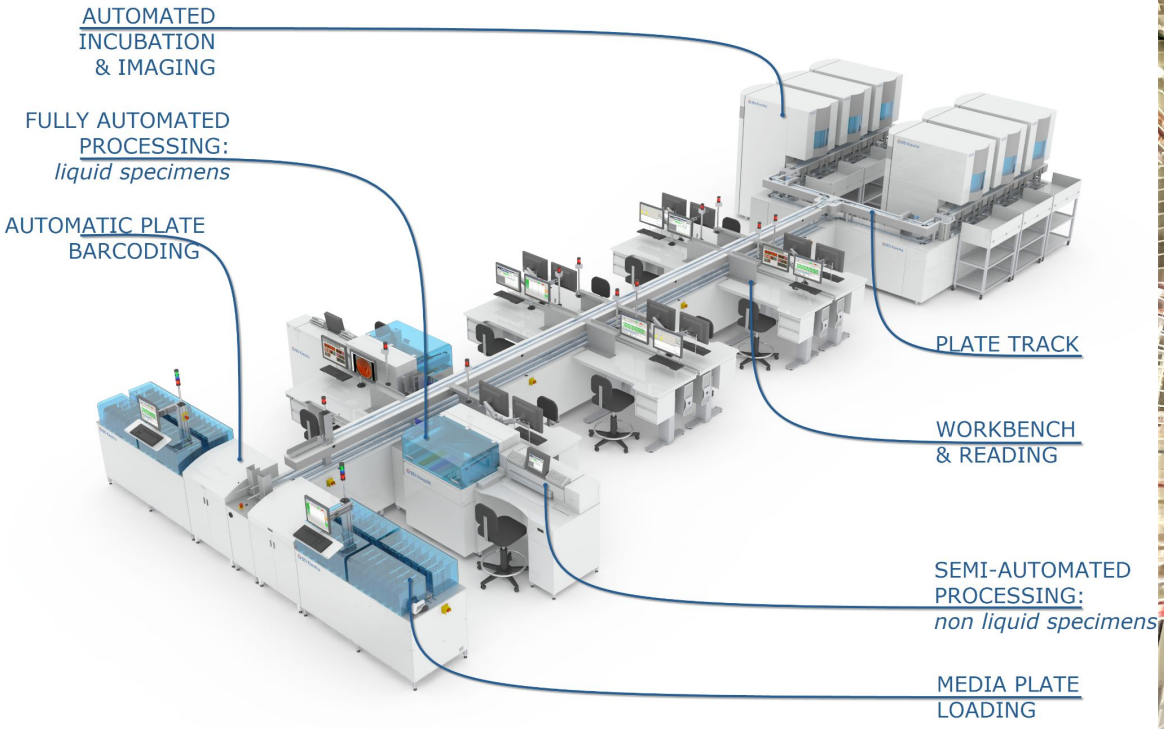


Sequencing Pathogen Genomes



How do we do this at scale?

Automation of labour intensive steps



https://www.bd.com/scripts/europe/labautomation/productsdrilldown.asp?CatID=455&SubID=1836&siteID=20309&d=&s=europe%2Flabautomation&sTitle=Lab+Automation&metaTitle=Total+Lab+Automation&dc=europe&dcTitle=Europe

Dr. Vermeiren

Automation of labour intensive steps

MEDIA PLATE LOADING

https://www.bd.com/scripts/europe/labautomation/productsdrilldown.asp?CatID=455&SubID=1836&siteID=20309&d=&s=europe%2Flabautomation&sTitle=Lab+Automation&metaTitle=Total+Lab+Automation&dc=europe&dcTitle=Europe

Dr. Vermeiren

Sequencing technology has rapidly changed and improved

First generation

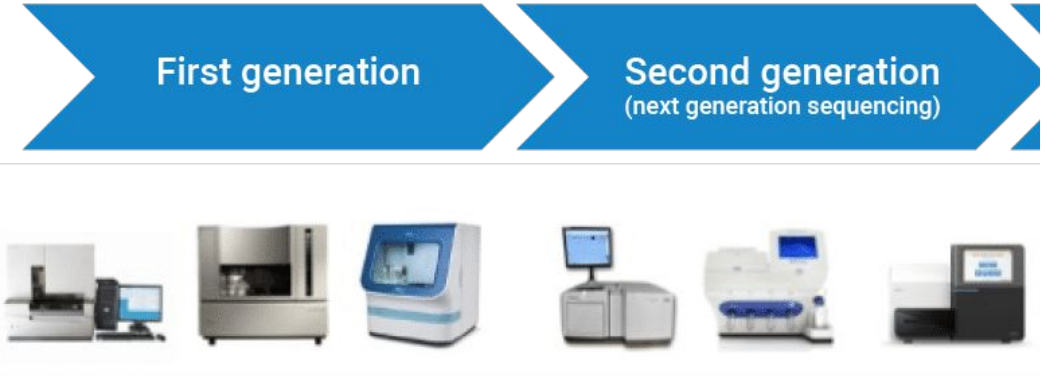
Sanger sequencing
Maxam and Gilbert
Sanger chain termination

Infer nucleotide identity using dNTPs, then visualize with electrophoresis

500–1,000 bp fragments

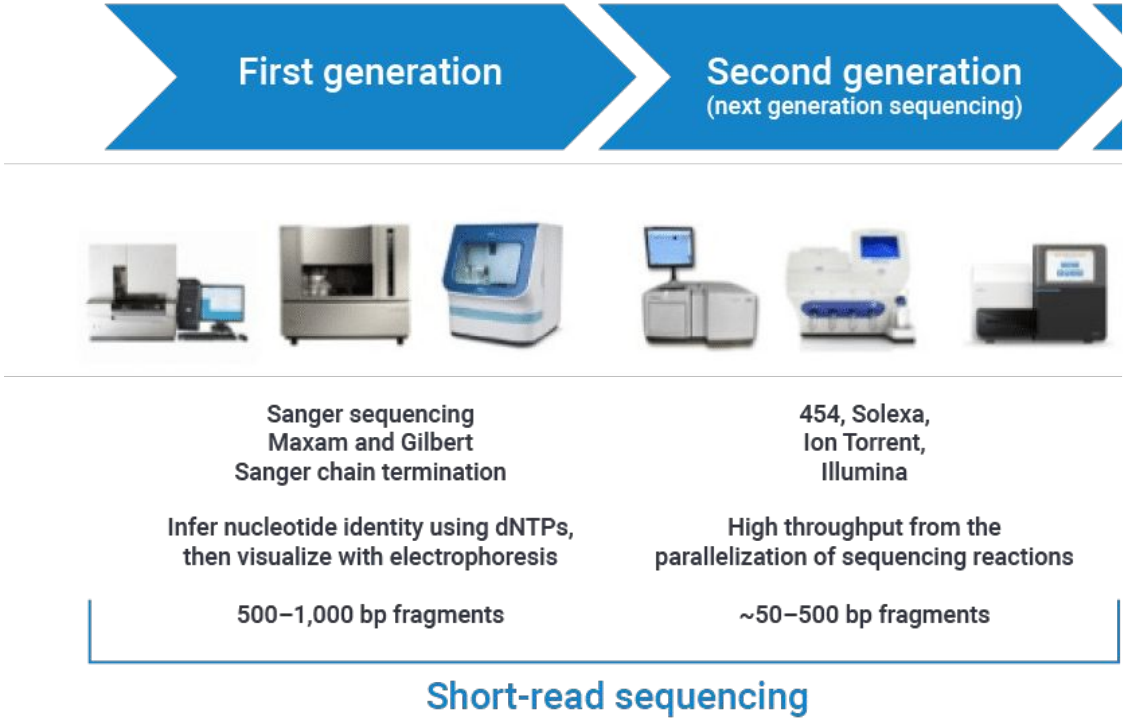
https://www.pacb.com/blog/the-evolution-of-dna-sequencing-tools/

Sequencing technology has rapidly changed and improved



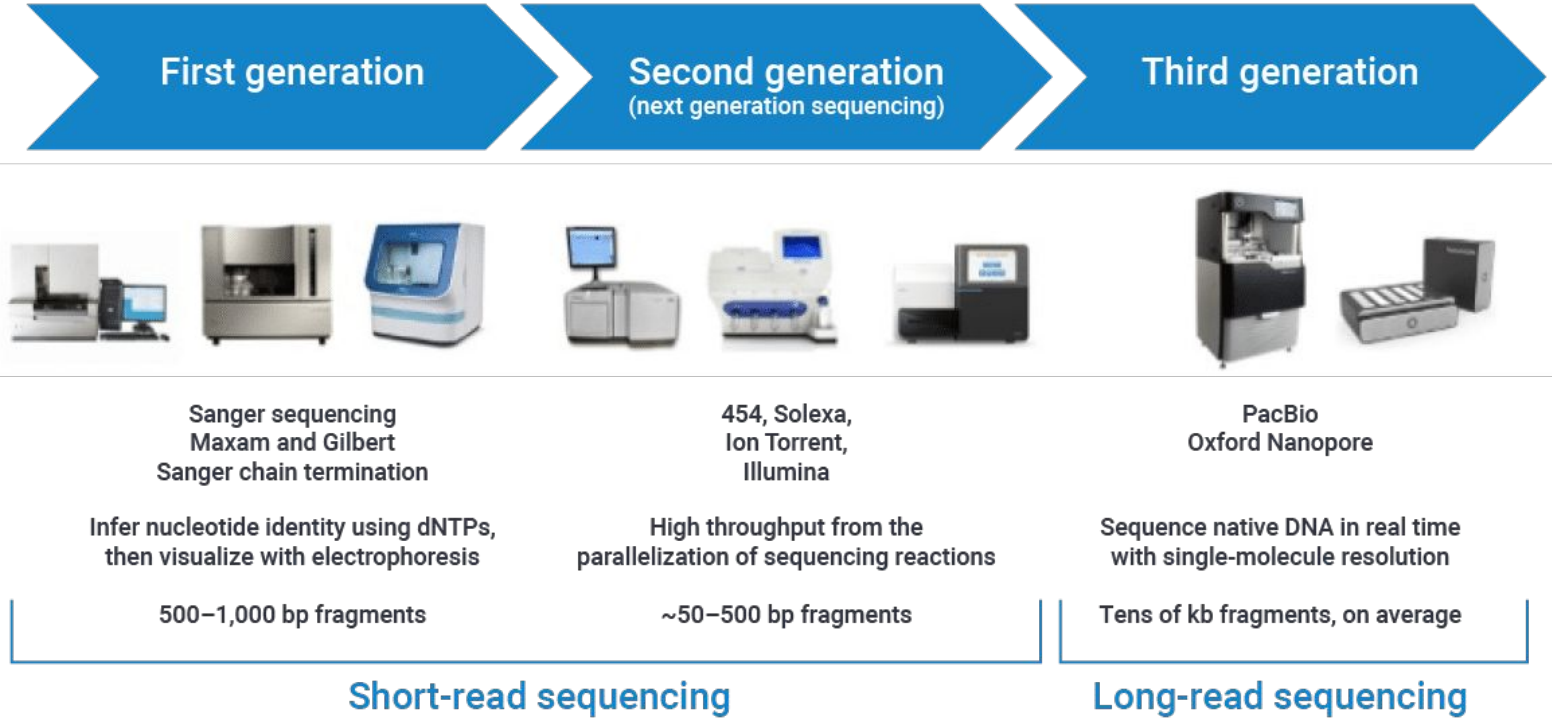
https://www.pacb.com/blog/the-evolution-of-dna-sequencing-tools/

Sequencing technology has rapidly changed and improved



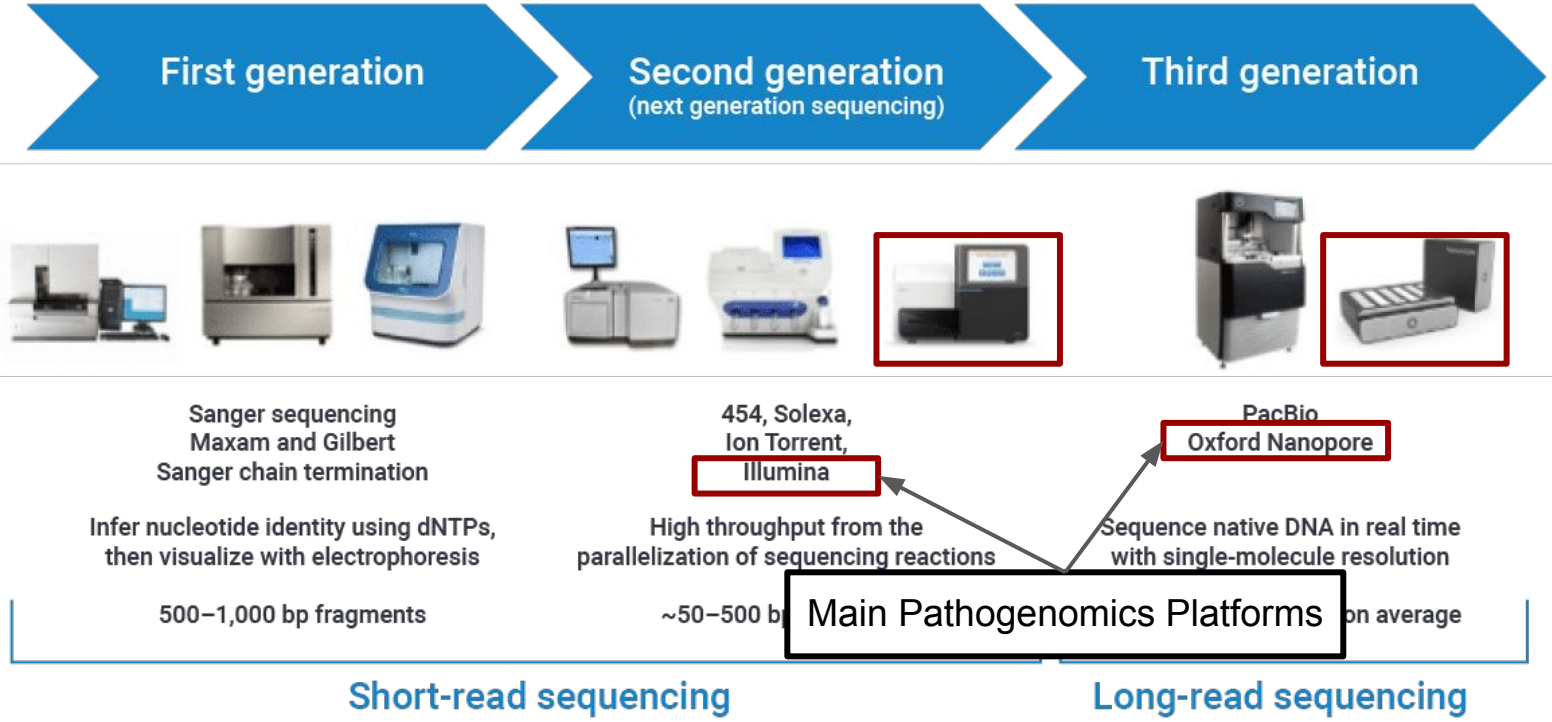
https://www.pacb.com/blog/the-evolution-of-dna-sequencing-tools/

Sequencing technology has rapidly changed and improved



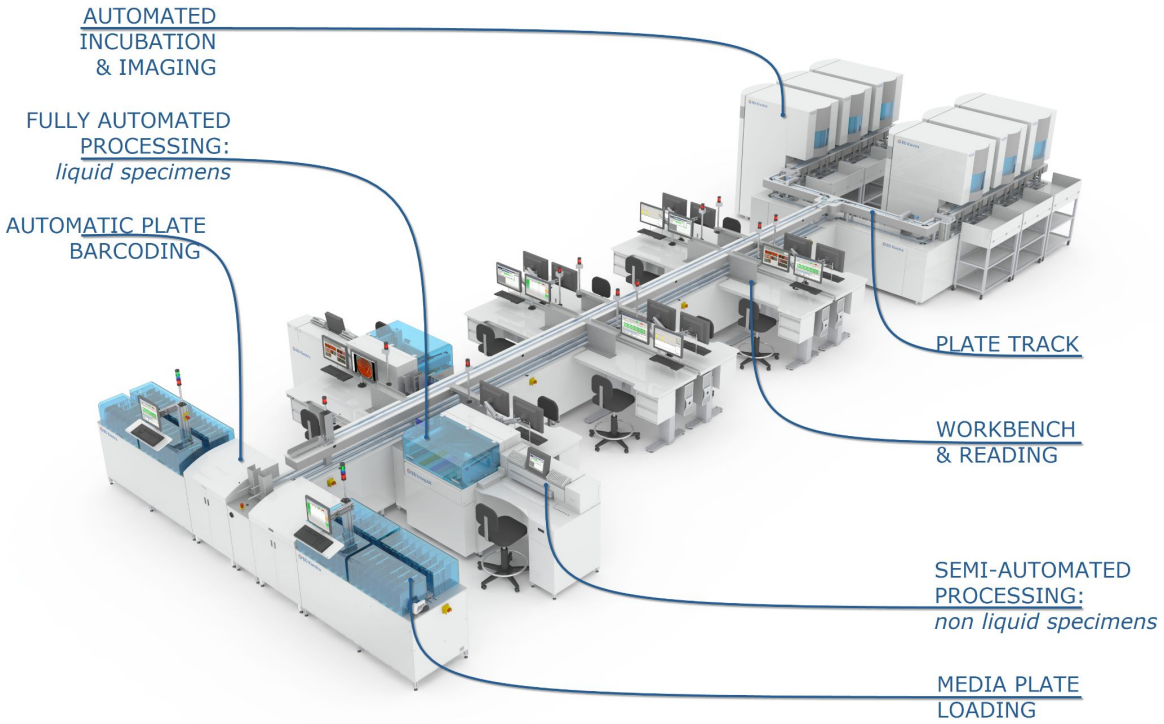
https://www.pacb.com/blog/the-evolution-of-dna-sequencing-tools/

Sequencing technology has rapidly changed and improved



https://www.pacb.com/blog/the-evolution-of-dna-sequencing-tools/

Mobile sequencing lab in a suitcase



https://www.bd.com/scripts/europe/labautomation/productsdrilldown.asp?CatID=455&SubID=1836&siteID=20309&d=&s=europe%2Flabautomation&sTitle=Lab+Automation&metaTitle=Total+Lab+Automation&dc=europe&dcTitle=Europe

Mobile sequencing lab in a suitcase

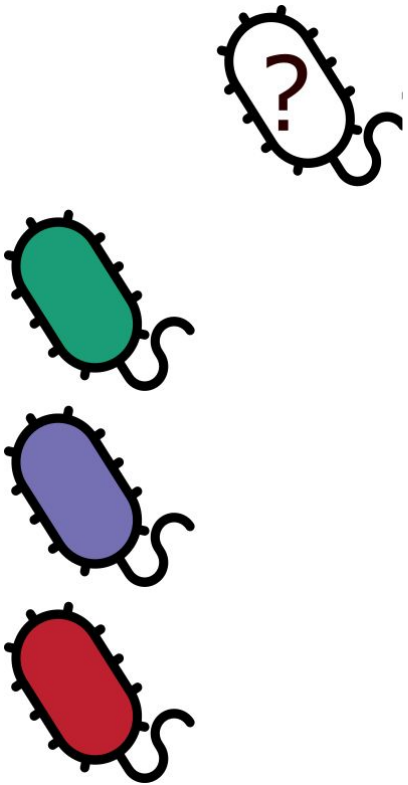
https://crain-platform-genomeweb-prod.s3.amazonaws.com/s3fs-public/styles/1200x630/public/lab_in_a_suitcase.jpeg

See lecture reading: Loman & Gardy 2017

Got a genome, now what?

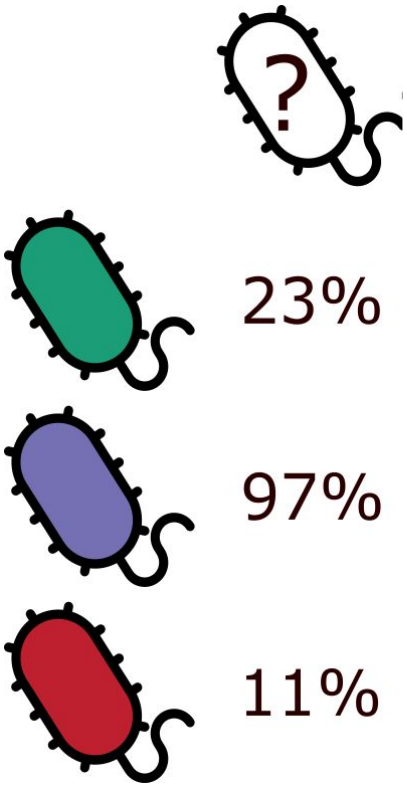
Genomic Diagnostics: What is the pathogen?

Genomic Diagnostics: What is the pathogen?



