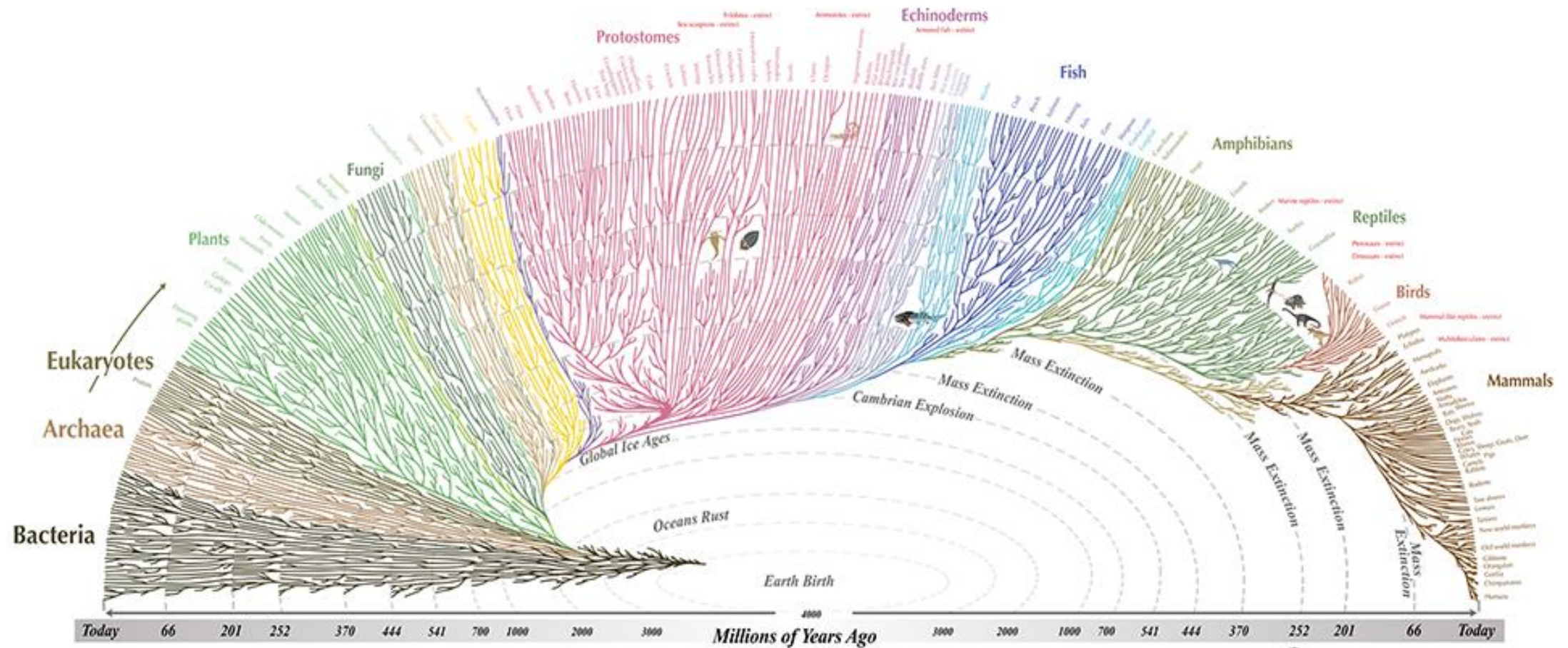


A background image featuring a complex phylogenetic tree with numerous nodes. The nodes are colored in various shades including purple, yellow, red, green, and blue, set against a light grey tree structure.

Phylogenetics & COVID Outbreak

Sara Fritz and Rebecca Wright

What is Phylogenetics?



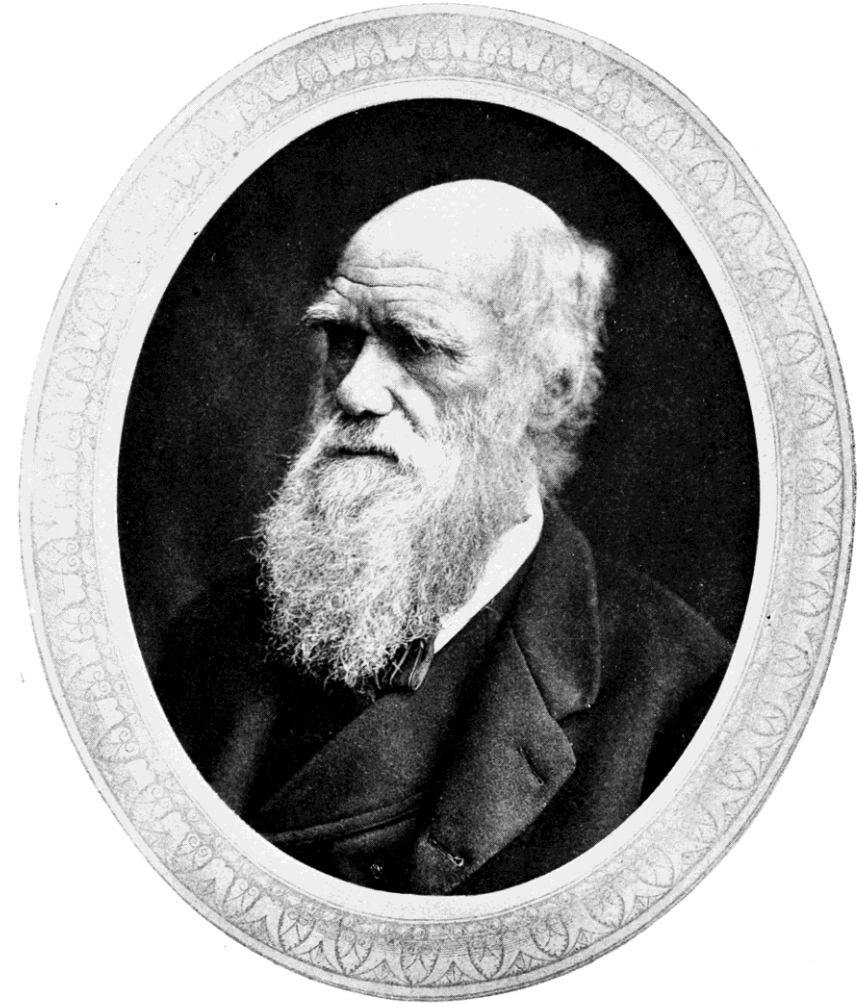
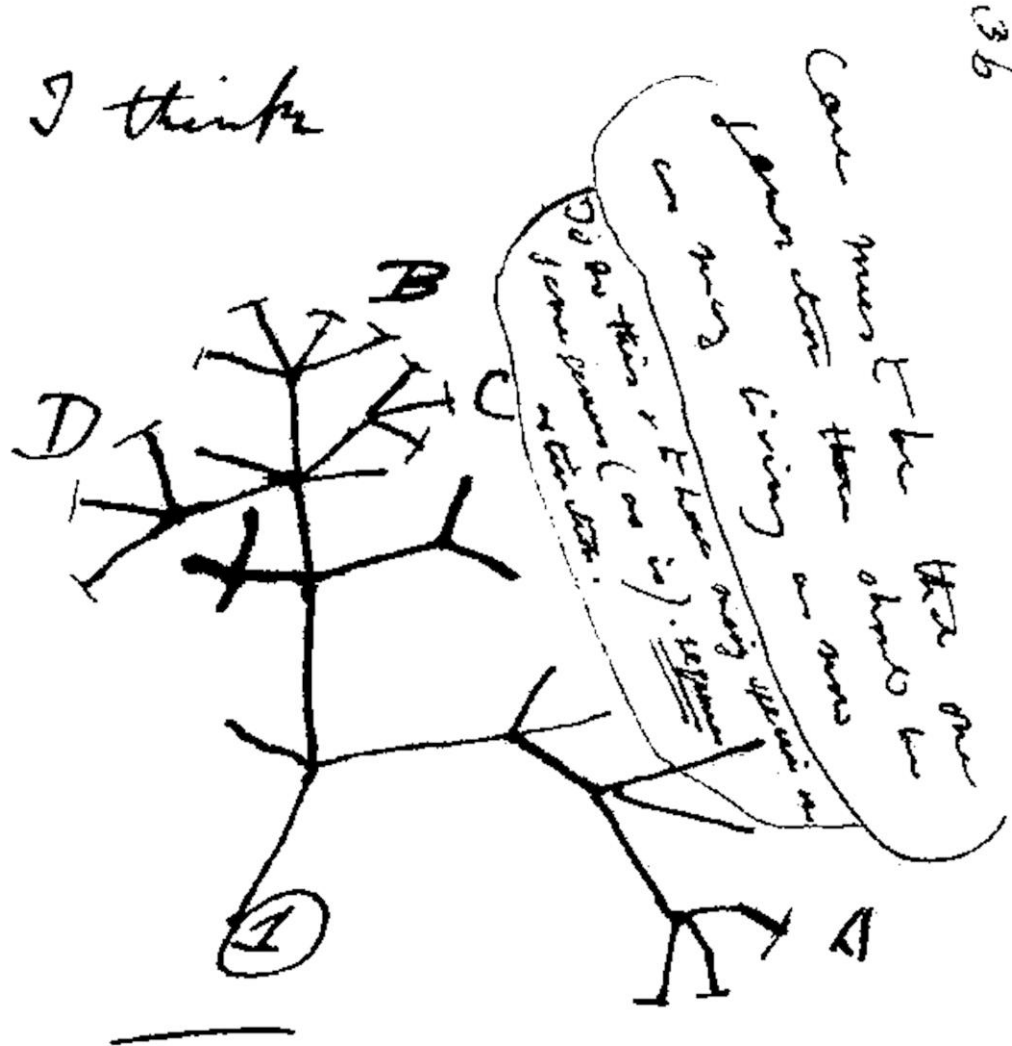
All the major and many of the minor living branches of life are shown on this diagram, but only a few of those that have gone extinct are shown. Example: Dinosaurs - extinct



© 2008, 2017 Leonard Eiseberg. All rights reserved.
eiseberg.com

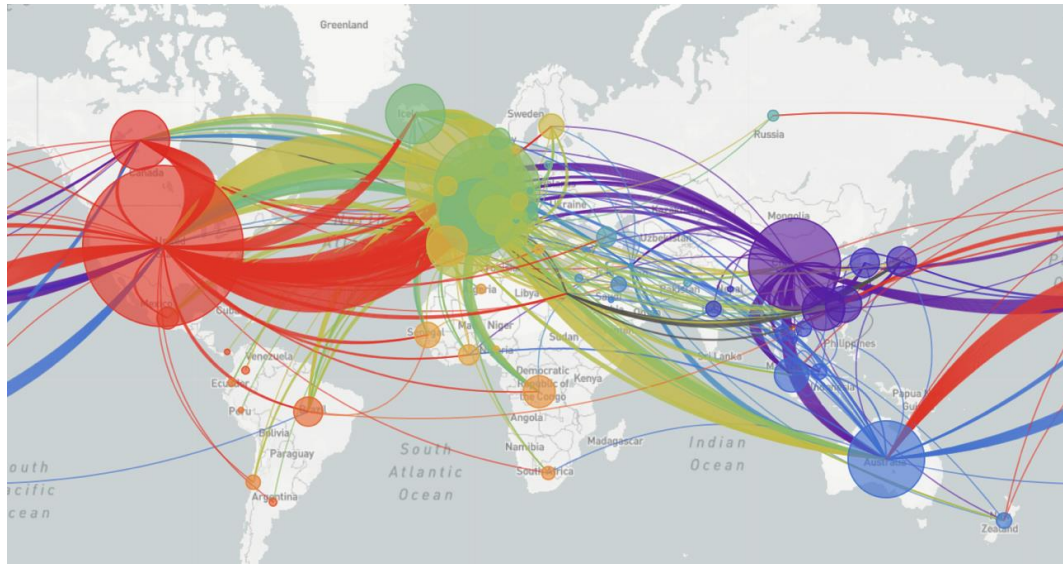
The study evolutionary relationship of biological entities

Who created modern phylogenetics?