- Compare to genomes in database from known organisms

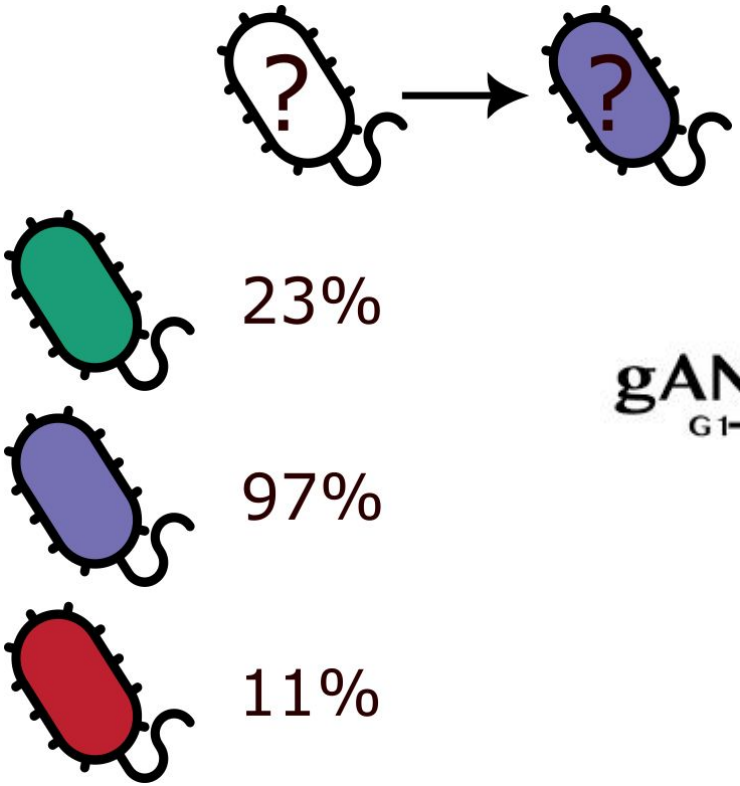
Genomic Diagnostics: What is the pathogen?



- Compare to genomes in database from known organisms
- Average Nucleotide Identity (ANI) is an example of a similarity metric

$$gANI_{G1 \rightarrow G2} = \frac{\sum_{bbh} (\text{Percent Identity} * \text{Alignment length})}{\text{lengths of BBH genes}}$$

Genomic Diagnostics: What is the pathogen?



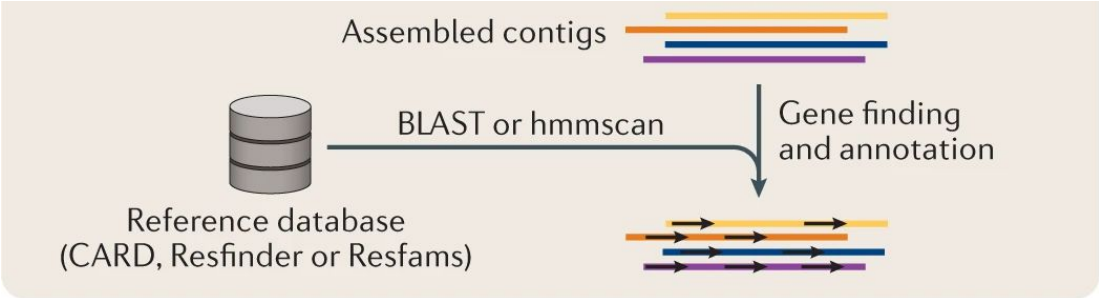
- Compare to genomes in database from known organisms
- Average Nucleotide Identity (ANI) is an example of a similarity metric

$$gANI_{G1 \rightarrow G2} = \frac{\sum_{bbh} (\text{Percent Identity} * \text{Alignment length})}{\text{lengths of BBH genes}}$$

- Identify pathogen as closest reference genome taxa
- Use identity to drive treatment (if x then treat by y)

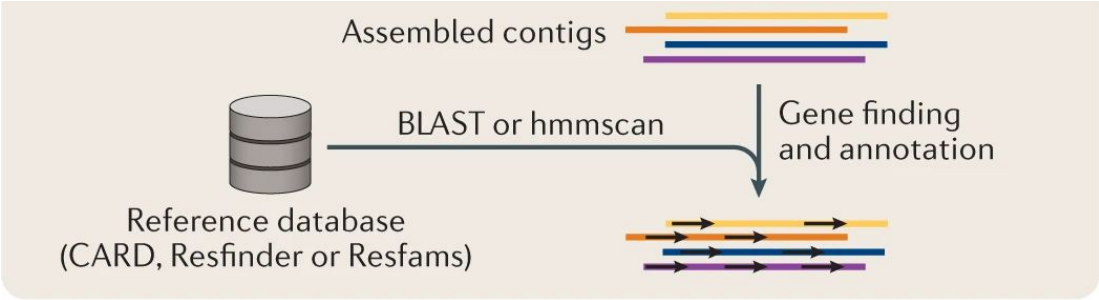
- Typing for outbreak investigation linkage

Genomic Diagnostics: What drugs will work?



10.1038/s41576-019-0108-4

Genomic Diagnostics: What drugs will work?



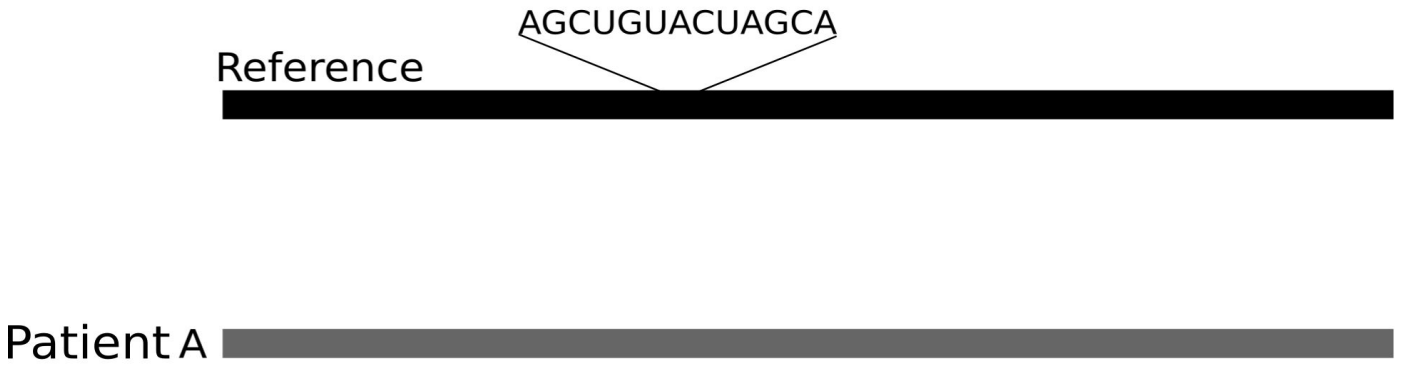
10.1038/s41576-019-0108-4

- Detect NDM-1 carbapenemase gene
- Pathogen protein that destroys many antibiotics (beta-lactams)
- => Treat with alternative class of antibiotics (e.g., colistin)

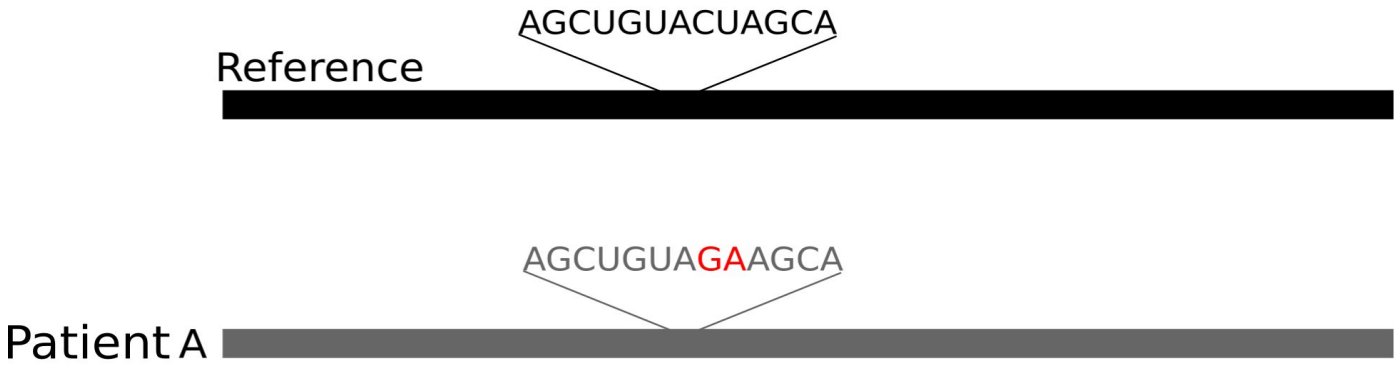
- Hours vs weeks for some pathogens (TB)

OK, but what else can we do with genomes?

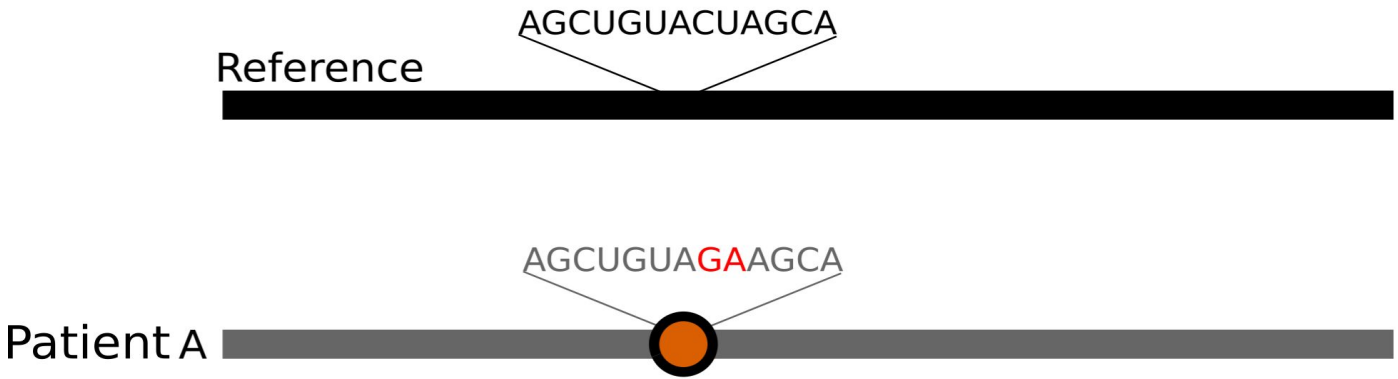
Detection mutations relative to reference



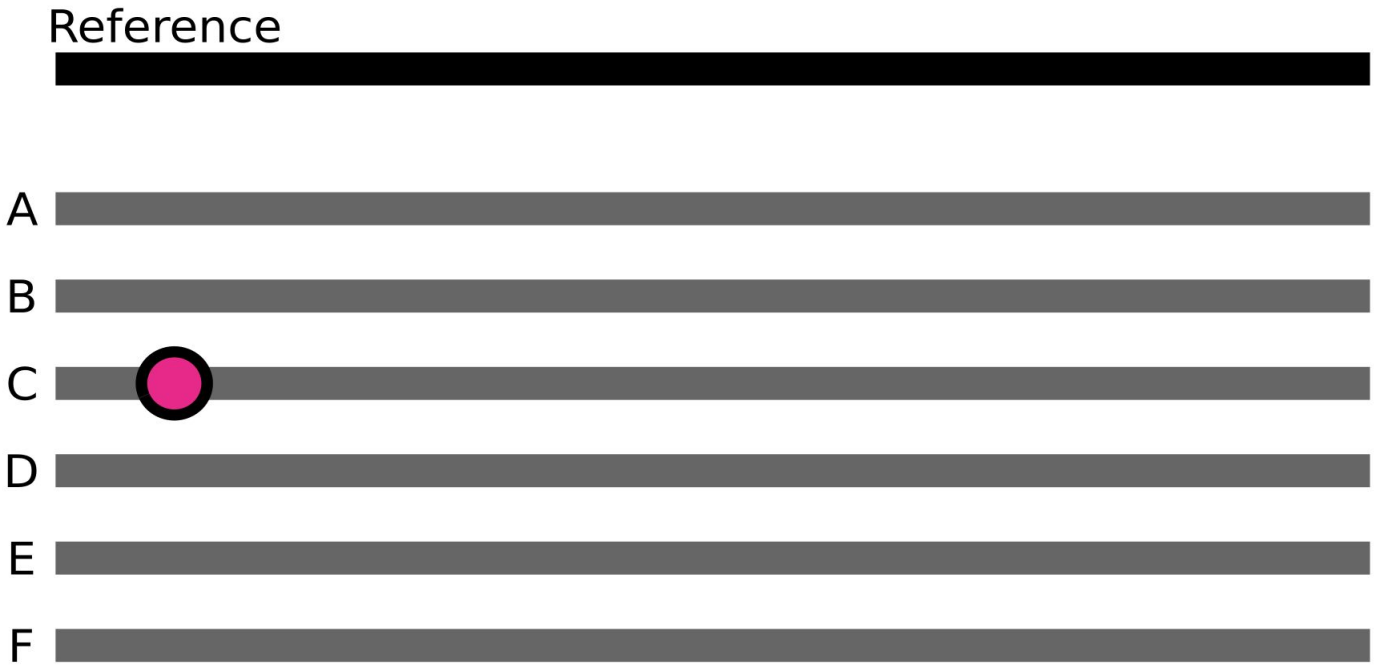
Detection mutations relative to reference



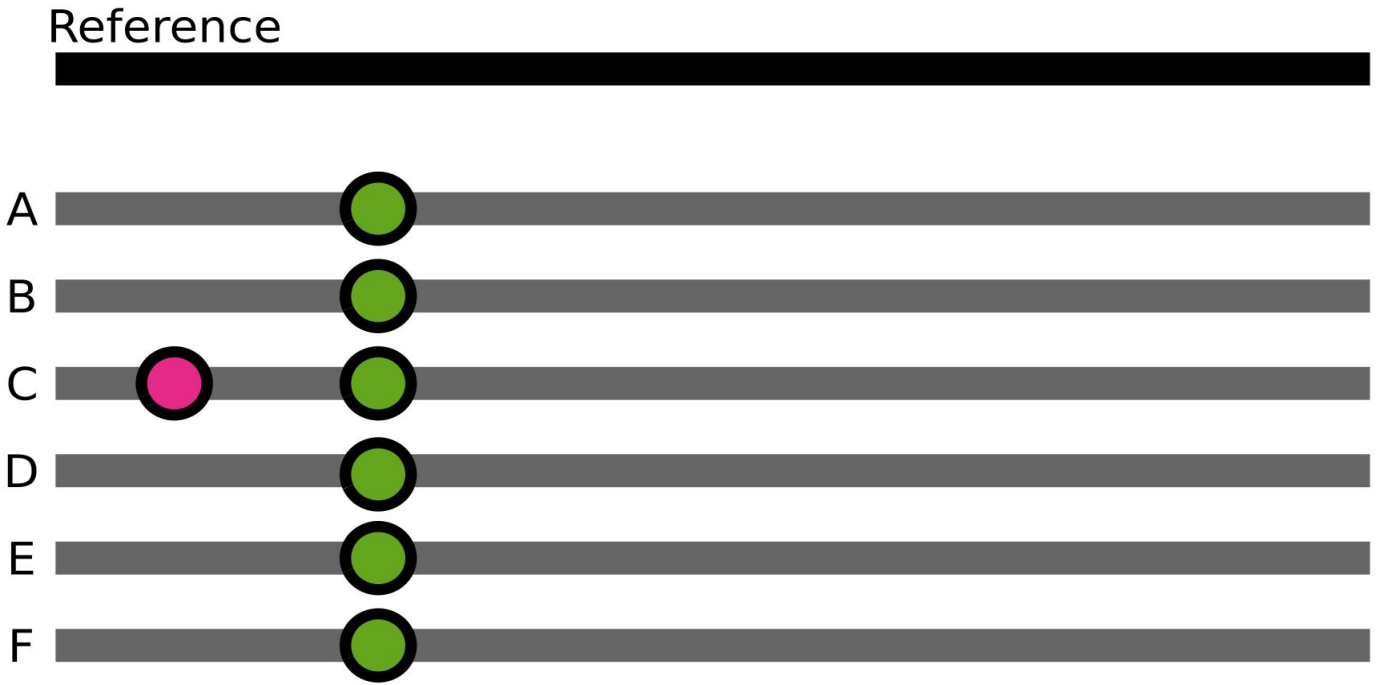
Detection mutations relative to reference