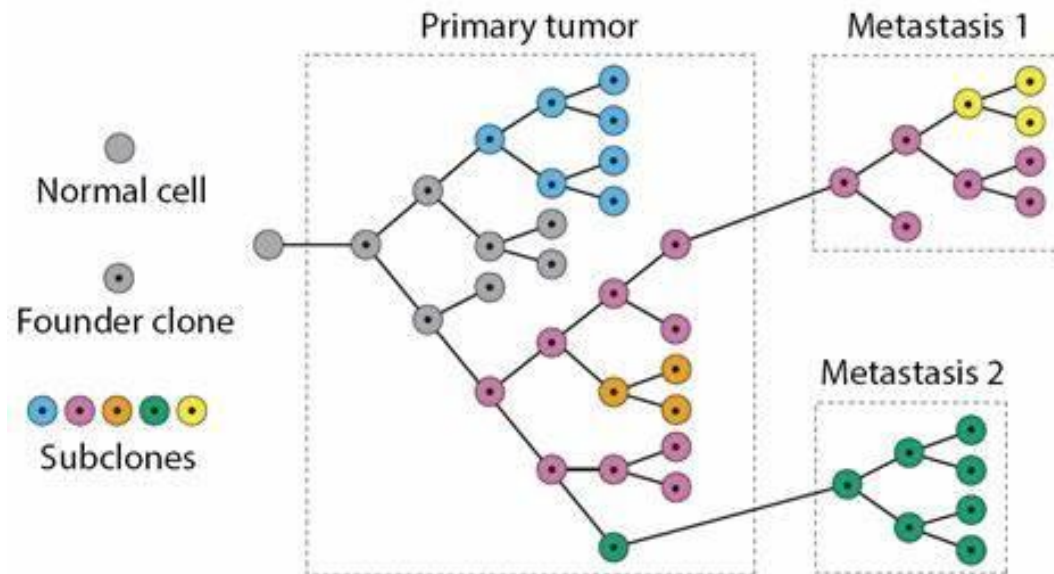


Why is phylogenetics important?