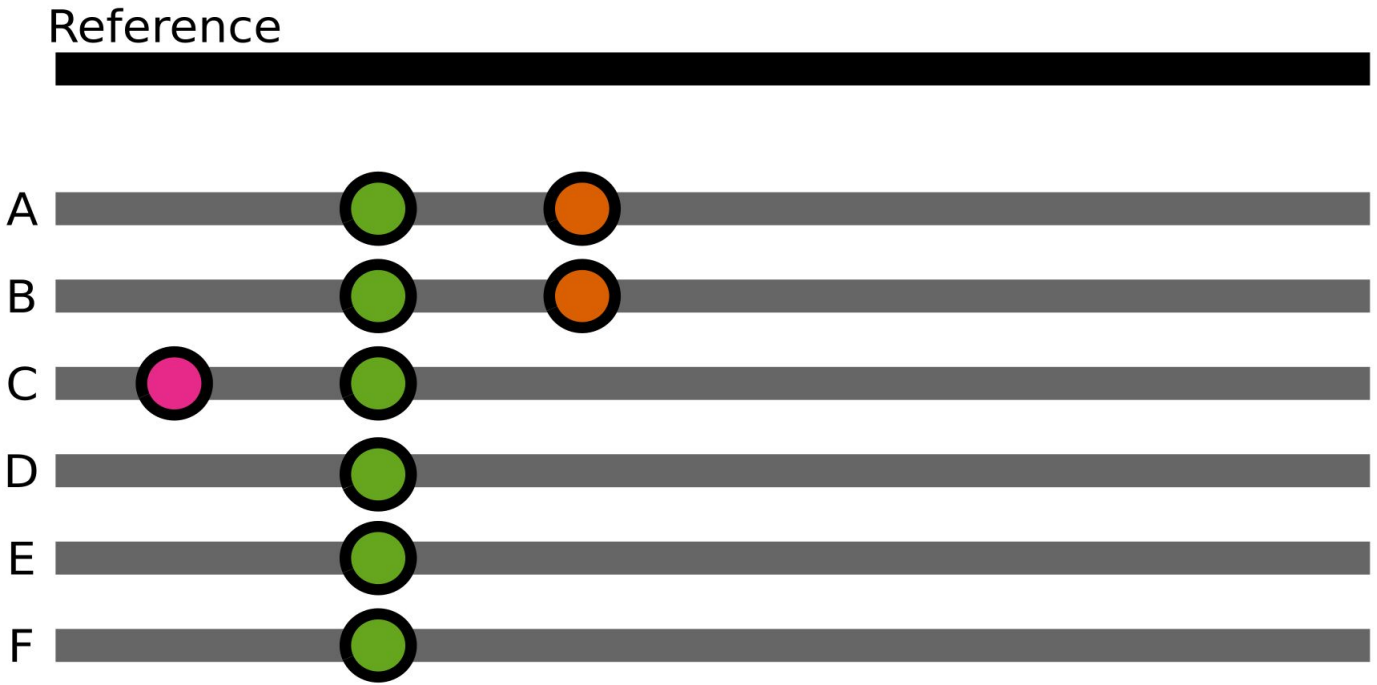
Compare mutations across patients



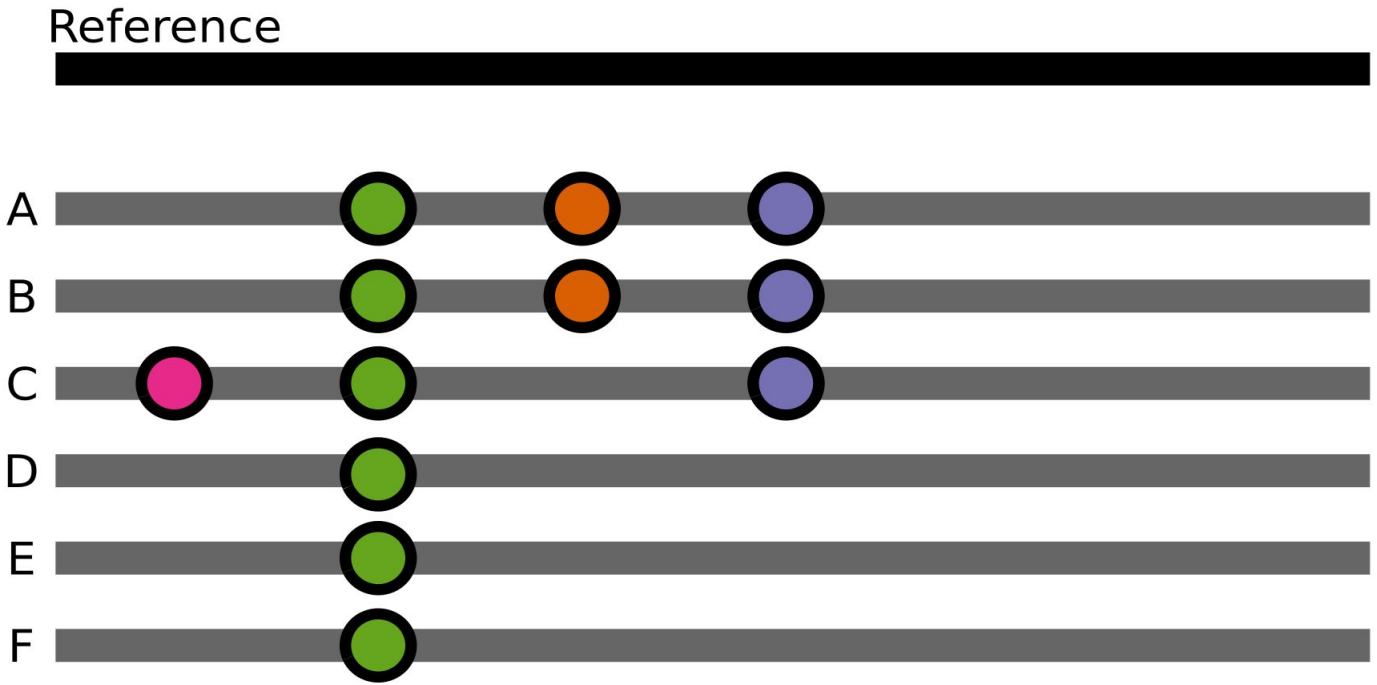
Compare mutations across patients



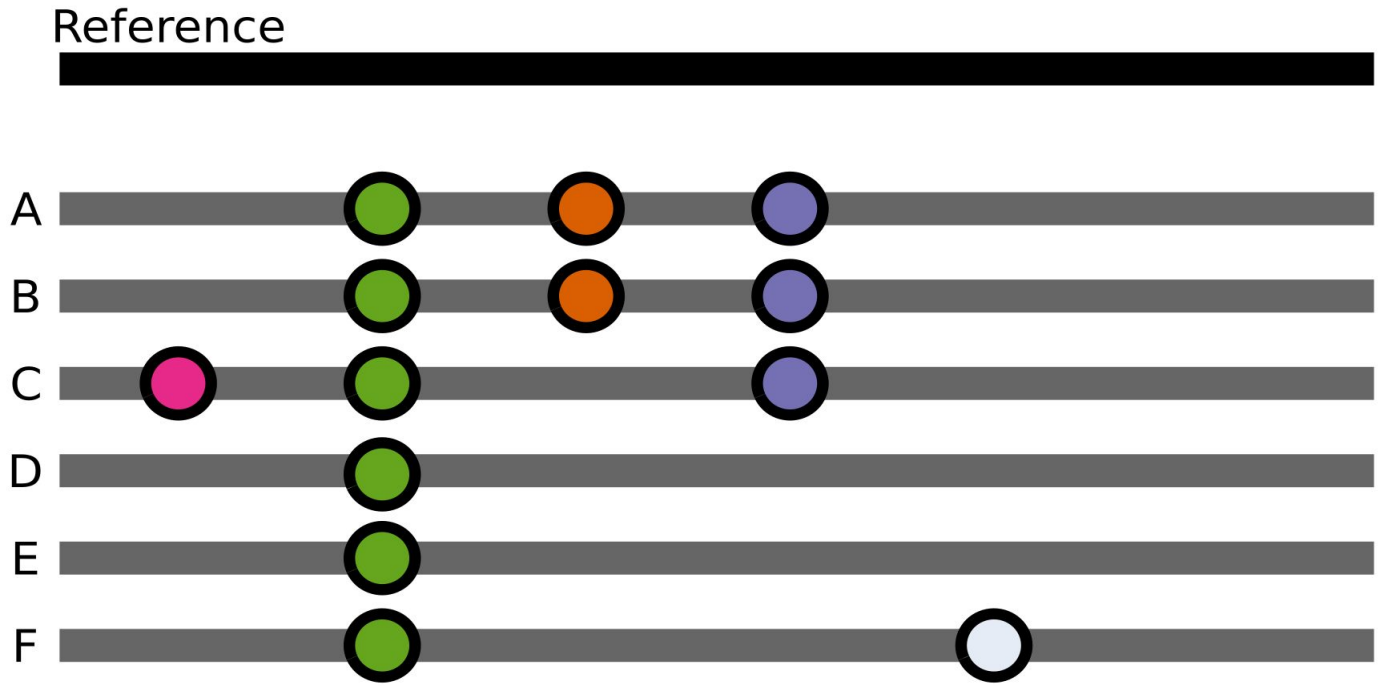
Compare mutations across patients



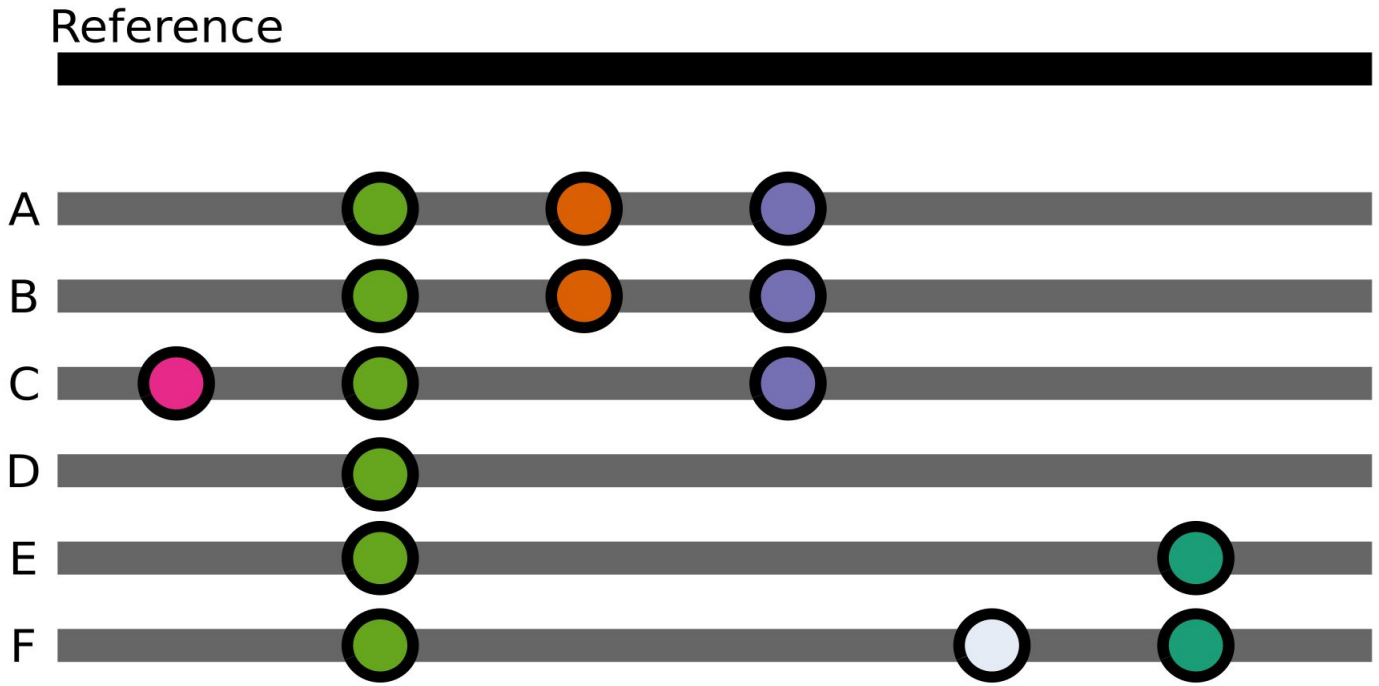
Compare mutations across patients



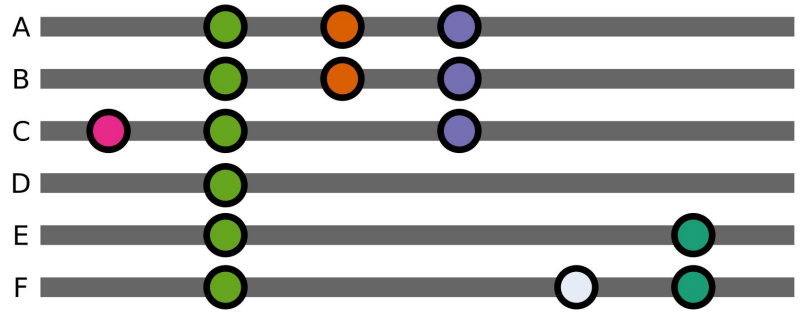
Compare mutations across patients



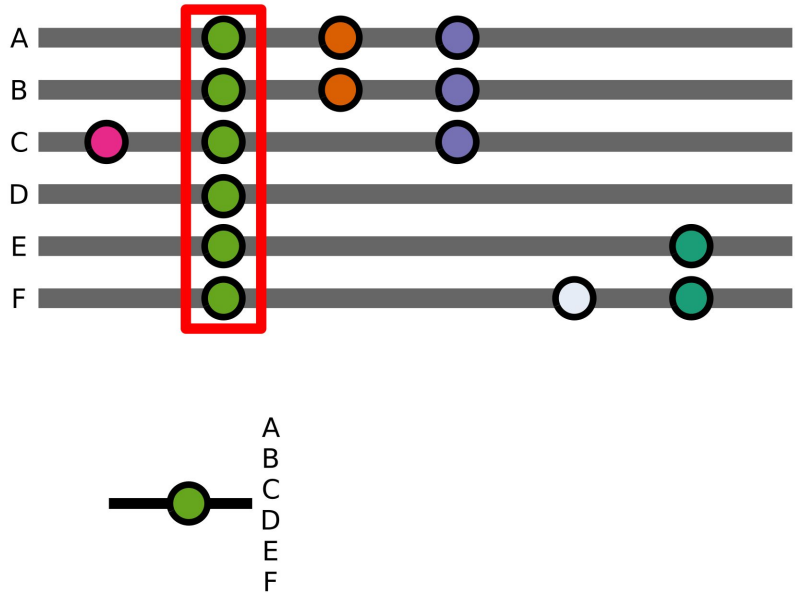
Compare mutations across patients



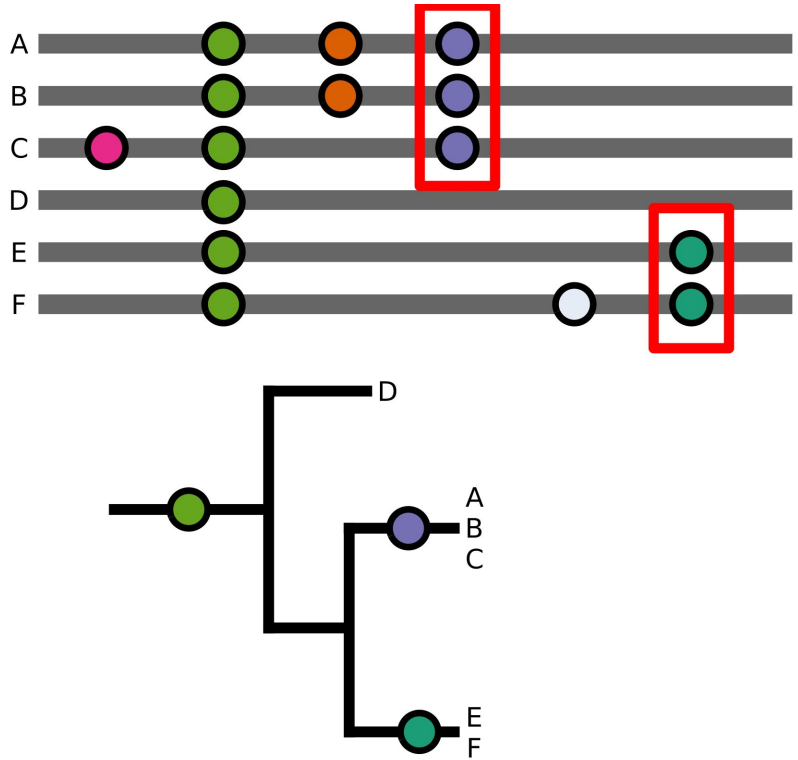
Using pattern of mutations to infer relationships



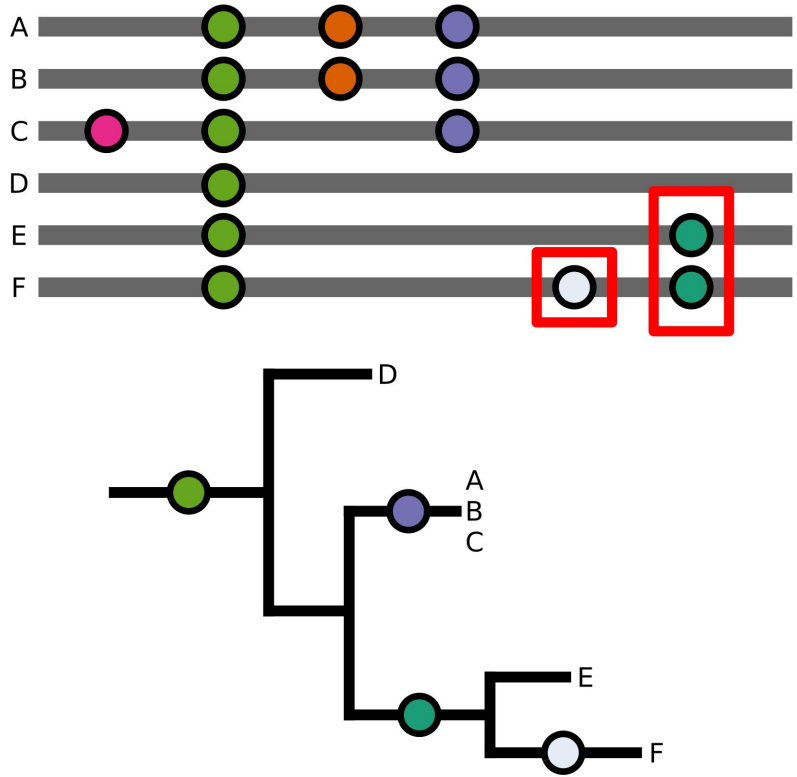
Using pattern of mutations to infer relationships



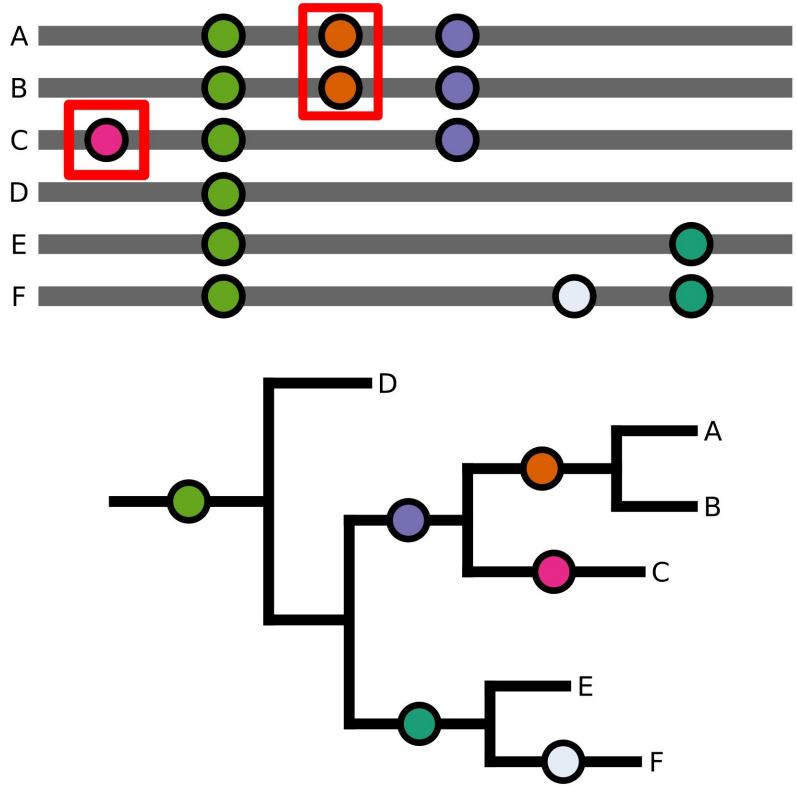
Using pattern of mutations to infer relationships



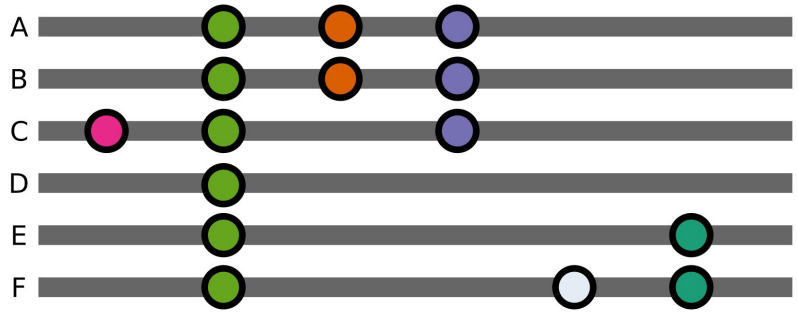
Using pattern of mutations to infer relationships

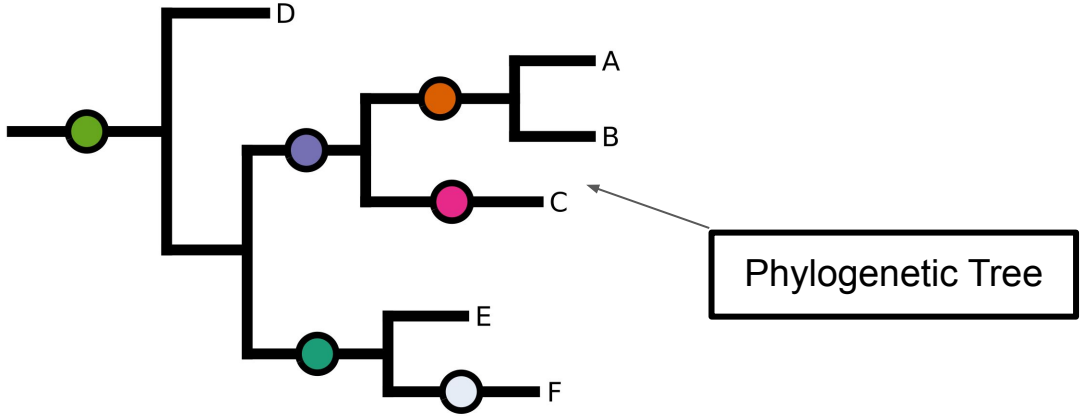


Using pattern of mutations to infer relationships



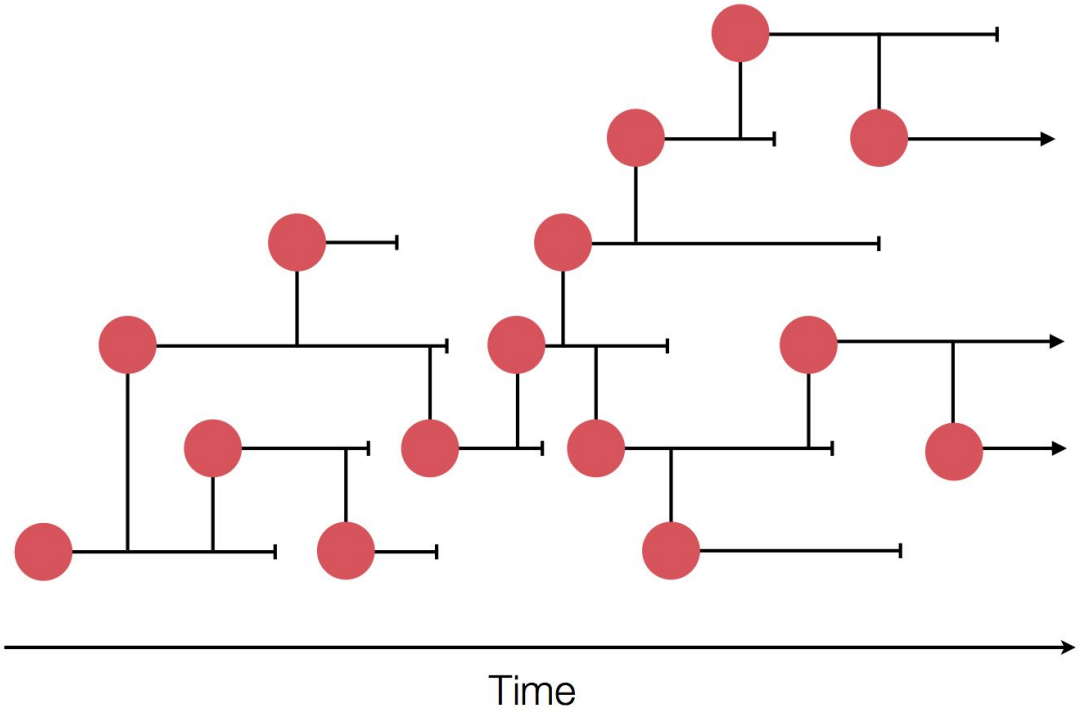
Using pattern of mutations to infer relationships





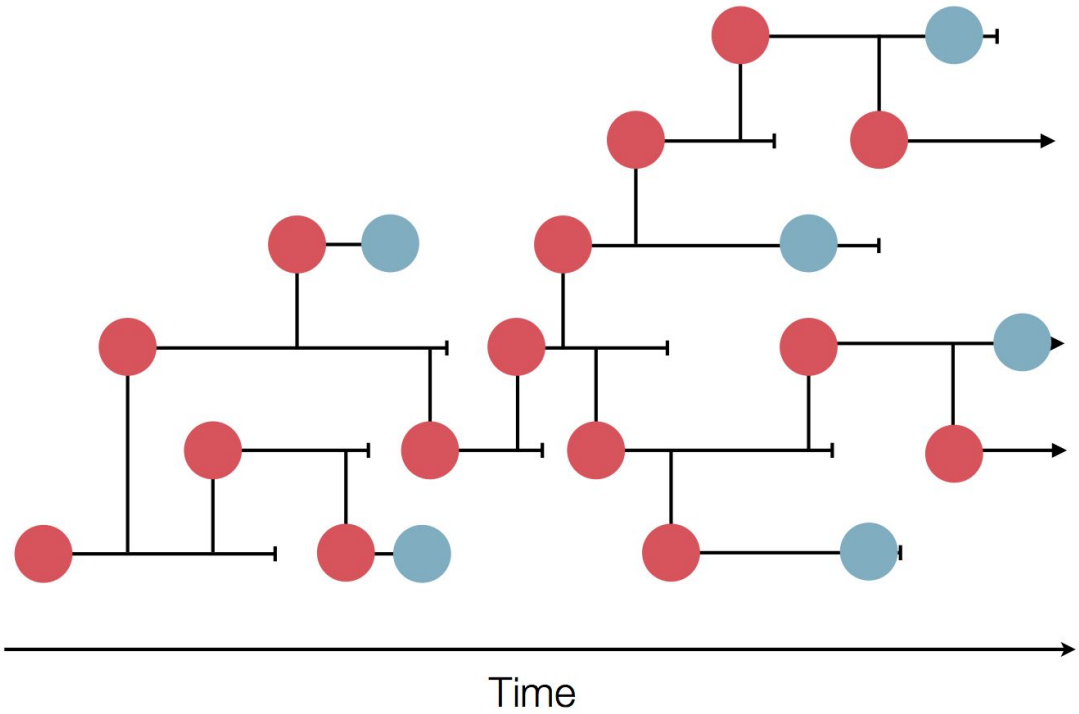
What does this tree actually represent?