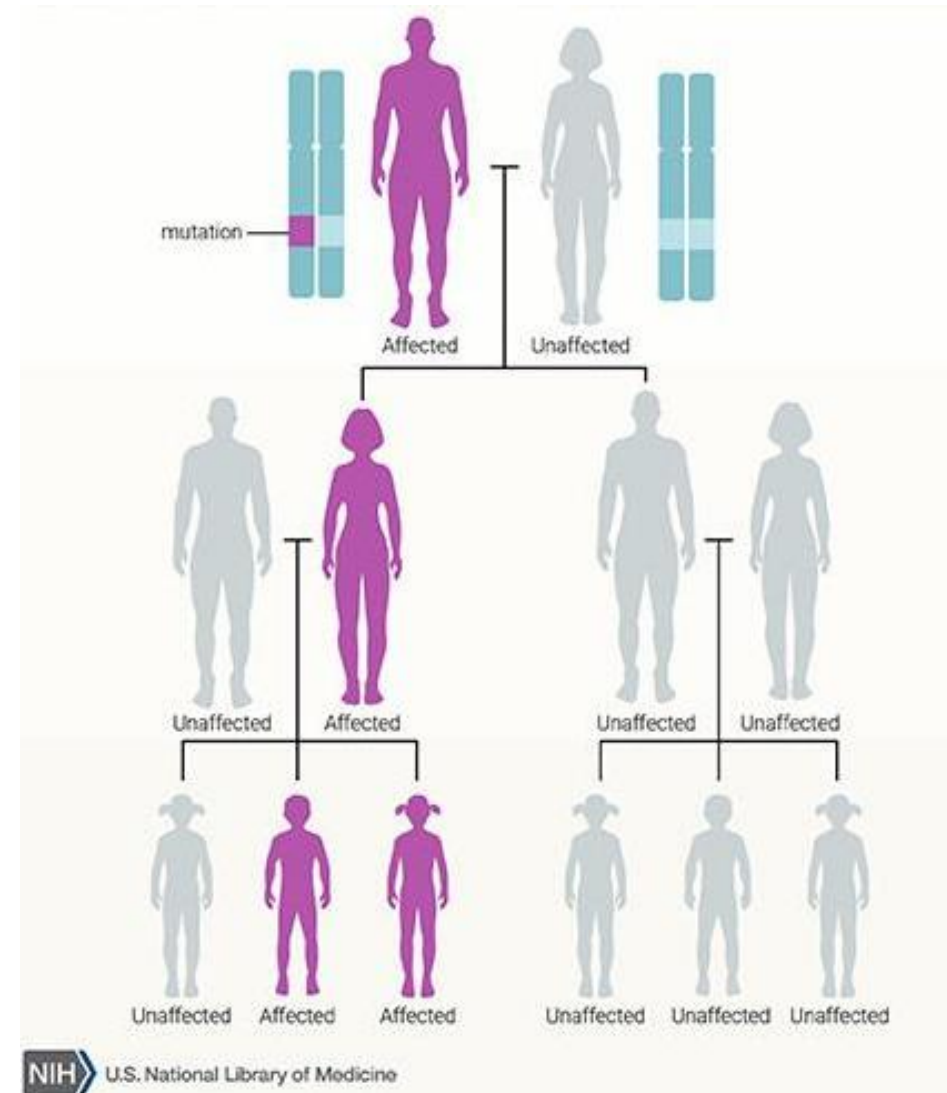
Viral Outbreaks



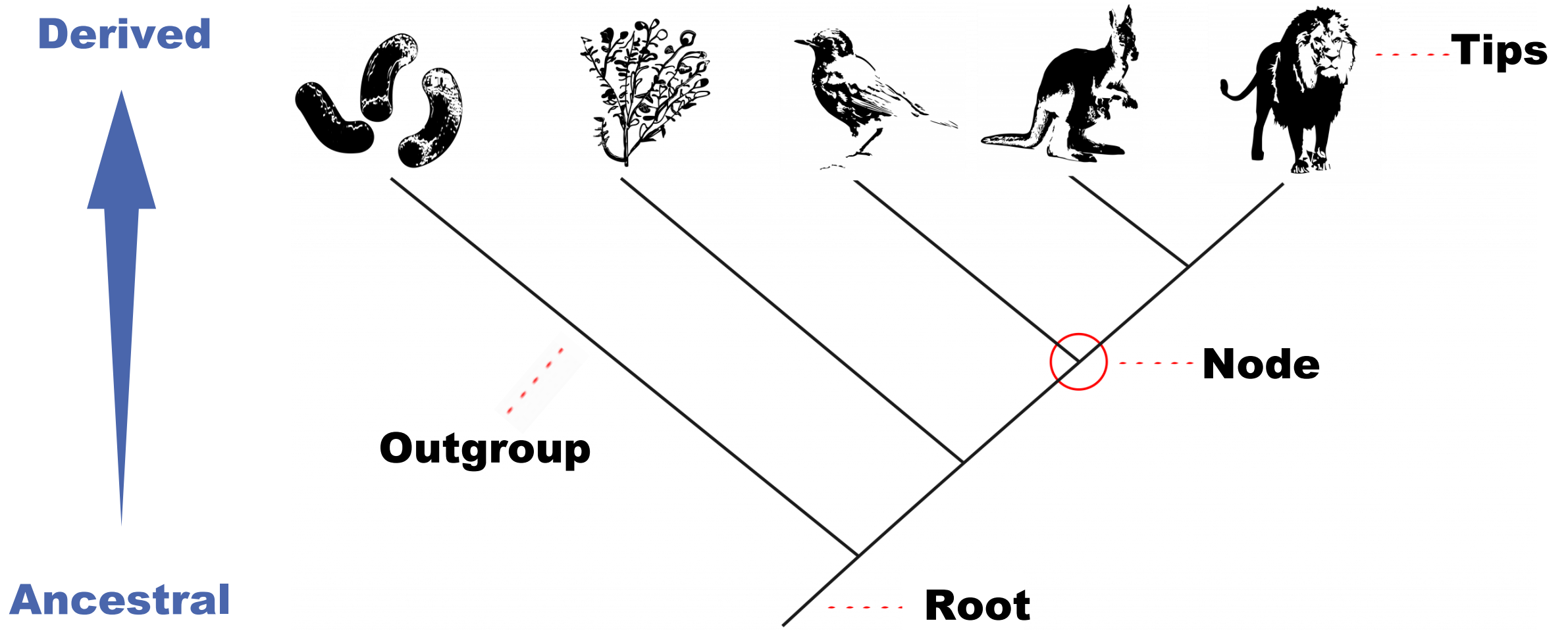
Studying Cancer



Disease genes

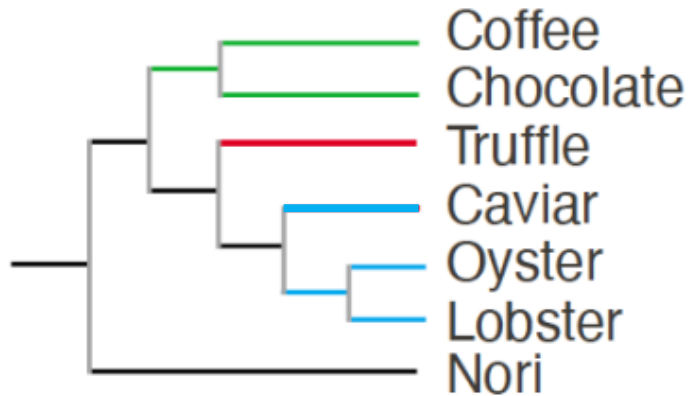
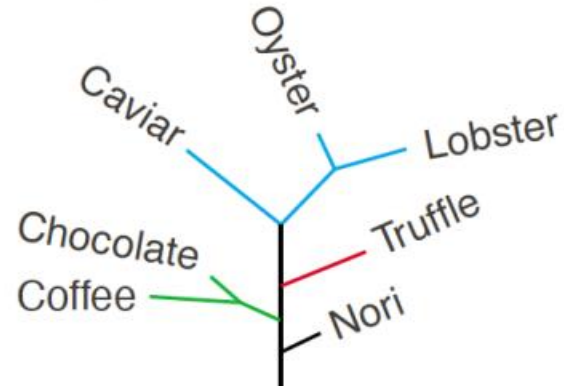
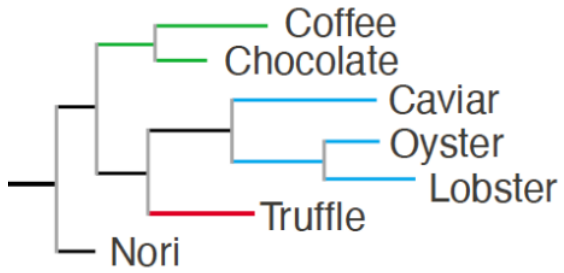


What is the purpose a phylogenetic tree?