Sampling & partially reconstructing underlying epidemic process



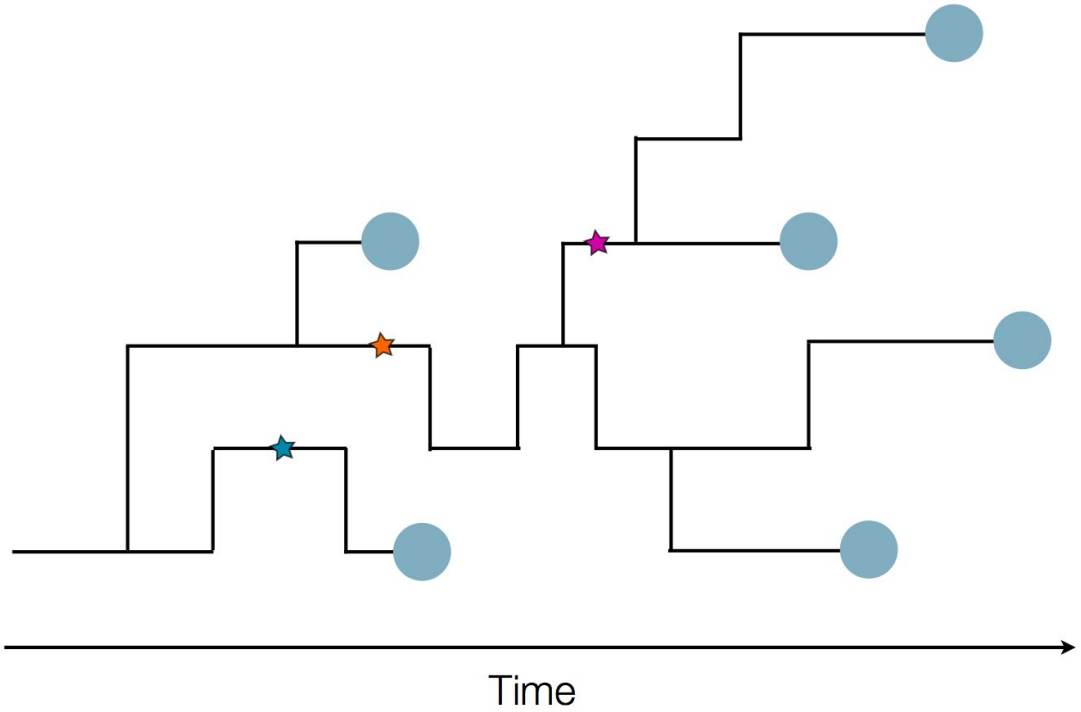
https://github.com/trvrb/phylodynamics-lecture

Sampling & partially reconstructing underlying epidemic process



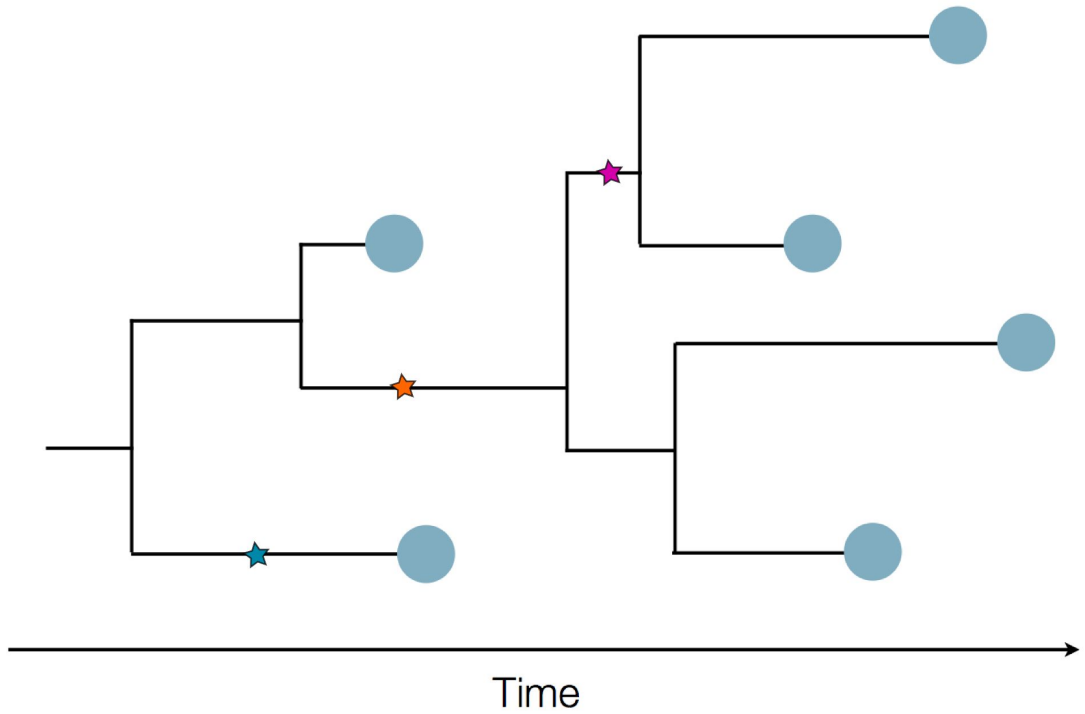
https://github.com/trvrb/phylodynamics-lecture

Sampling & partially reconstructing underlying epidemic process



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

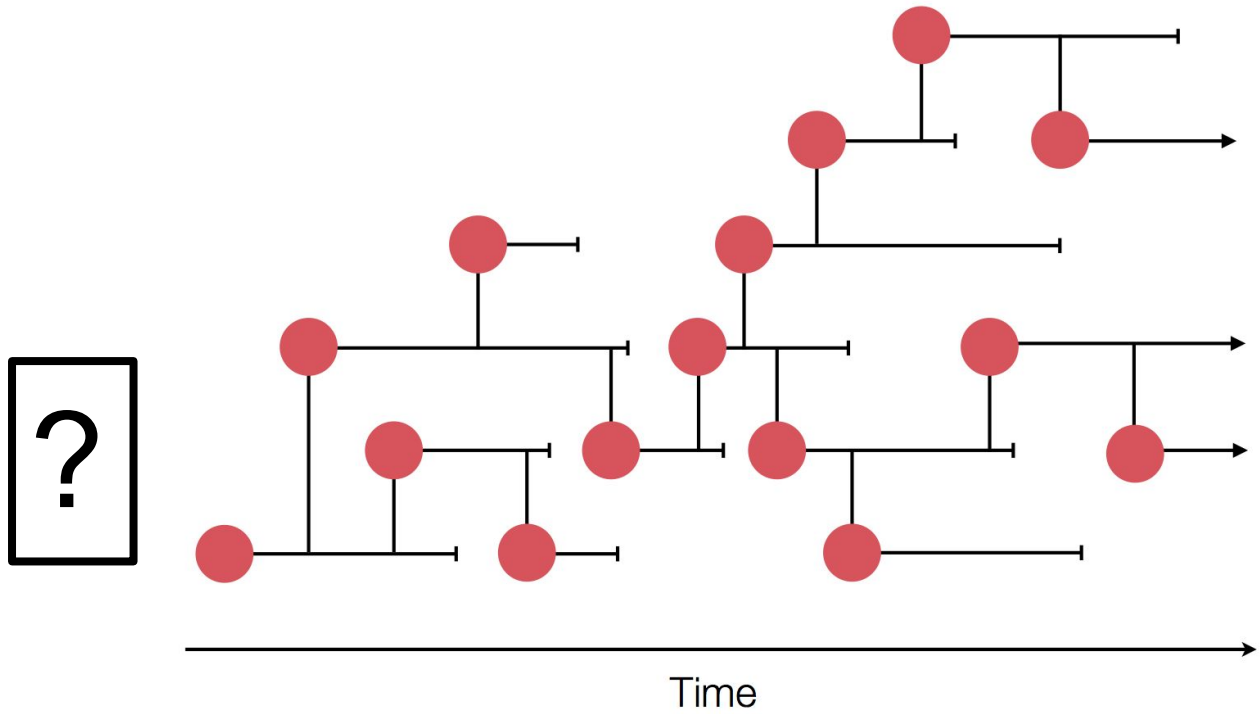
Sampling & partially reconstructing underlying epidemic process



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

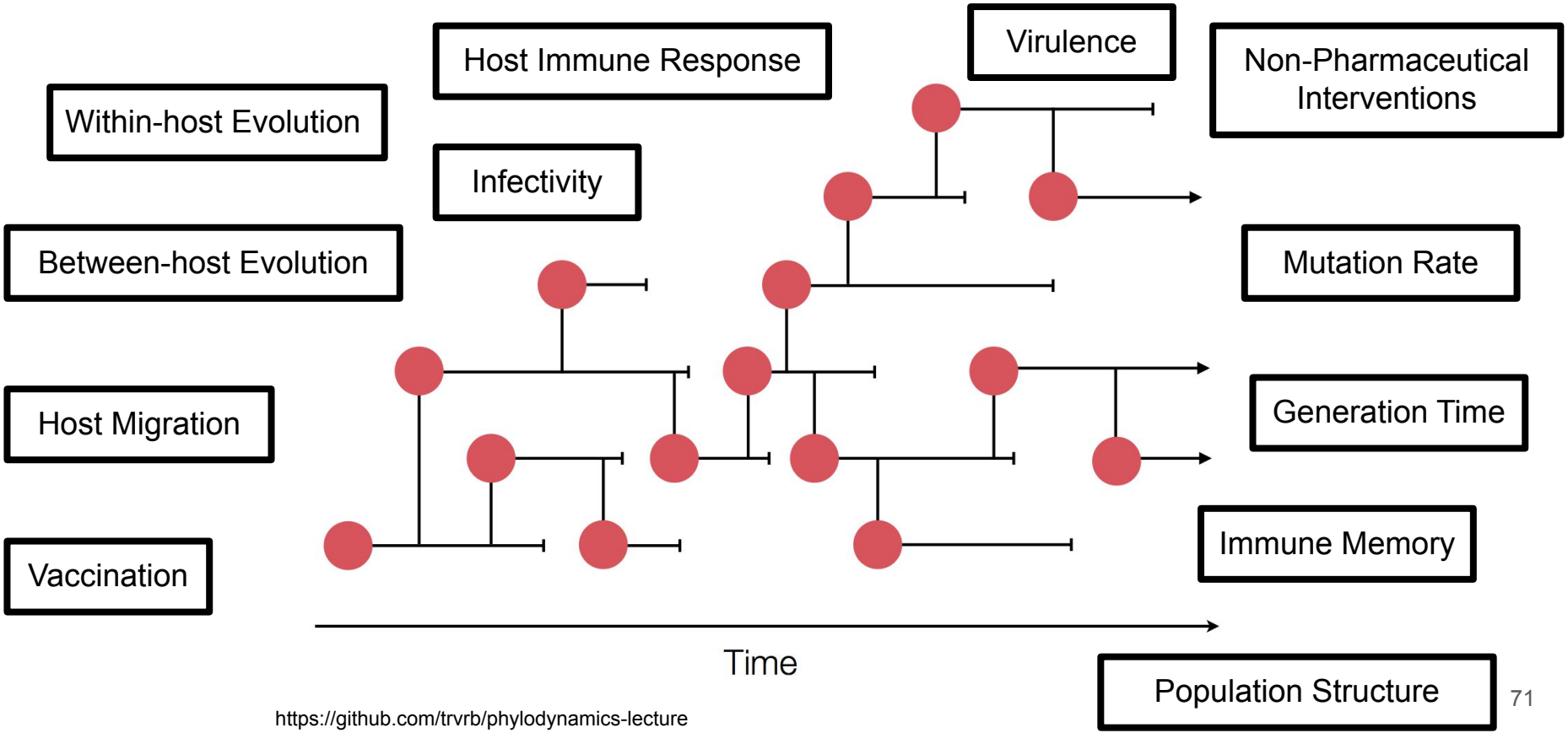
What determines underlying process?

Many forces shaping underlying process



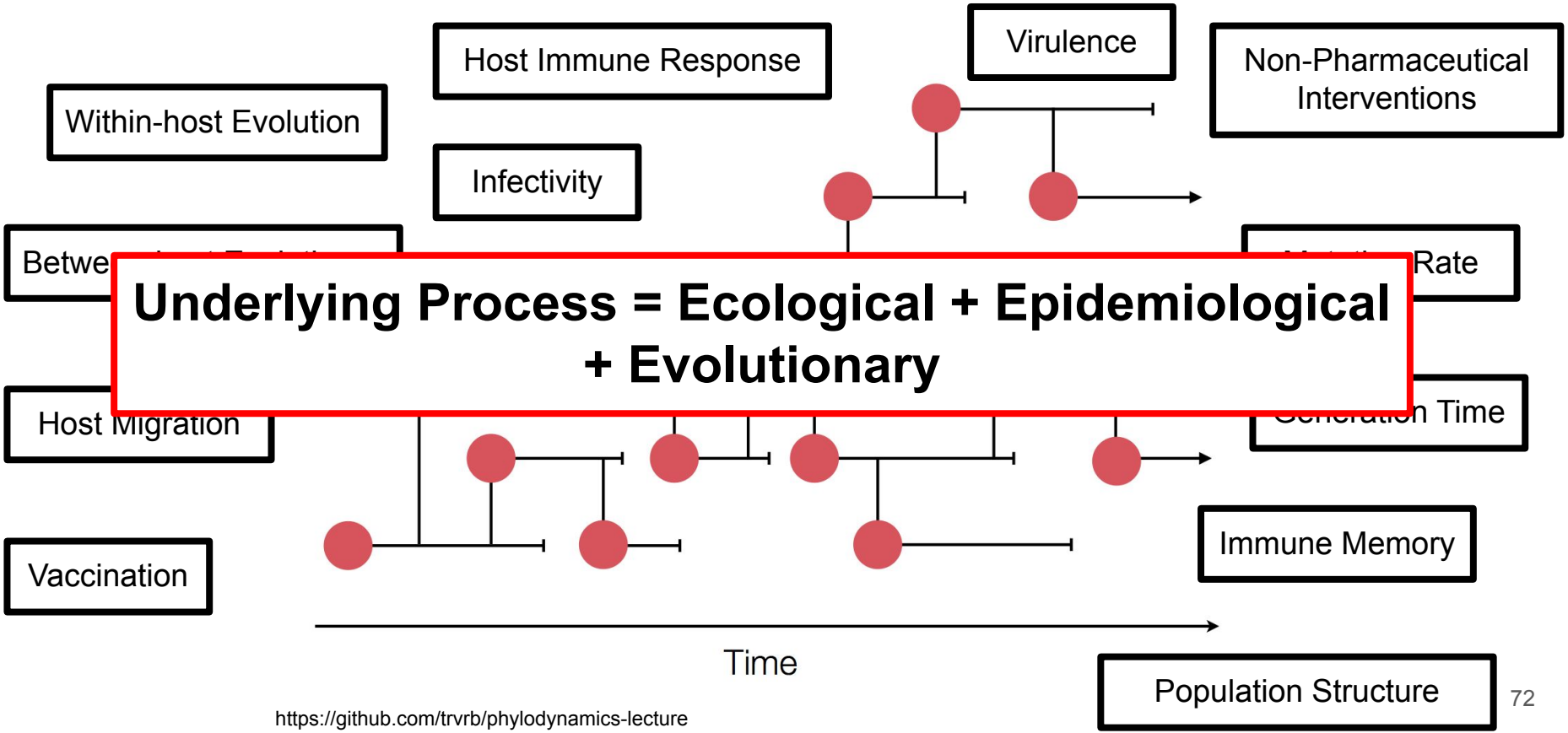
https://github.com/trvrb/phylodynamics-lecture

Many forces shaping underlying process



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

Many forces shaping underlying process



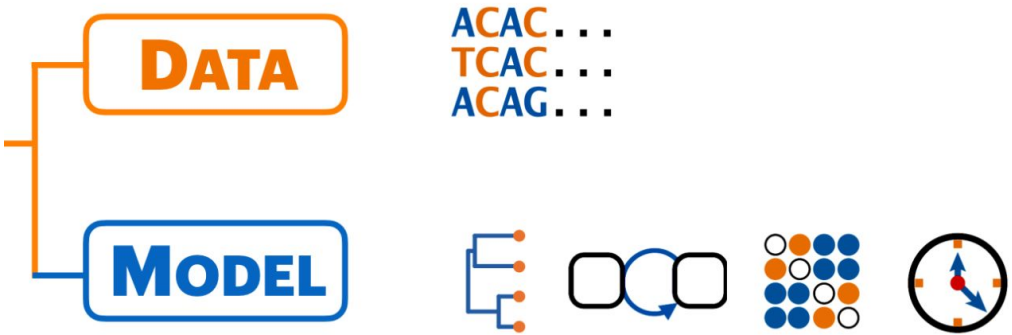
https://github.com/trvrb/phylodynamics-lecture

How can we can model this process?

Bayesian inference is a key tool in genomic epidemiology

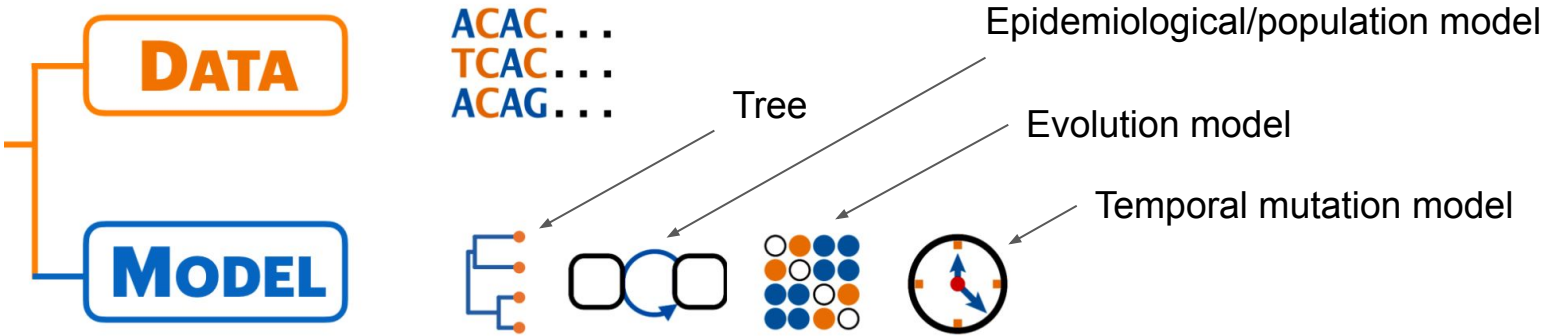
https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/raw/master/Day4_Structured_Phylodynamics_-_Denise_Kuehnert.pdf

Bayesian inference is a key tool in genomic epidemiology



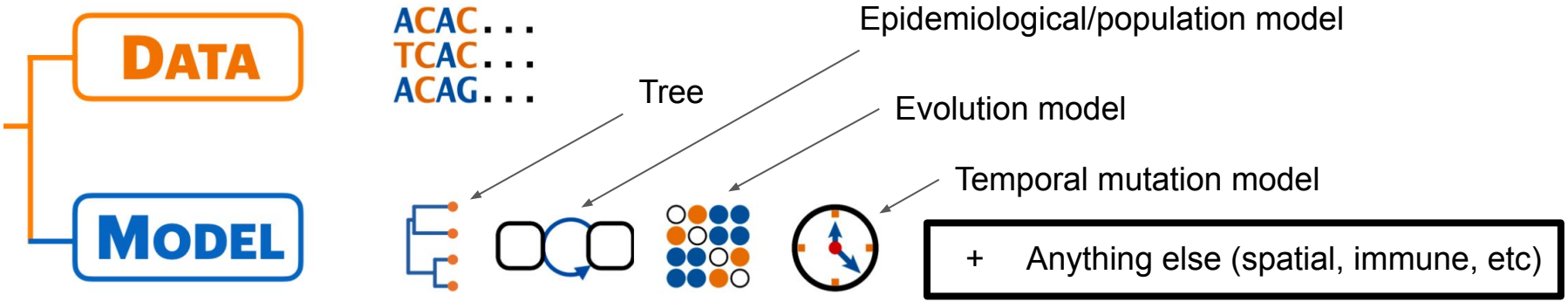
https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/raw/master/Day4_Structured_Phylodynamics_-_Denise_Kuehnert.pdf

Bayesian inference is a key tool in genomic epidemiology



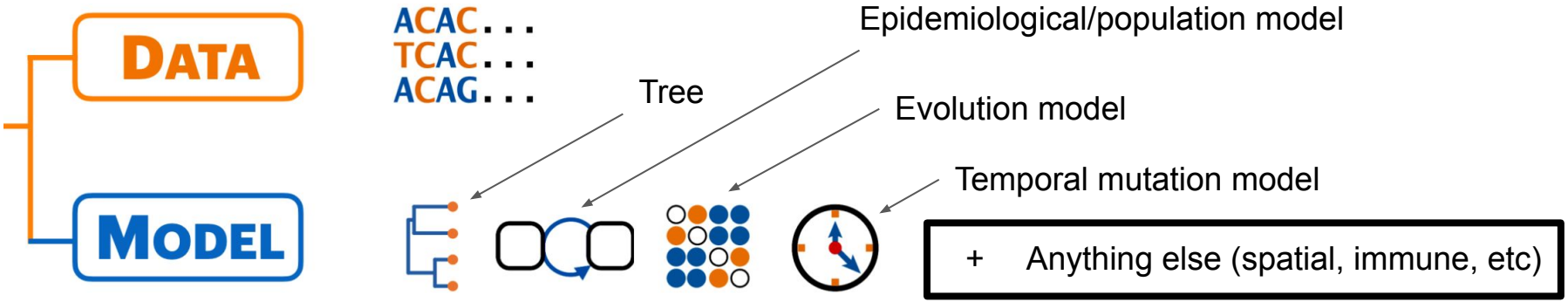
https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/raw/master/Day4_Structured_Phylodynamics_-_Denise_Kuehnert.pdf

Bayesian inference is a key tool in genomic epidemiology

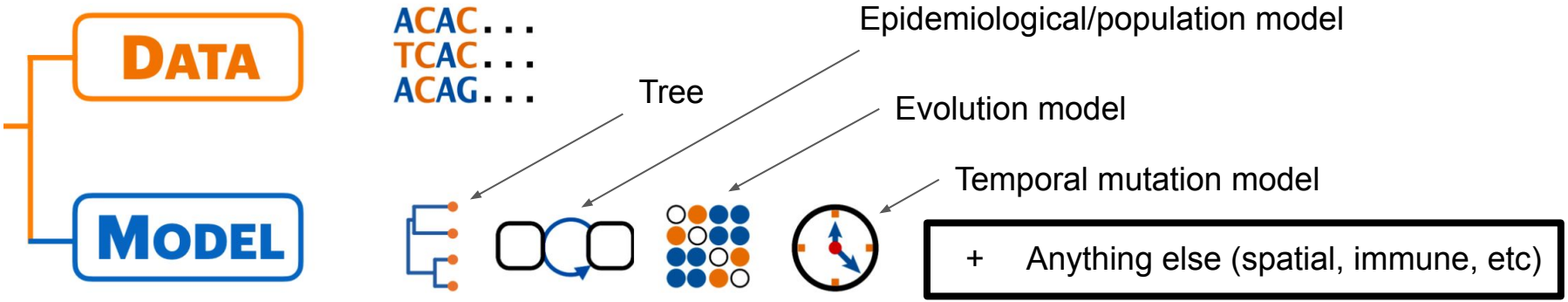


https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/raw/master/Day4_Structured_Phylodynamics_-_Denise_Kuehnert.pdf

Bayesian inference is a key tool in genomic epidemiology



Bayesian inference is a key tool in genomic epidemiology



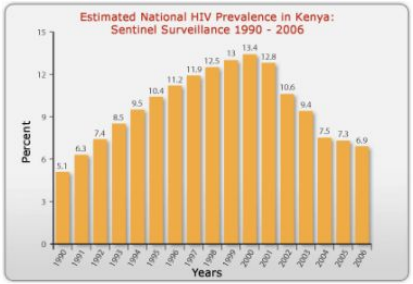
$$P(\text{tree}, \text{model} \mid \text{ACAC...}, \text{TCAC...}, \text{ACAG...}) = \frac{P(\text{ACAC...}, \text{TCAC...}, \text{ACAG...} \mid \text{tree}, \text{model})\, P(\text{tree}, \text{model})}{P(\text{ACAC...}, \text{TCAC...}, \text{ACAG...})}$$

https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/raw/master/Day4_Structured_Phylodynamics_-_Denise_Kuehnert.pdf

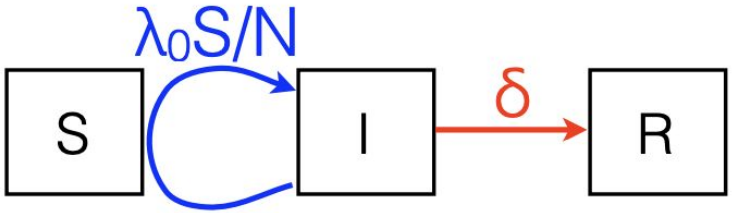
Why not just use case information?

Genomics can be used to infer unobserved events

of infected individuals through time



Population dynamics described by SIR models:



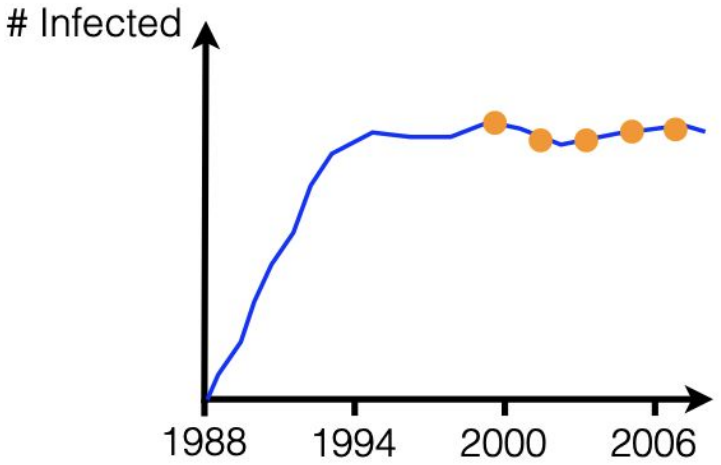
$$\frac{dS}{dt} = -\lambda_0 IS/N$$
$$\frac{dI}{dt} = \lambda_0 IS/N - \delta I$$
$$\frac{dR}{dt} = \delta I$$

https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/raw/master/Day2_Phylodynamics_-_Tanja_Stadler.pdf

Genomics can be used to infer unobserved events

If sampling in early epidemic was missed:

- **Time of epidemic outbreak?**
- **Basic reproductive number R_0?**



https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/raw/master/Day2_Phylodynamics_-_Tanja_Stadler.pdf

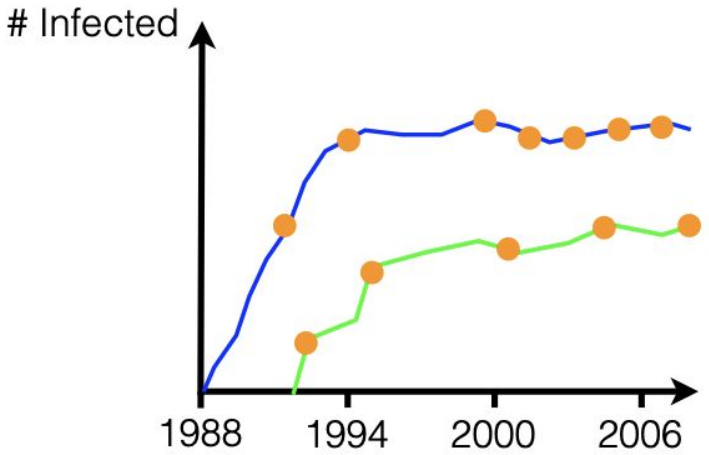
Genomics can be used to infer unobserved events

If sampling in early epidemic was missed:

- **Time of epidemic outbreak?**
- **Basic reproductive number R_0?**

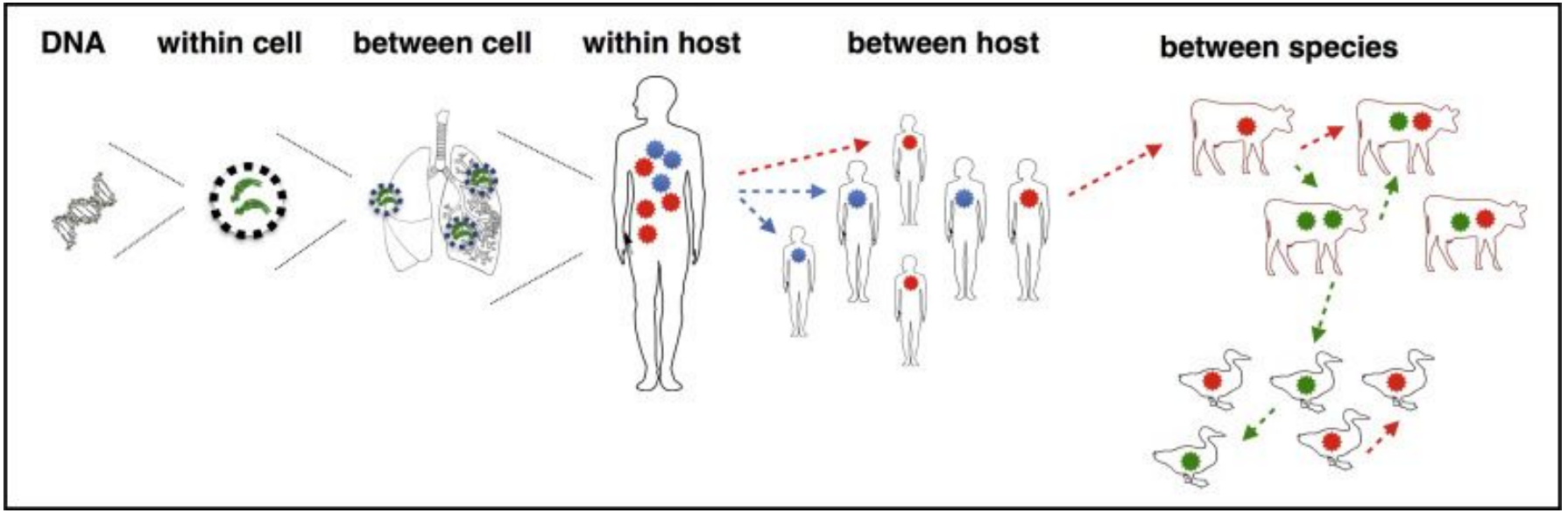
Data does not tell who infected whom:

- **Population structure?**



https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/raw/master/Day2_Phylodynamics_-_Tanja_Stadler.pdf

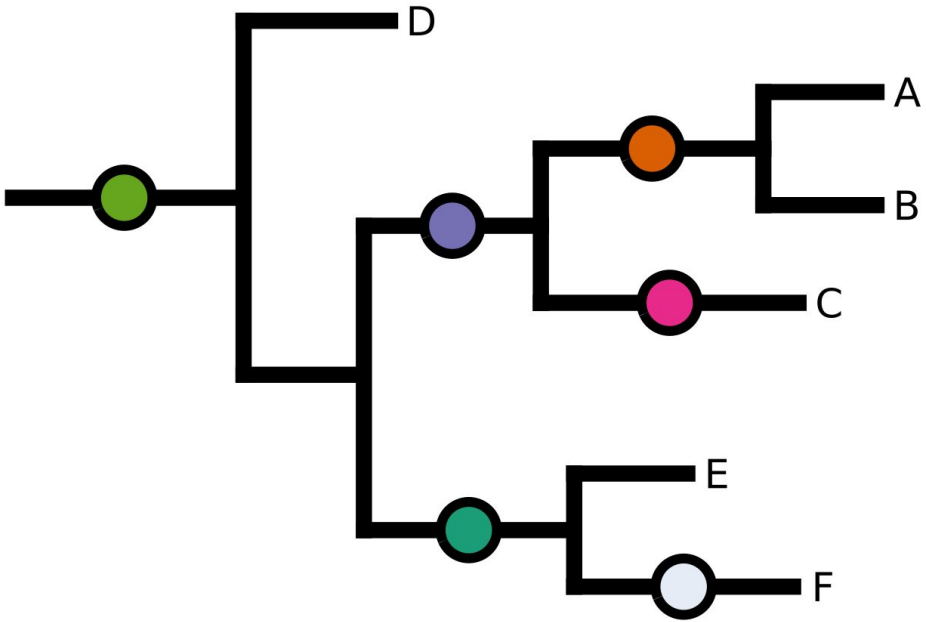
Cases don’t tell you (much) about pathogen evolution



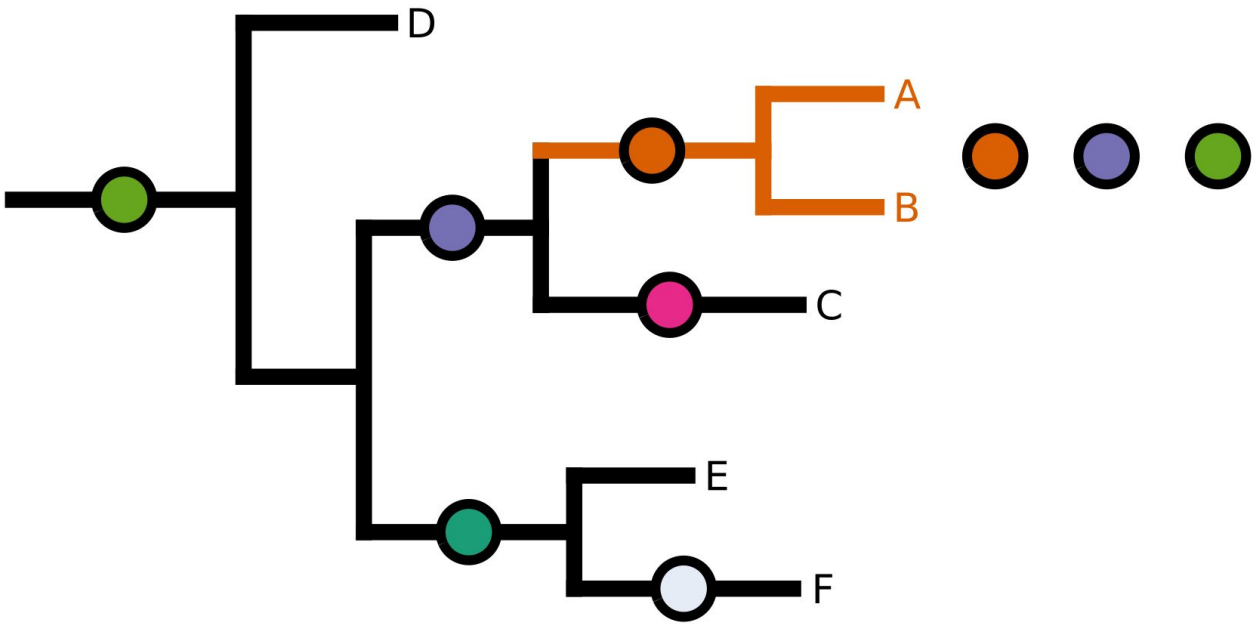
https://www.sciencedirect.com/science/article/pii/S1755436514000723

So, we can use genomes for modelling but how do we actually use all this in infectious disease epidemiology?