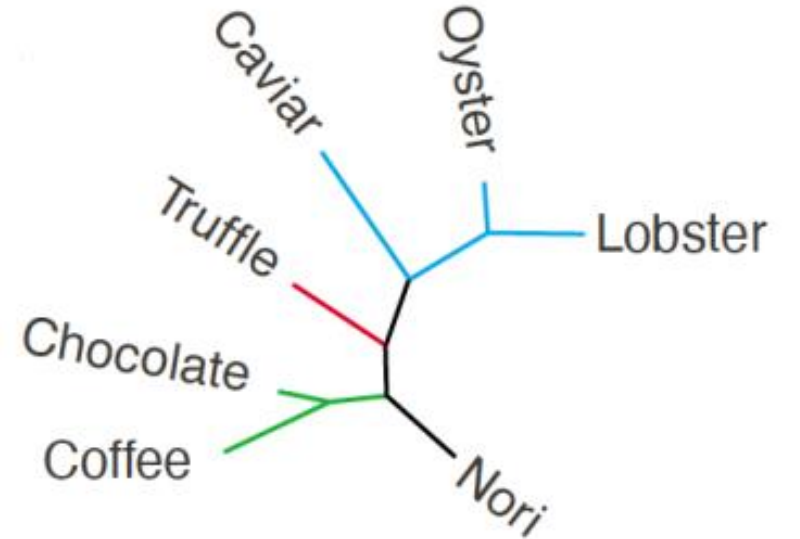


Multiple Tree Shapes

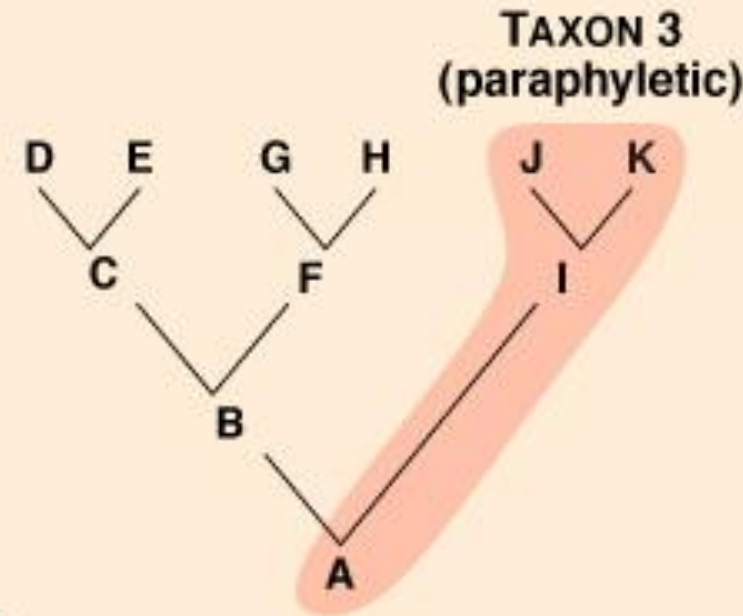
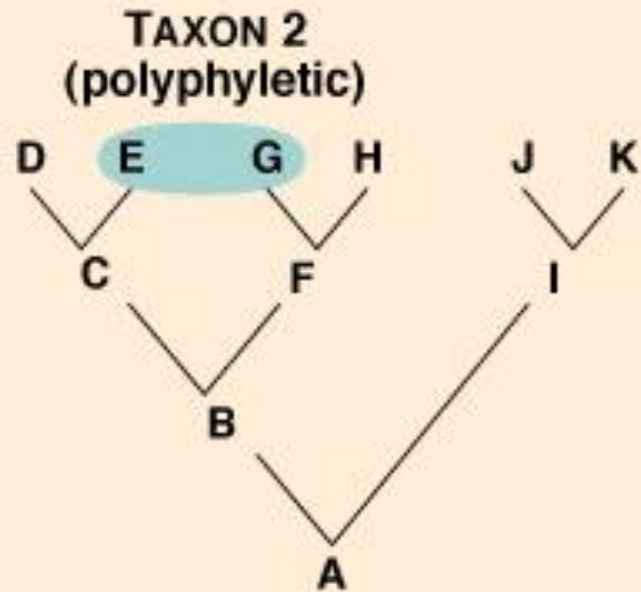
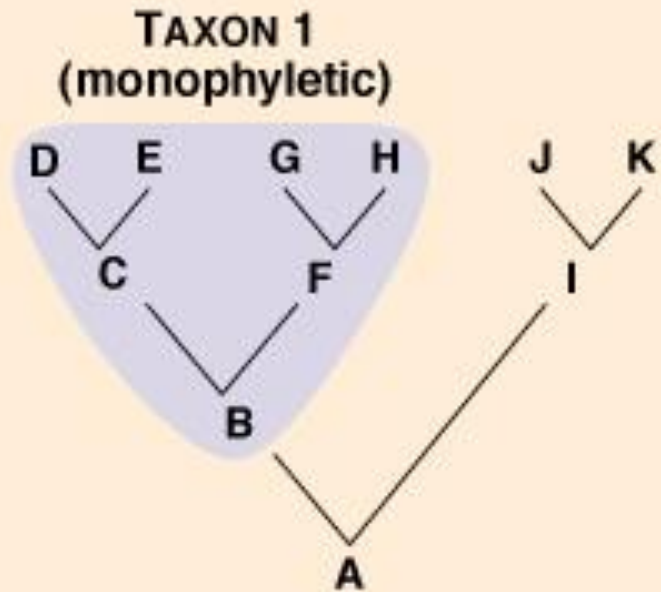
Rooted