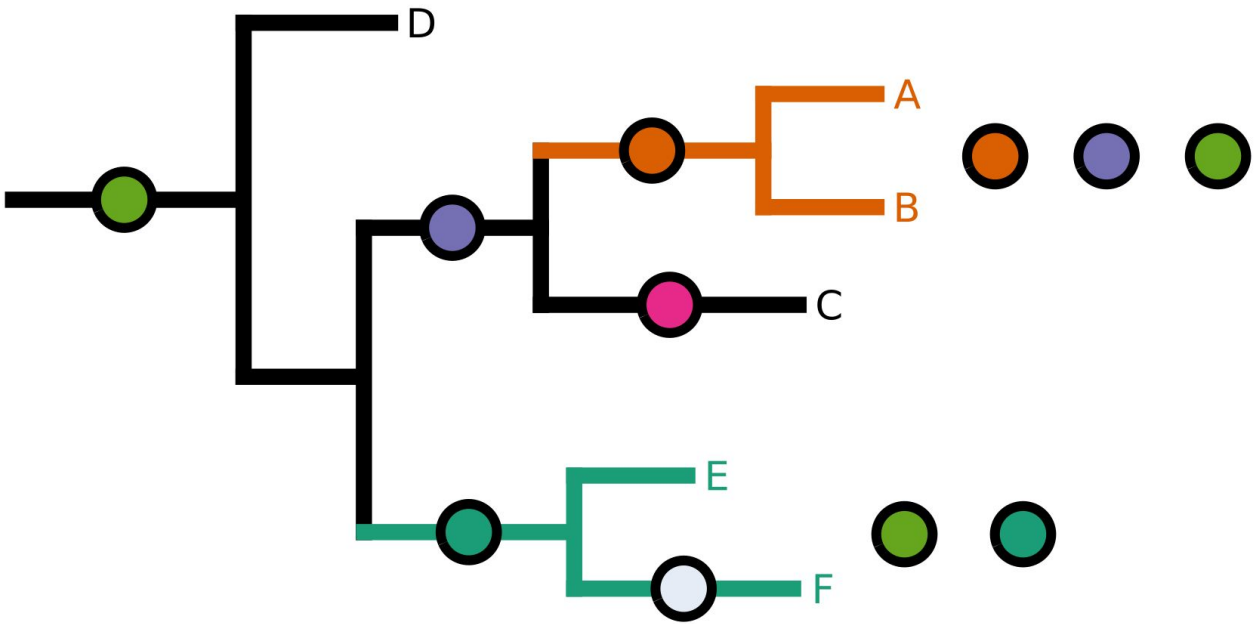
Define lineages (groups) of pathogens



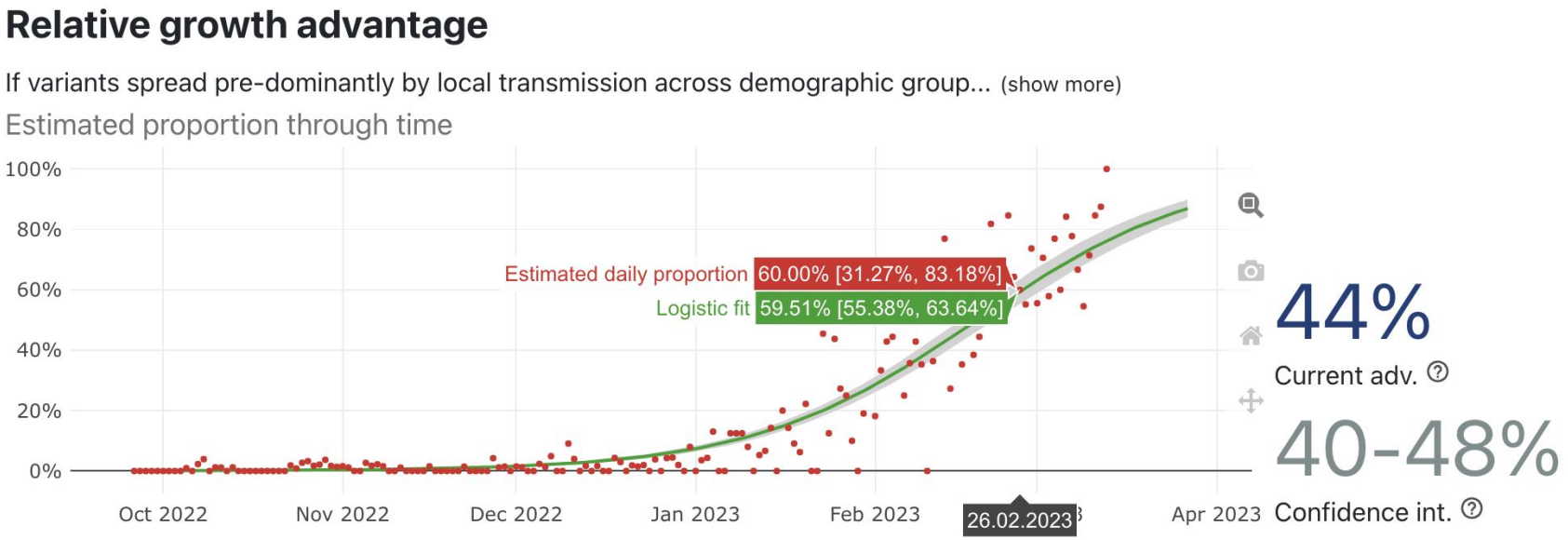
Define lineages (groups) of pathogens



Define lineages (groups) of pathogens

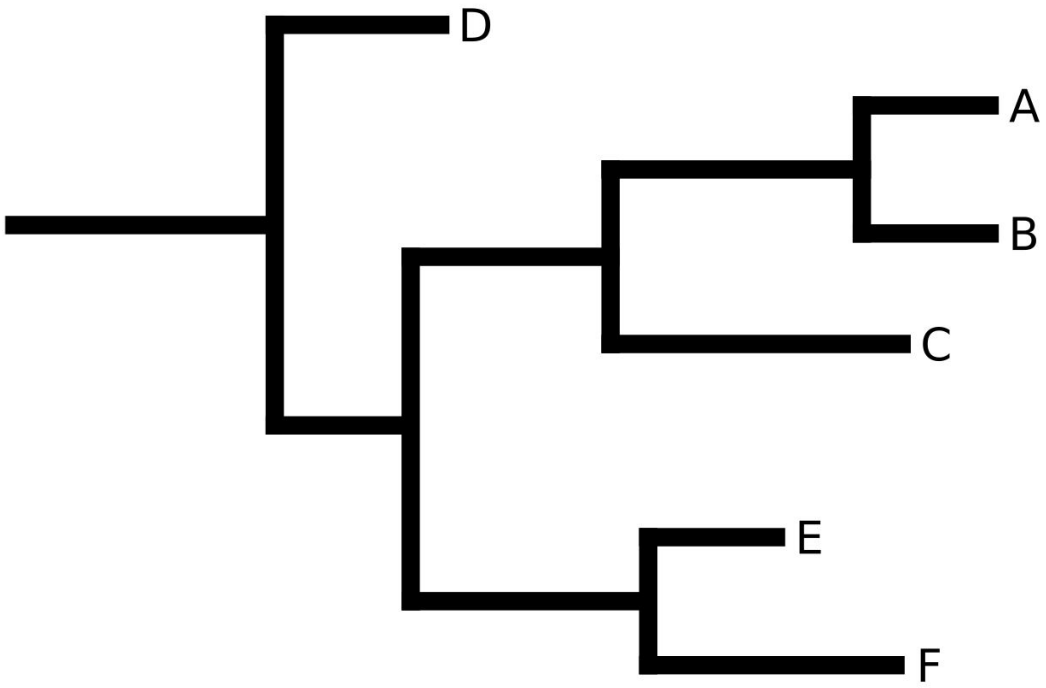


Early warning modelling lineage relative growth advantage

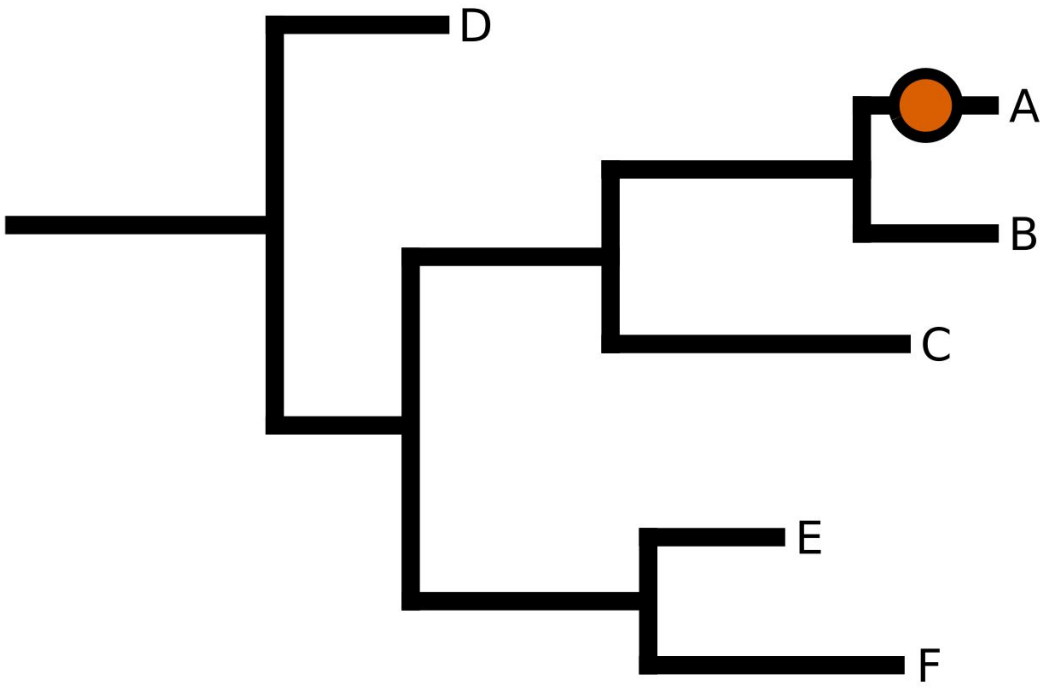


https://cov-spectrum.org/explore/Switzerland/Surveillance/Past6M/variants?nextcladePangoLineage=xbb*&

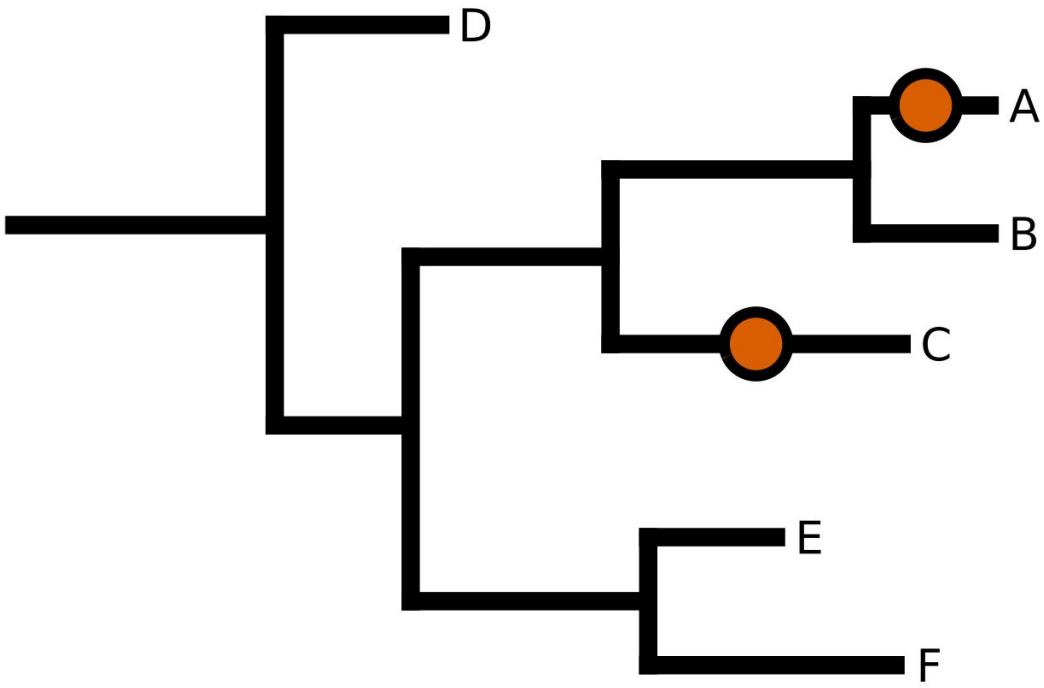
Identify clinically relevant mutations



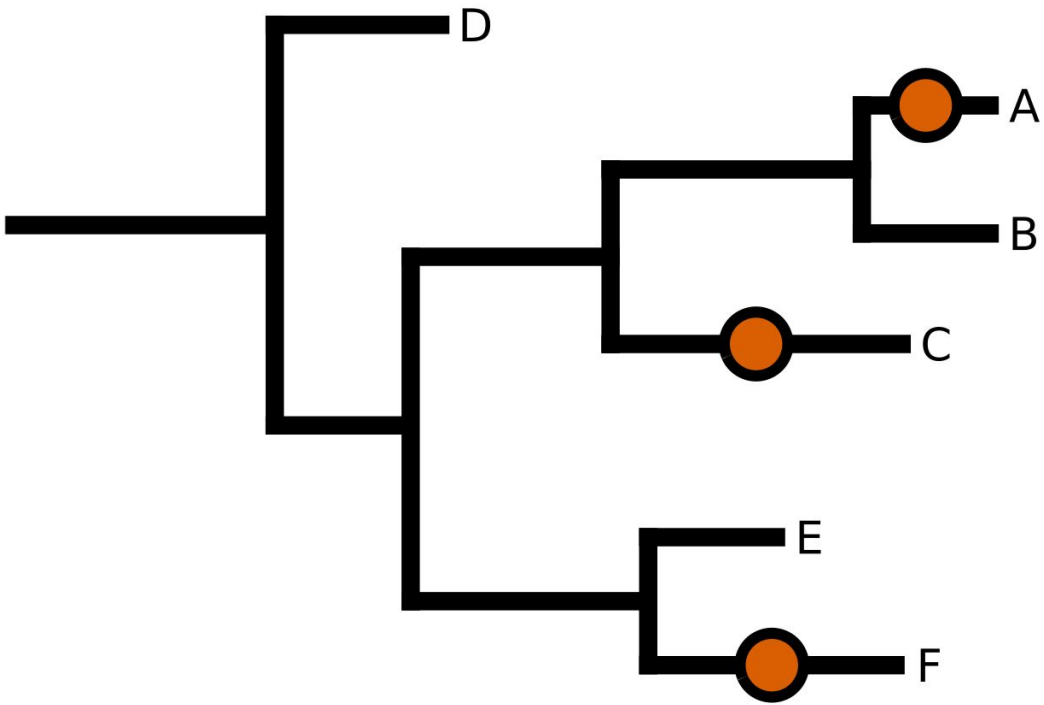
Identify clinically relevant mutations



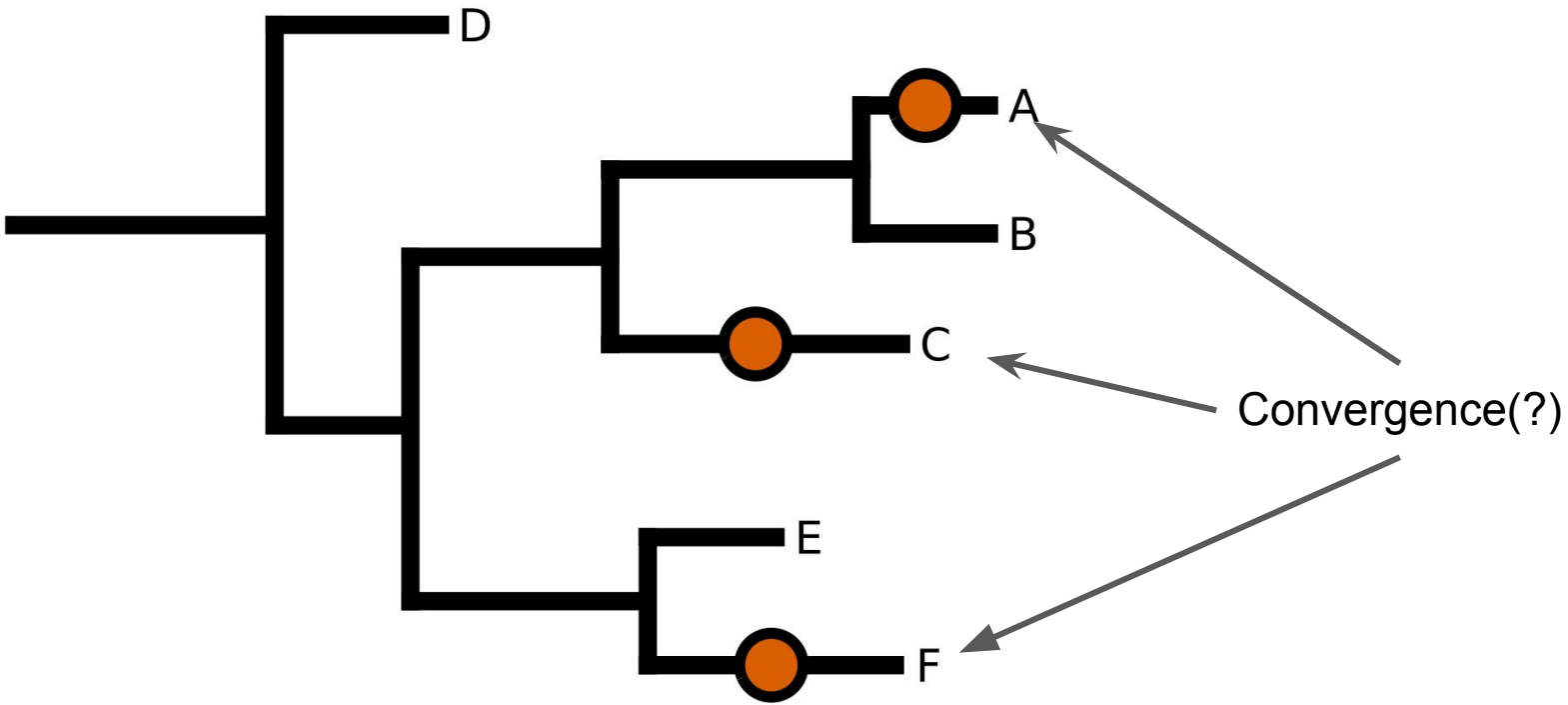
Identify clinically relevant mutations



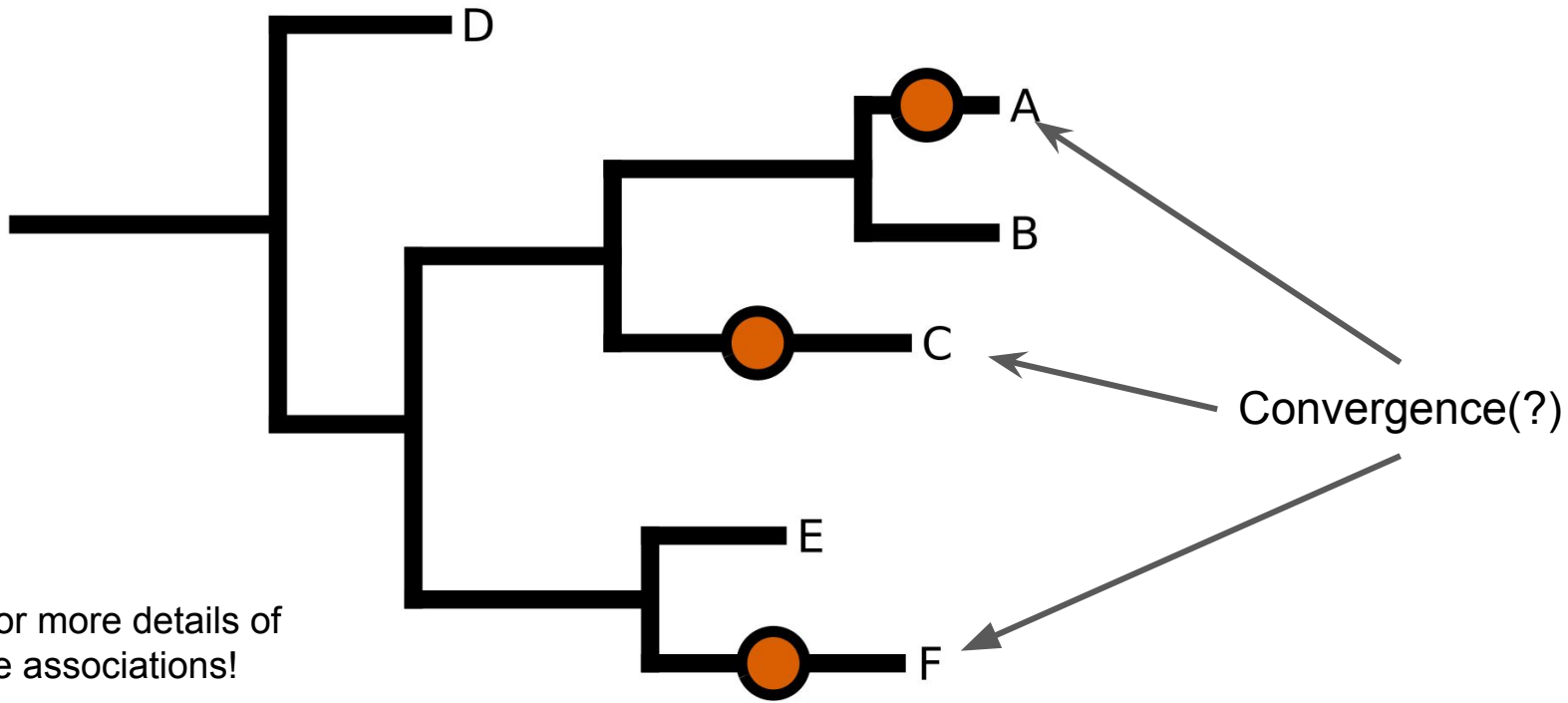
Identify clinically relevant mutations



Identify clinically relevant mutations

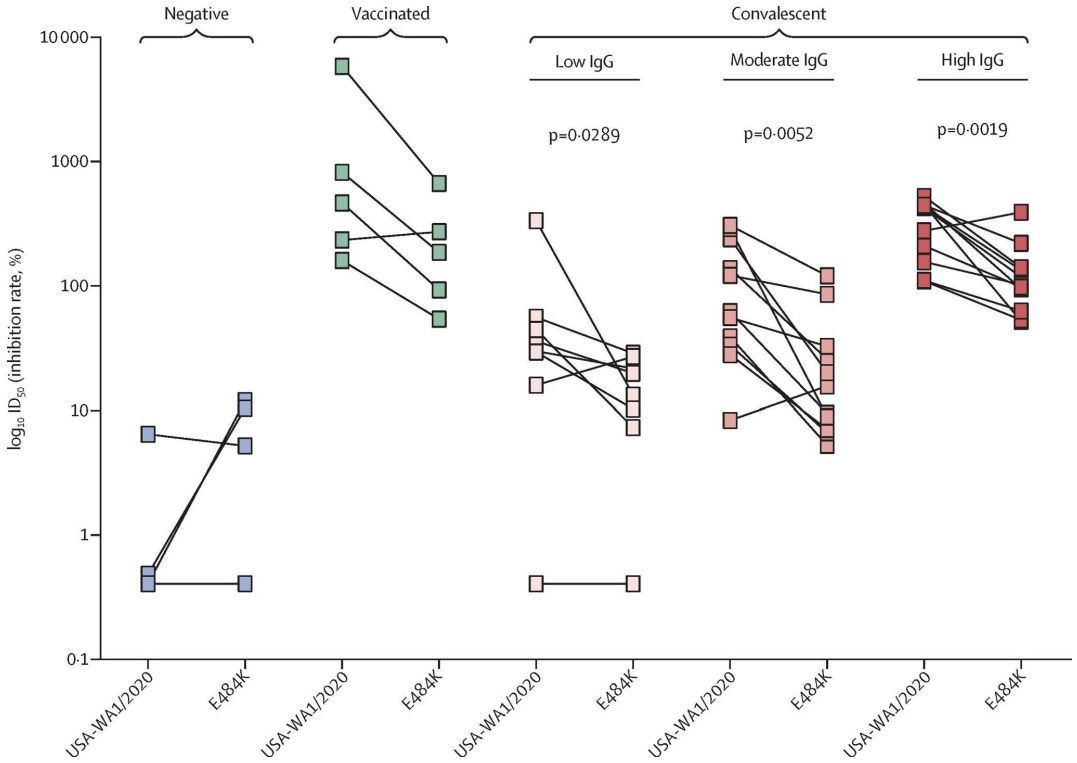


Identify clinically relevant mutations



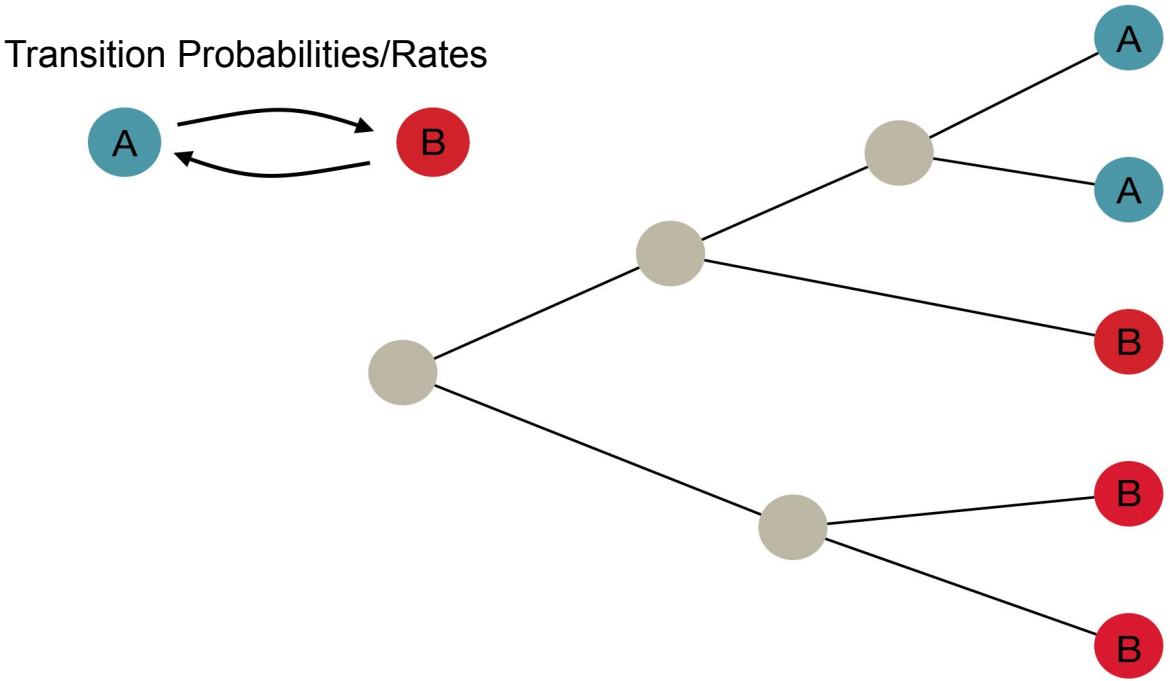
See part 2 for more details of testing these associations!