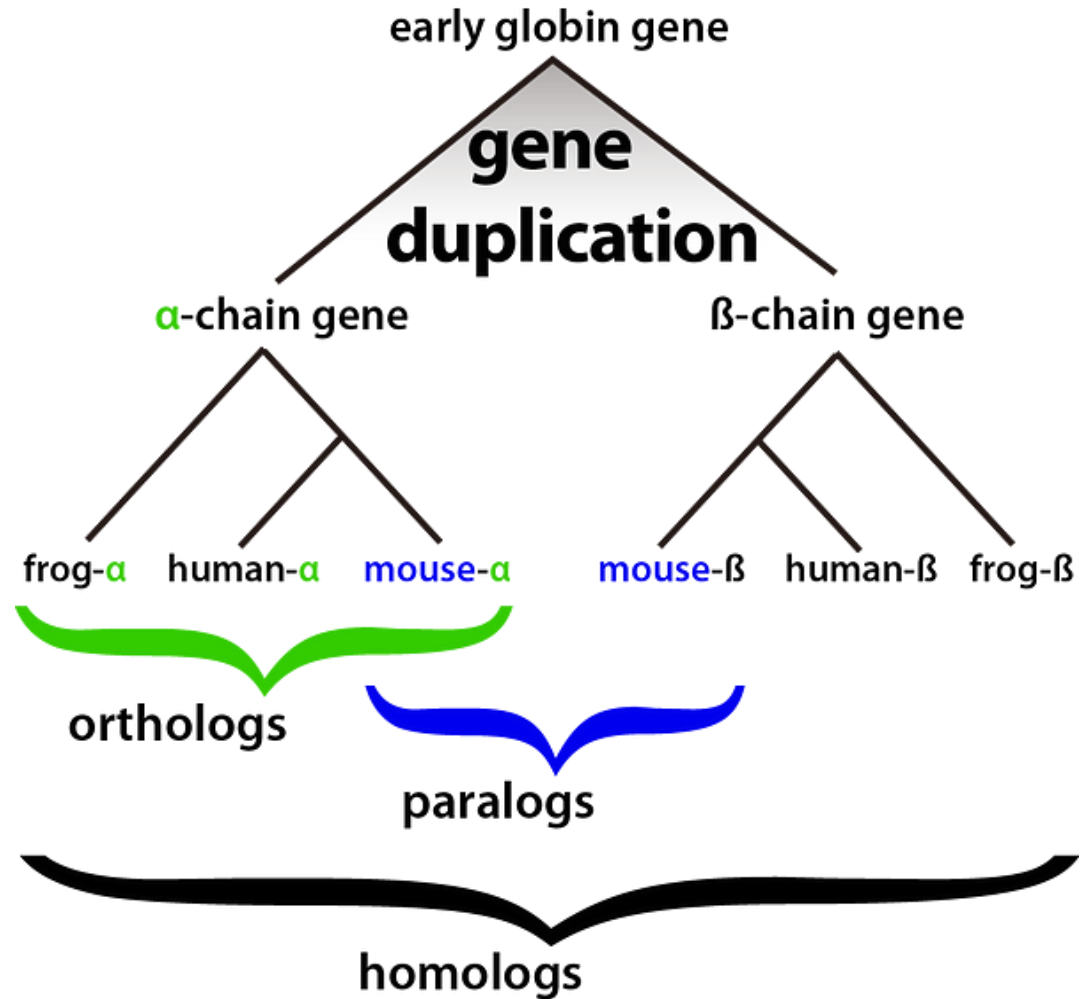
Unrooted



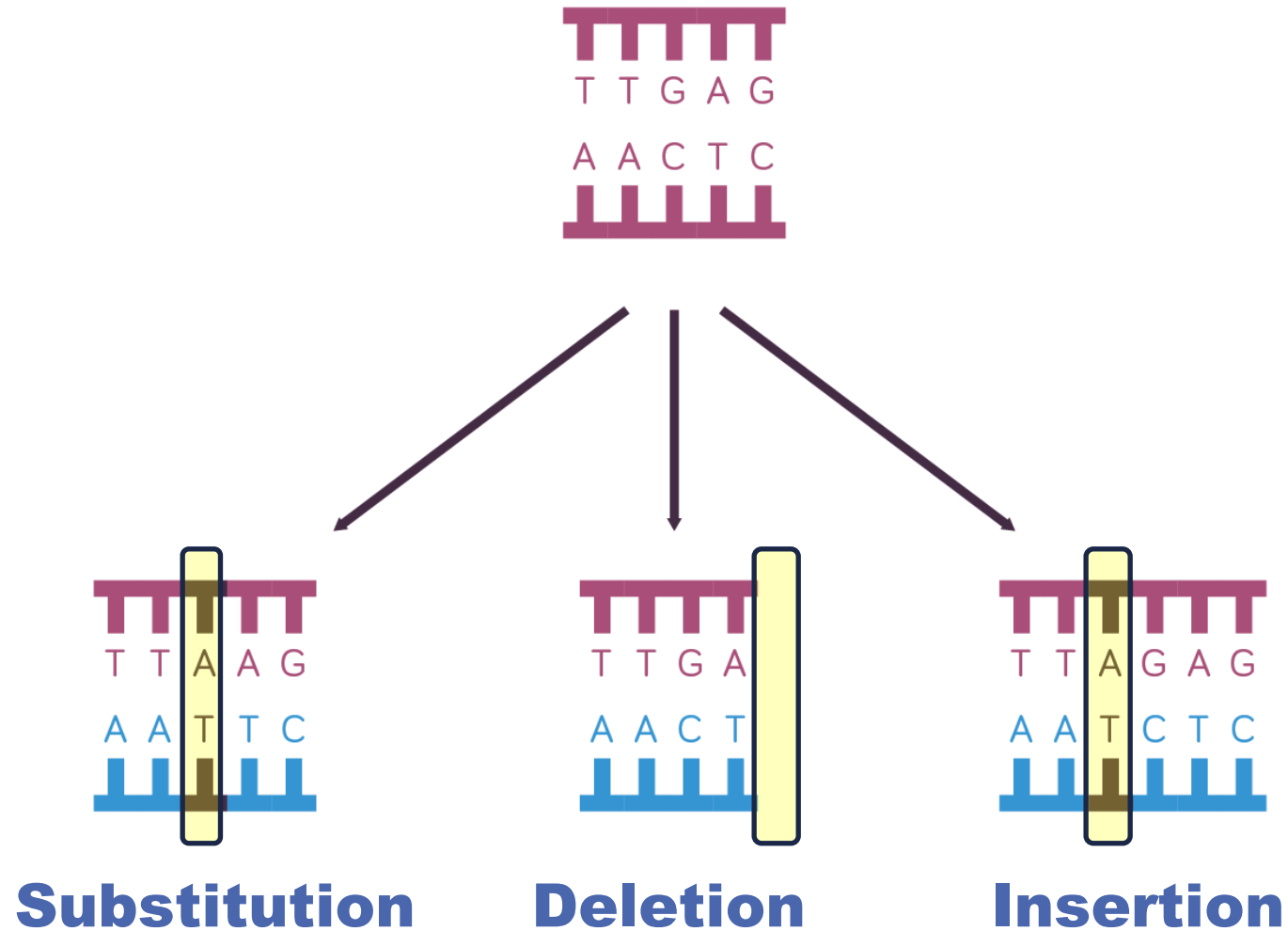
What is a **phyletic clade**?



What are orthologs, paralogs and homologs?



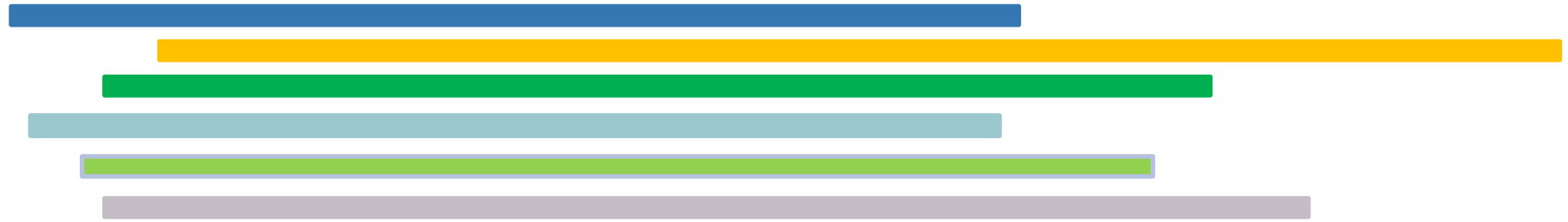
Why are mutations important?



Mutations are the molecular data used to create phylogenetic trees

What is multiple sequence alignment?

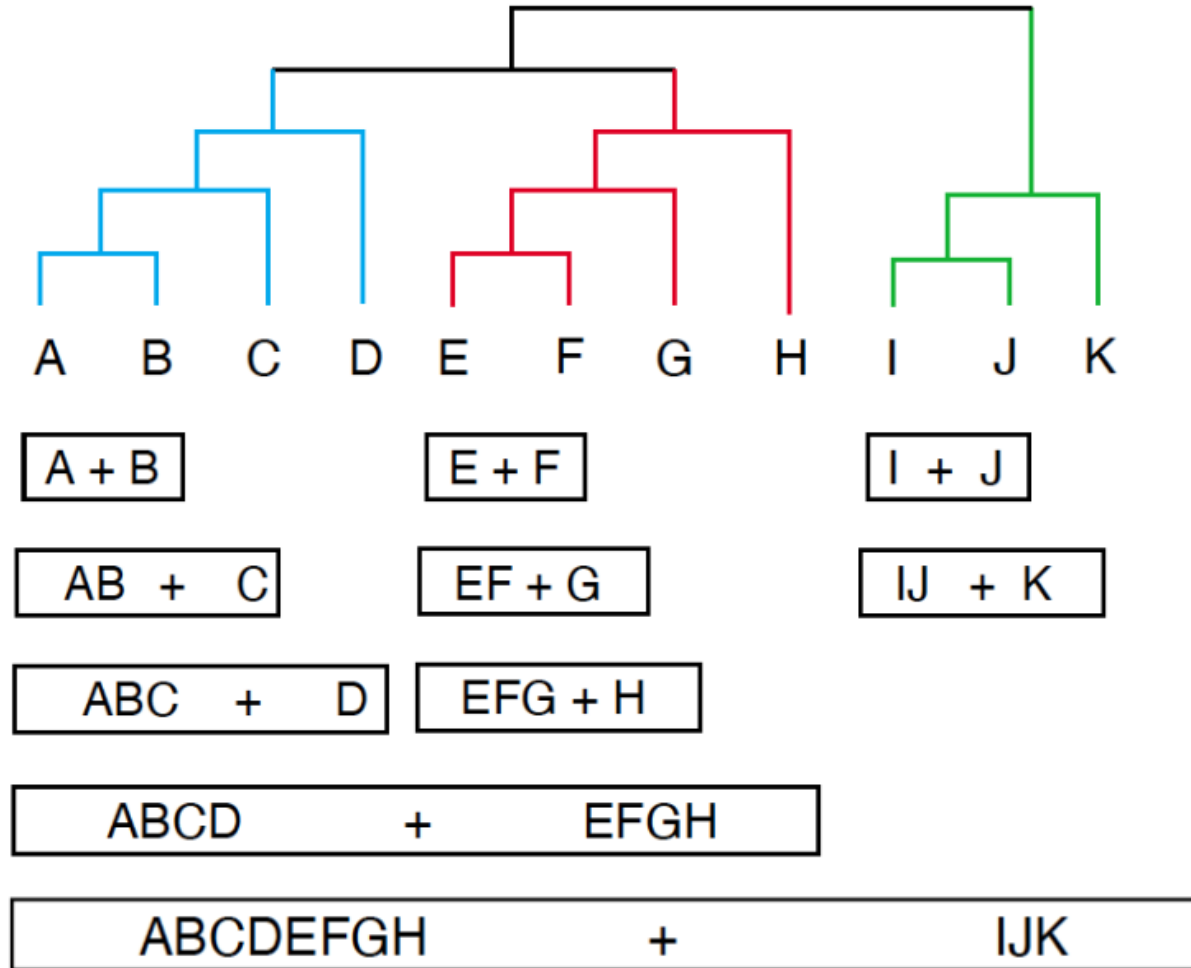
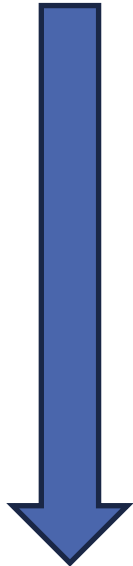
Sequences