Prioritise characterisation of mutations: S:E484K



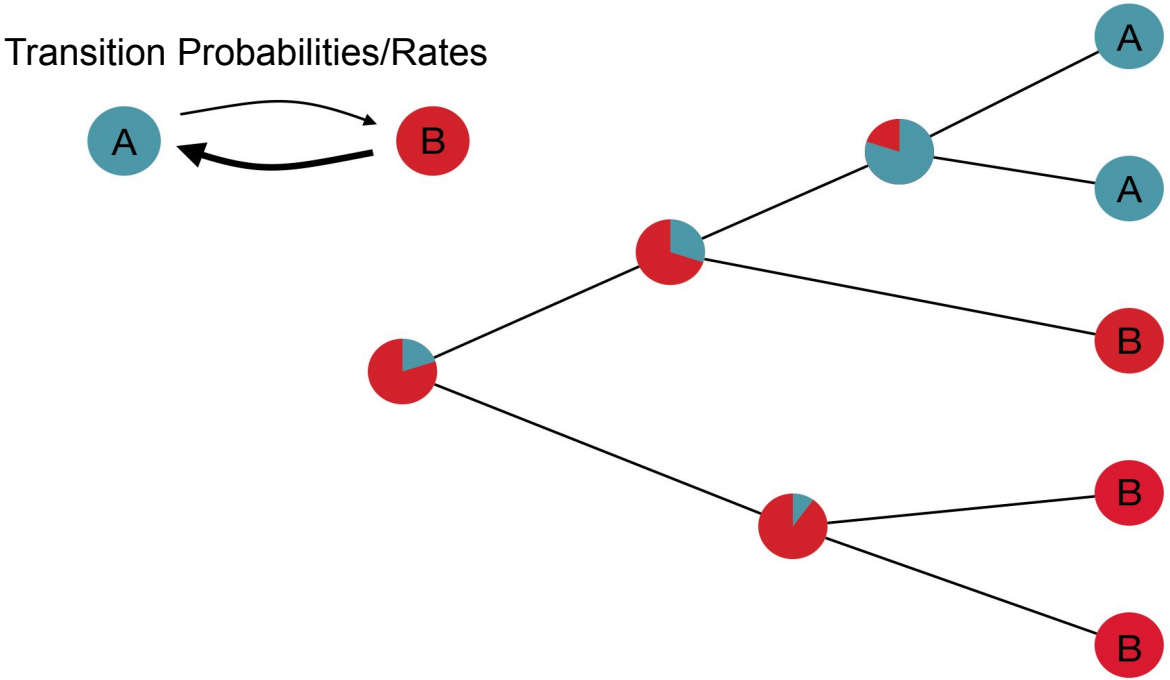
https://www.thelancet.com/journals/lanmic/article/PIIS2666-5247(21)00068-9

Inferring internal ancestral states from observed tips



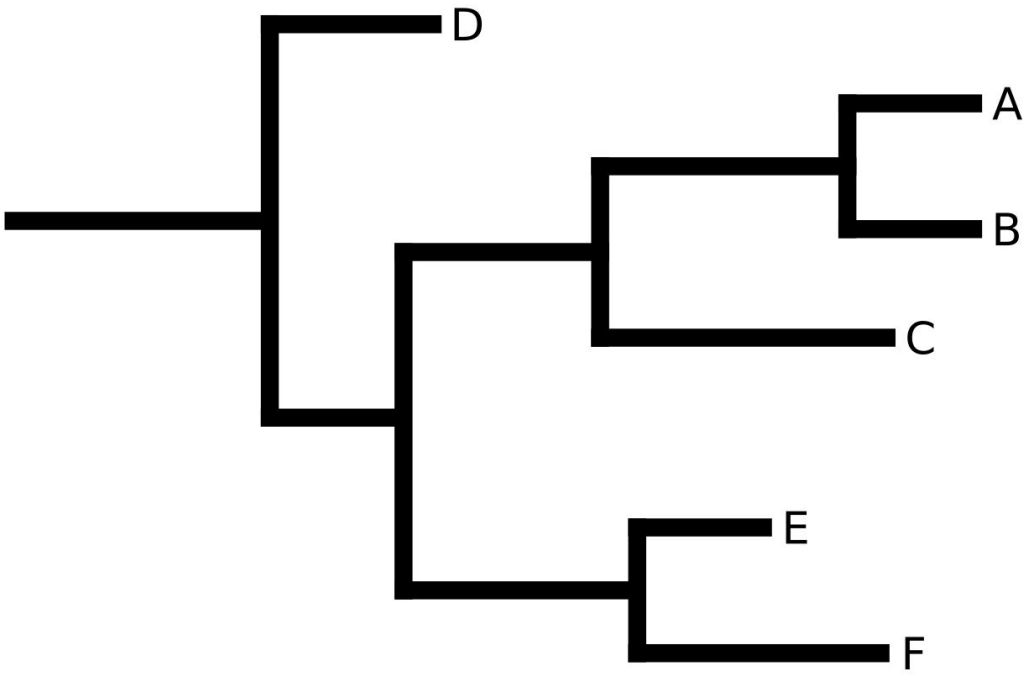
https://github.com/trvrb/gs541-phylodynamics

Inferring internal ancestral states from observed tips

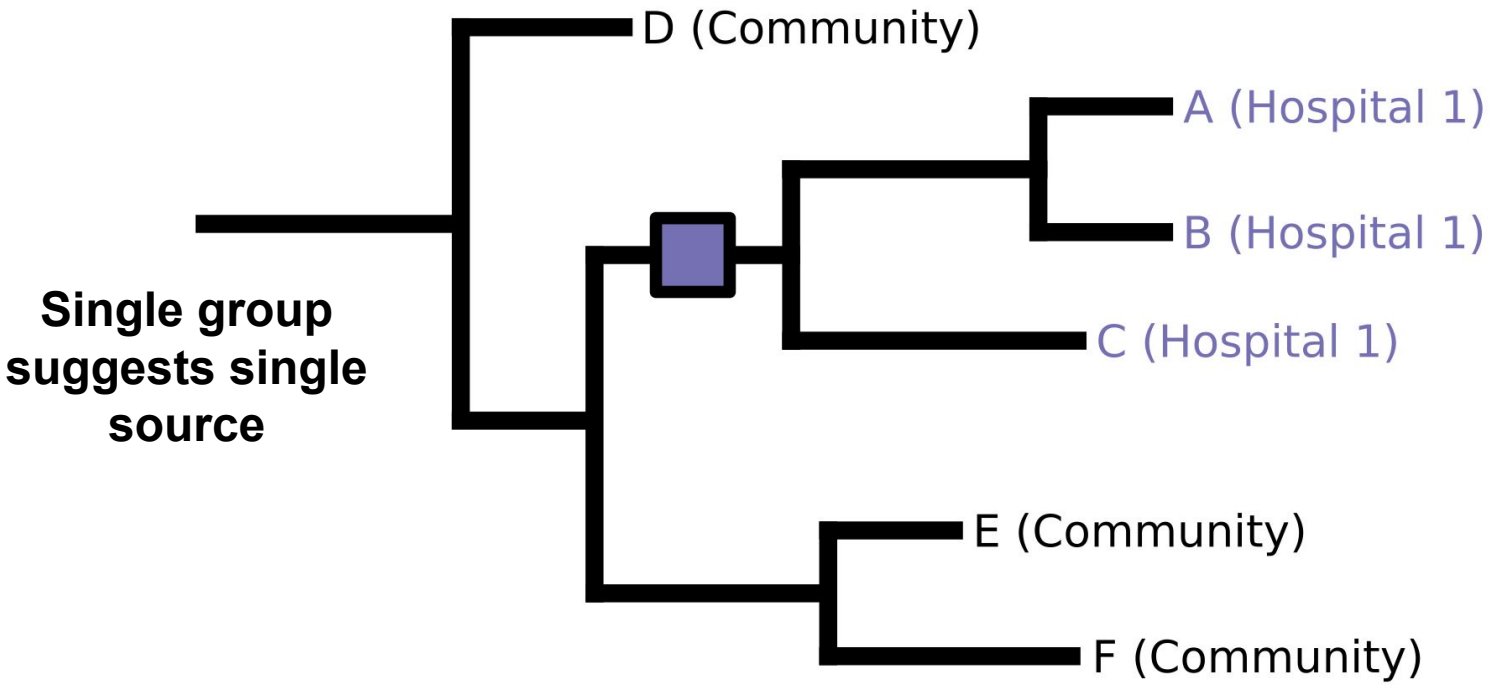


https://github.com/trvrb/gs541-phylodynamics

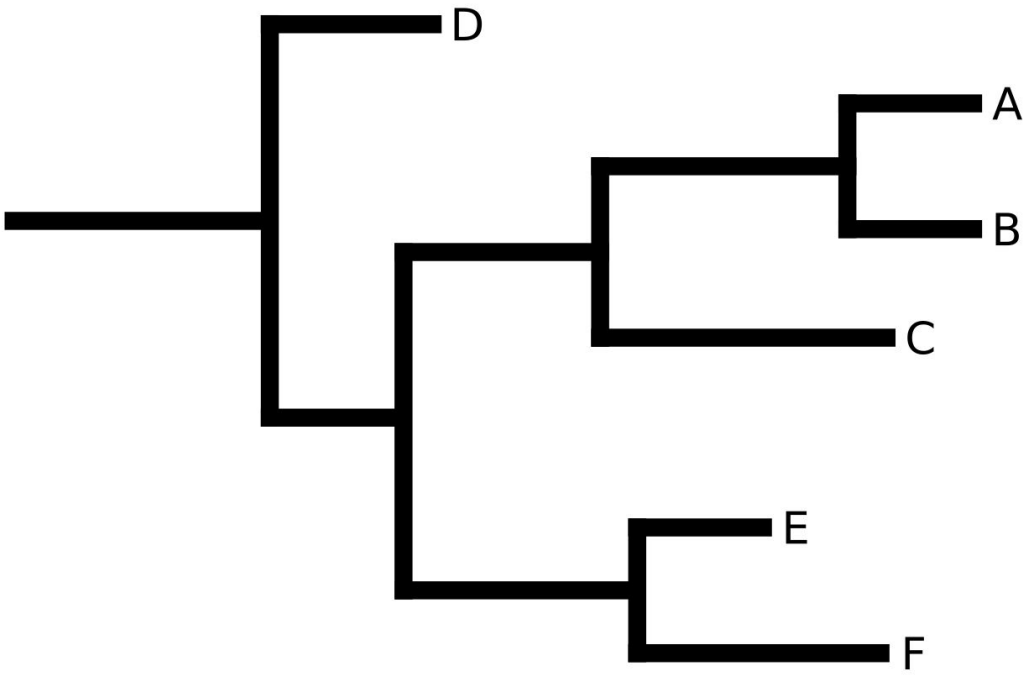
Trace sources of outbreaks



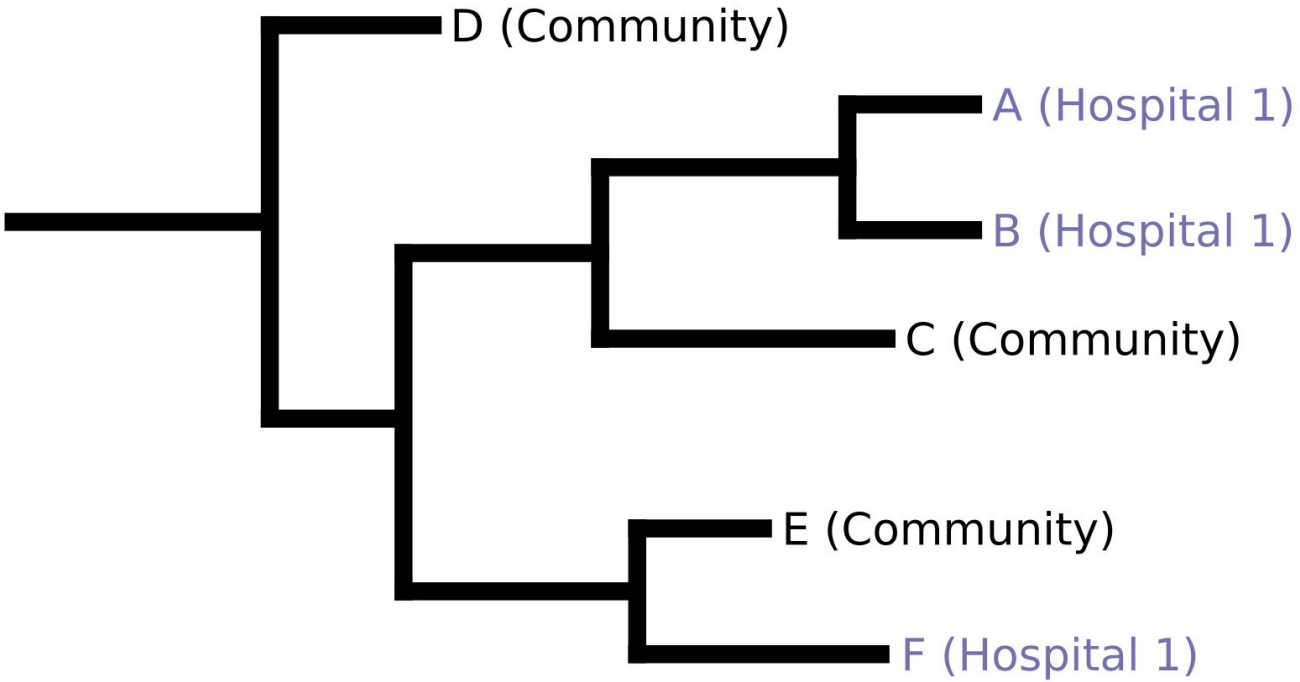
Trace sources of outbreaks



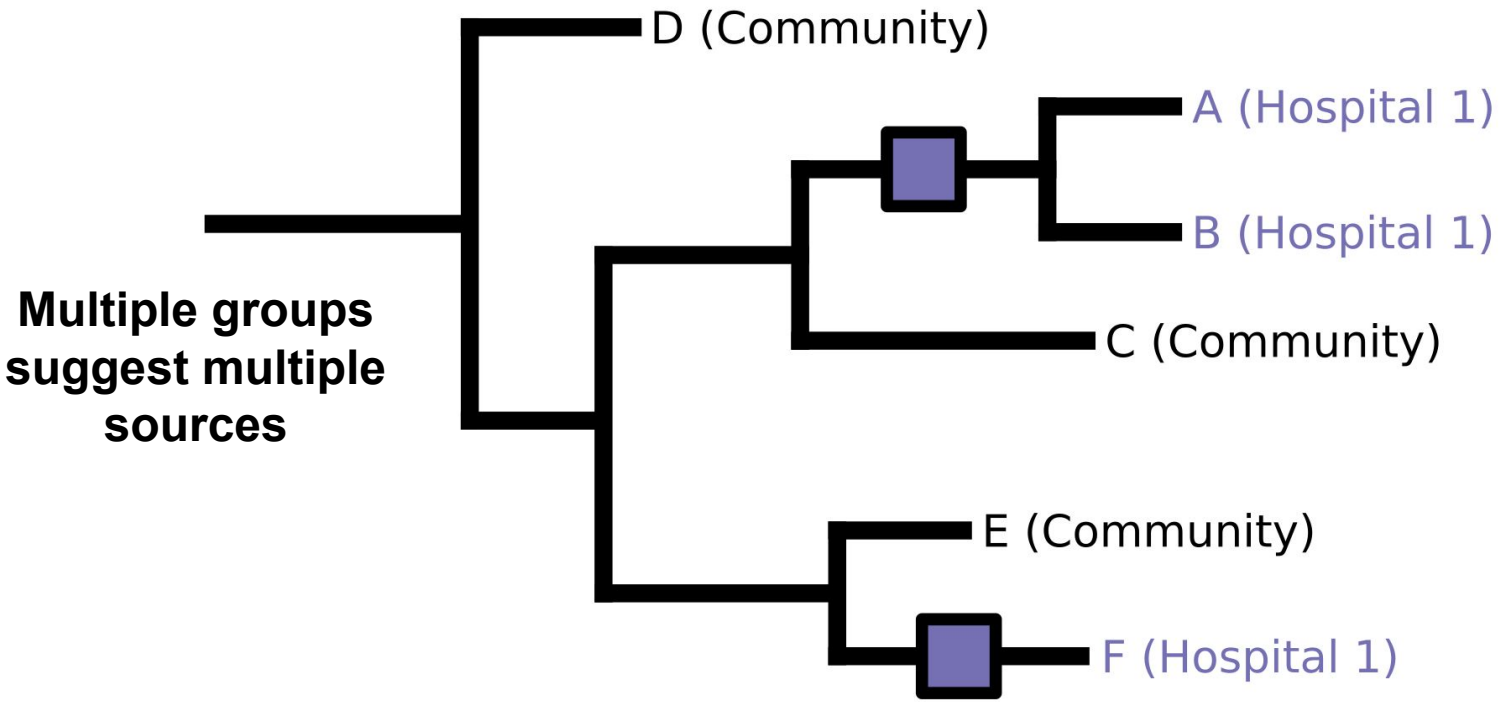
Trace sources of outbreaks



Trace sources of outbreaks

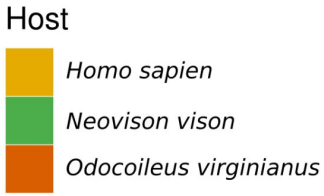


Trace sources of outbreaks



Multiple groups suggest multiple sources

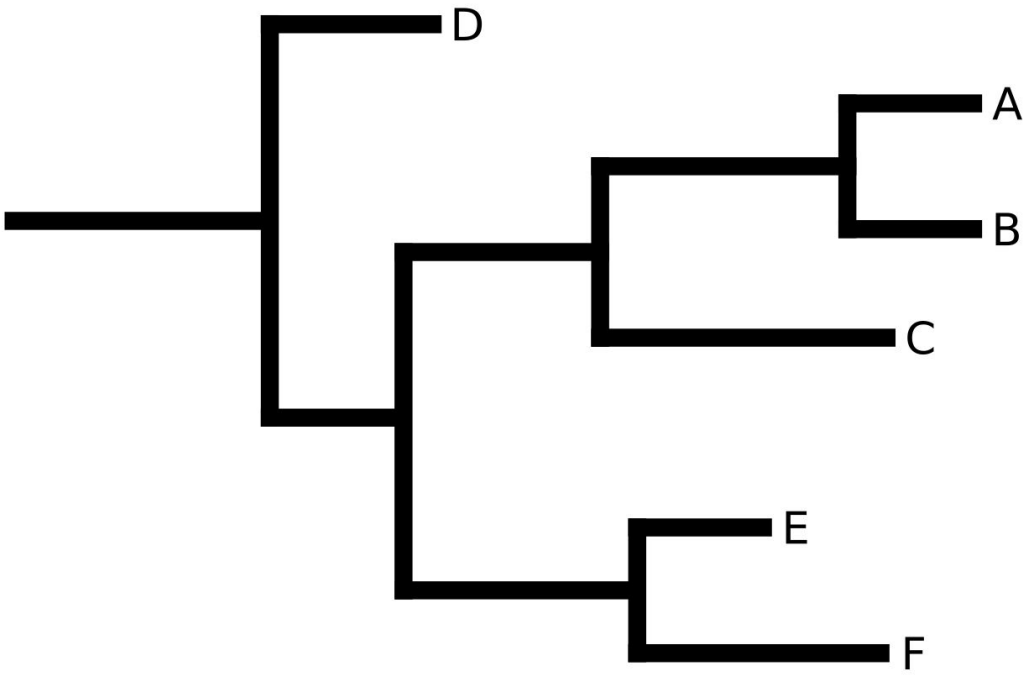
Finding Deer-to-Human transmission



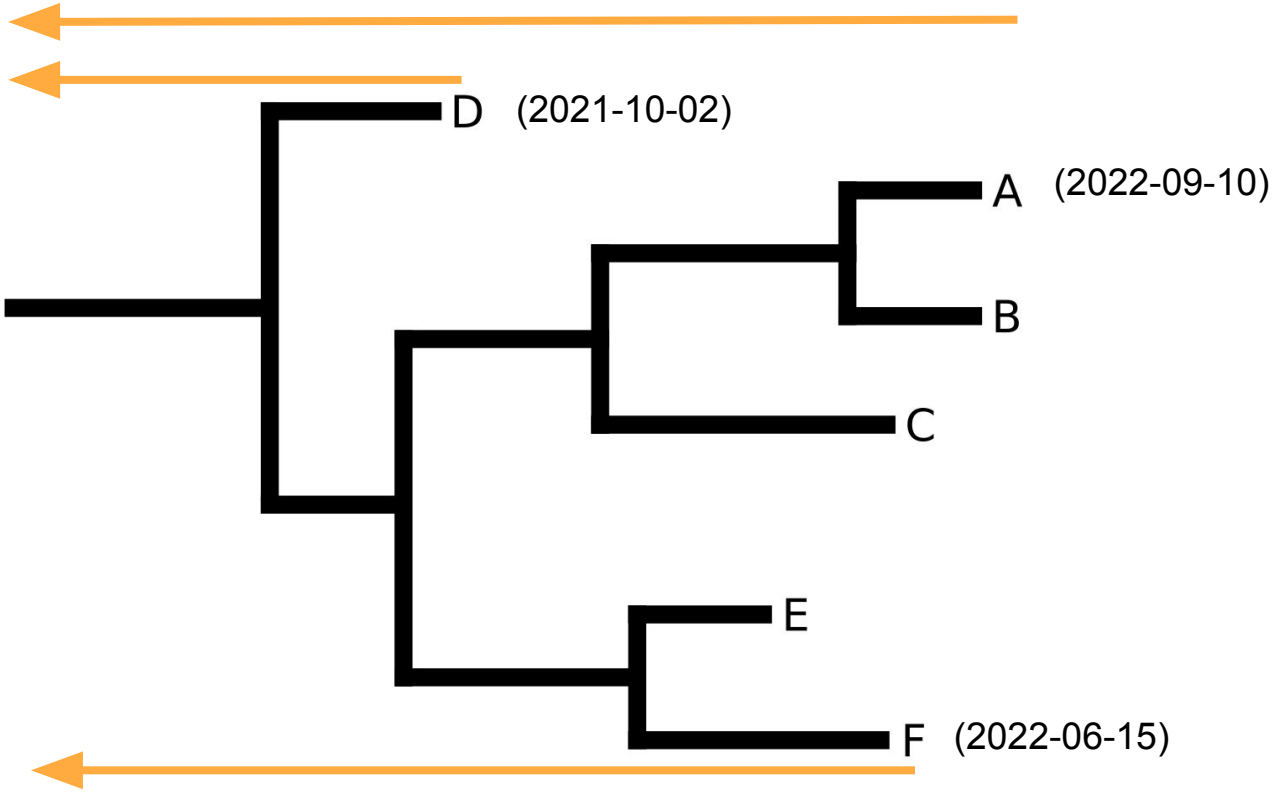
Supporting Circumstantial Evidence:

- Spatially congruent
- Temporally congruent
- Plausible Epidemiological Link

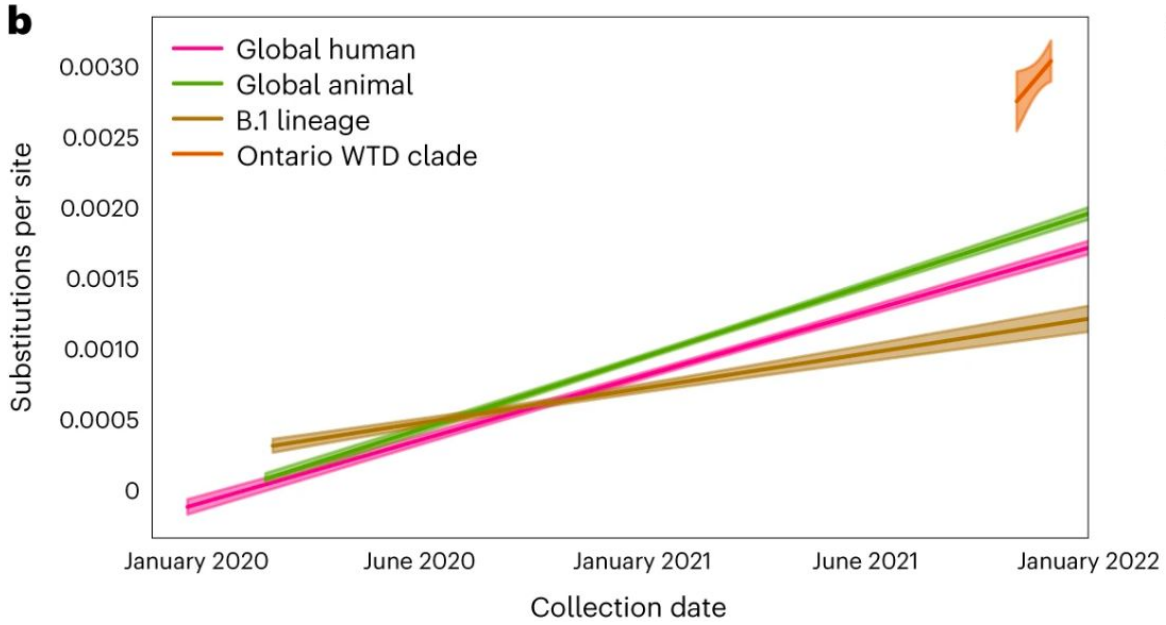
Convert genomic distance to time



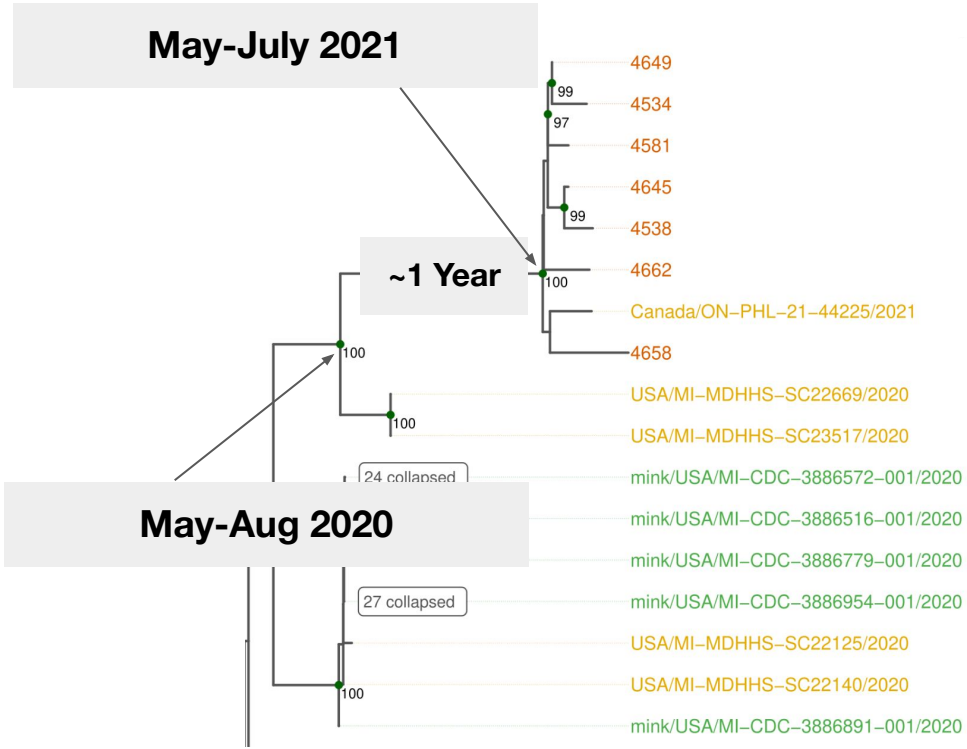
Convert genomic distance to time

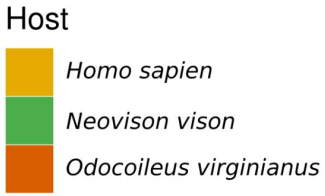


Mutation rates from Root to Tip Regression

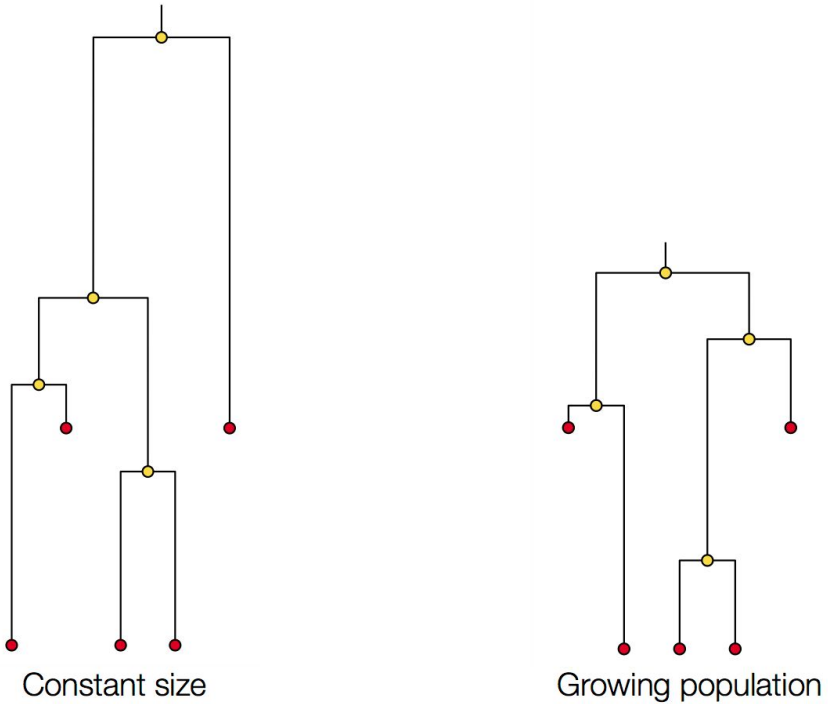


Mutations rates let us time of unobserved events



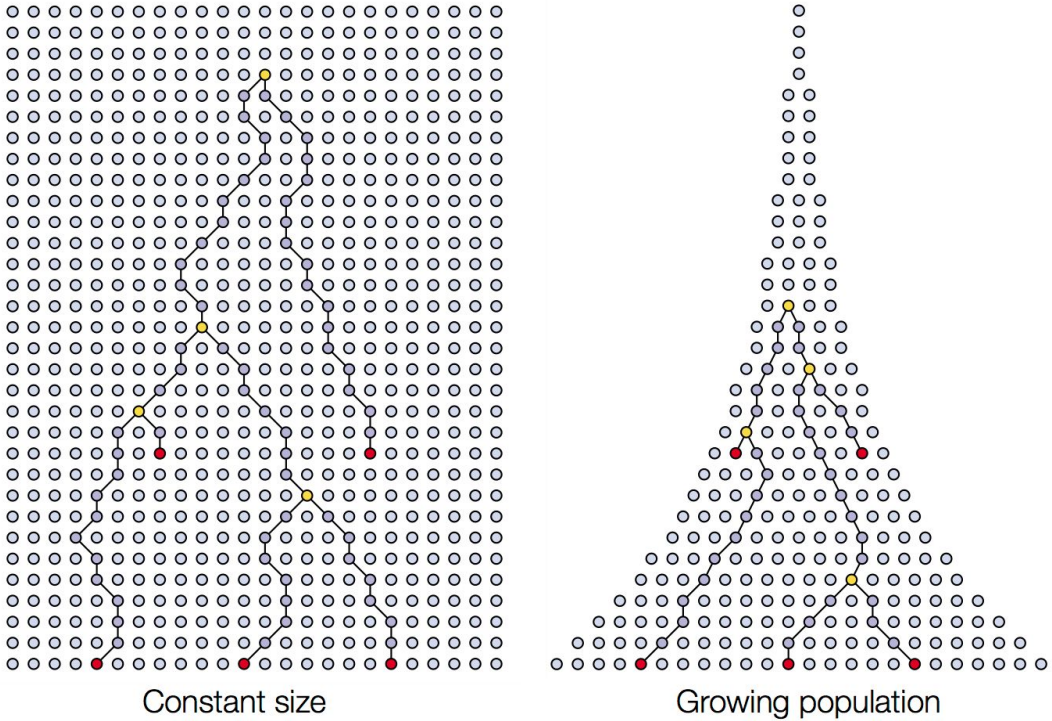


Tree shape tells us about population size



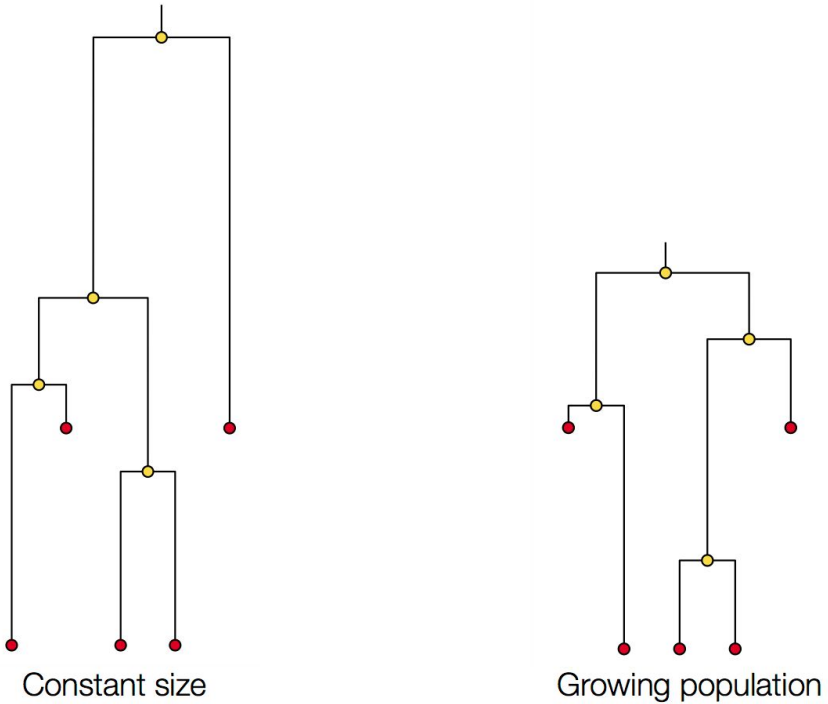
https://github.com/trvrb/gs541-phylodynamics

Tree shape tells us about population size



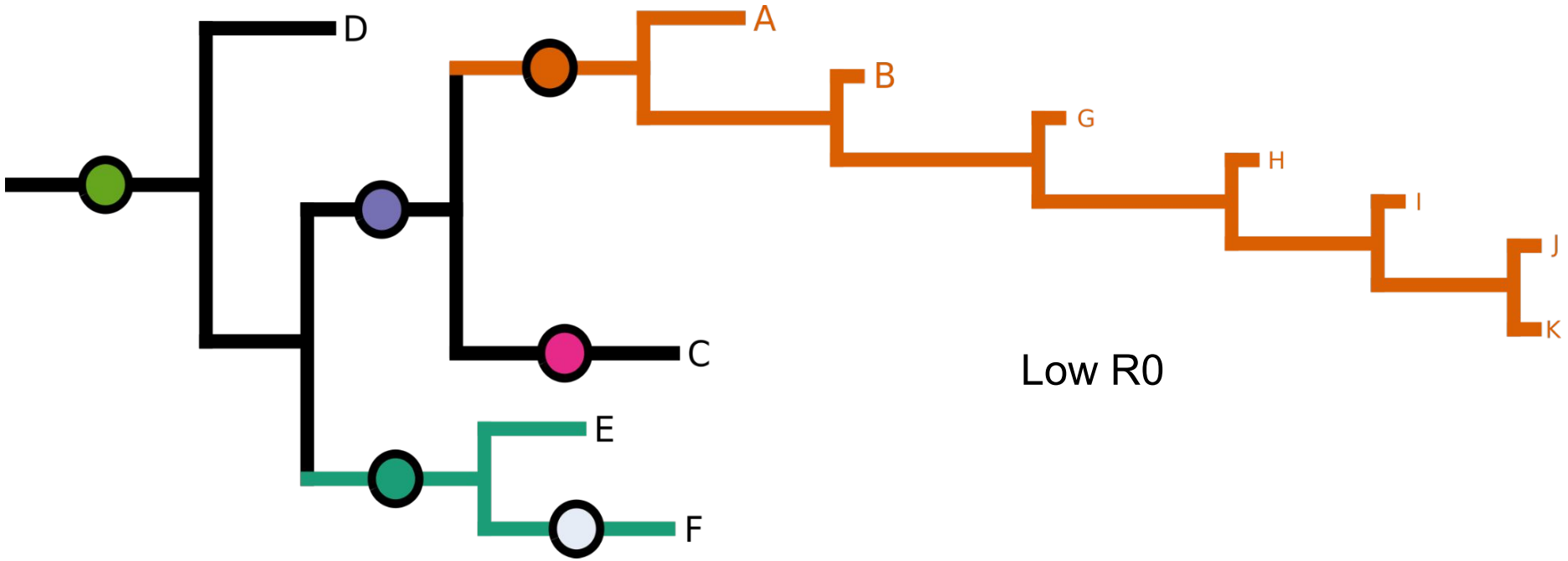
https://github.com/trvrb/gs541-phylodynamics

Tree shape tells us about population size

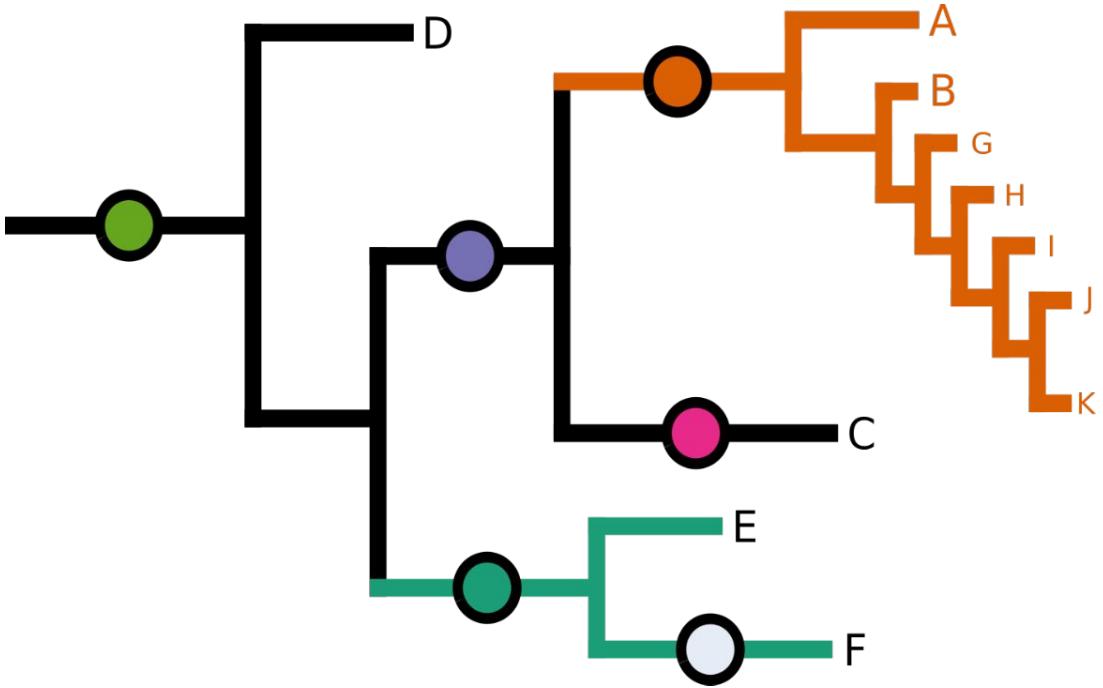


https://github.com/trvrb/gs541-phylodynamics

Can calculate epidemiological parameters from shape

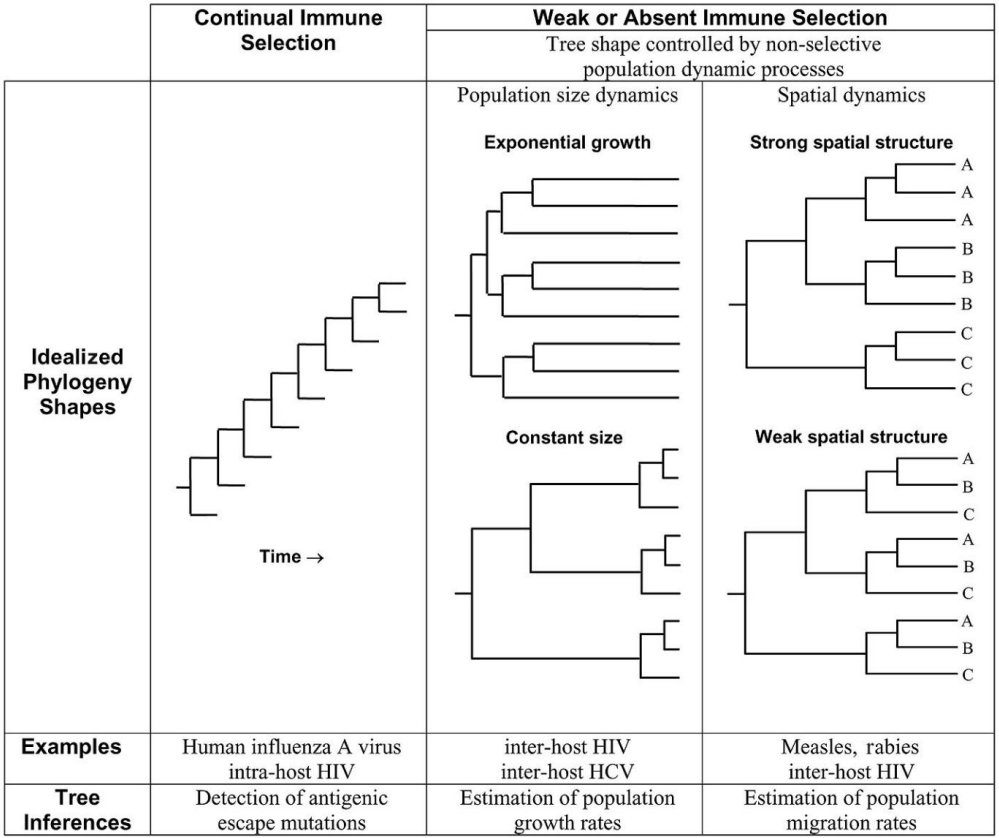


Can calculate epidemiological parameters from shape



High R0

Many other analyses are possible



Massive area:

- Birth-death models
- Coalescent models
- Bayesian skyline/skygrid models
- Spatiotemporal models (phylogeography)
- Recombination
- Inference of selection pressure

Grenfell, B. T. (2004). Unifying the Epidemiological and Evolutionary Dynamics of Pathogens. Science, 303(5656), 327–332. doi:10.1126/science.1090727

Summary

- Pathogen **evolution** and **epidemiology** are intrinsically linked
- Genomics provides insights into **evolution and unobserved events**
- Comparison of DNA sequences to databases can be used for **diagnostics**
- Pattern of mutations across genomes can be used to generate **phylogenies**
- Phylogenies are structured by **sampling**, **ecology, evolution, and epidemiology**
- Probabilistic **Bayesian phylogenetic inference** is a key tool
- Can use these approaches to do many things including:
  - Identify **lineages**
  - Surveille **evolution**
  - Infer **timing/location** of outbreaks/events
  - Determine **epidemiological parameters**