				115				120					125						
<i>Sequence A</i>	A	G	T	T	G	A	C	T	T	C	T	C	A	G	G	T	A	T	T
<i>Sequence B</i>	A	G	G	T	A	A	C	T	T	C	A	G	A	T	G	A	A	A	T
<i>Sequence C</i>	A	G	G	T	C	A	C	-	-	G	A	C	A	G	G	C	A	T	T
<i>Sequence D</i>	A	G	G	T	C	A	C	-	-	G	A	C	A	G	G	C	A	-	T
<i>Sequence E</i>	A	G	G	T	C	A	C	T	T	G	A	G	A	-	G	C	A	-	T
<i>Sequence F</i>	A	G	G	T	C	A	C	T	T	G	A	C	A	G	G	C	A	T	T

What is progressive sequence alignment?

Guide Tree



T-COFFEE
SIMPLE MSA

It aligns sequences pairwise, starting with the most similar sequences

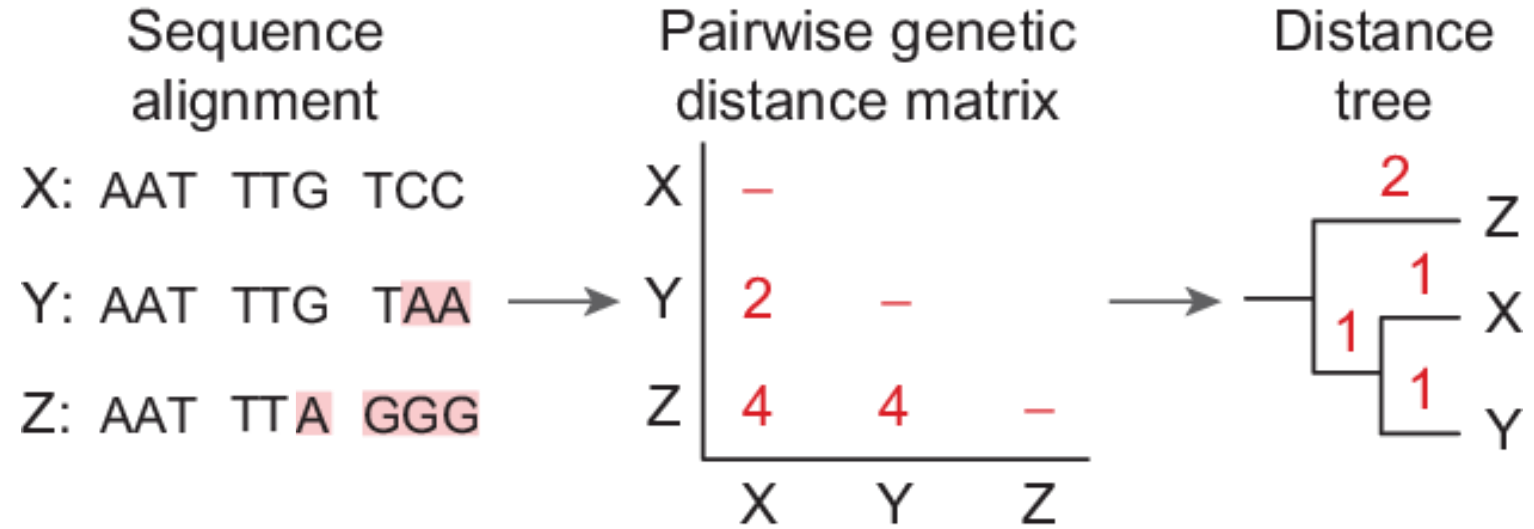
Where is sequence data found?

e!Ensembl

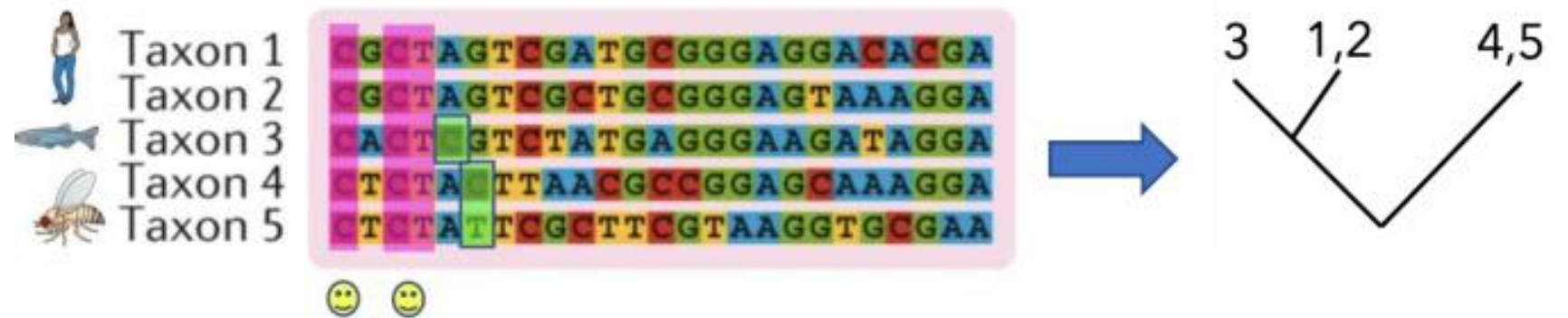


How does sequence alignment create a tree?

Distance Based

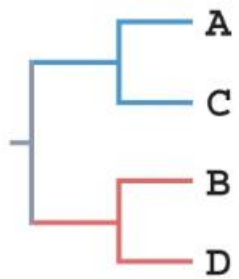
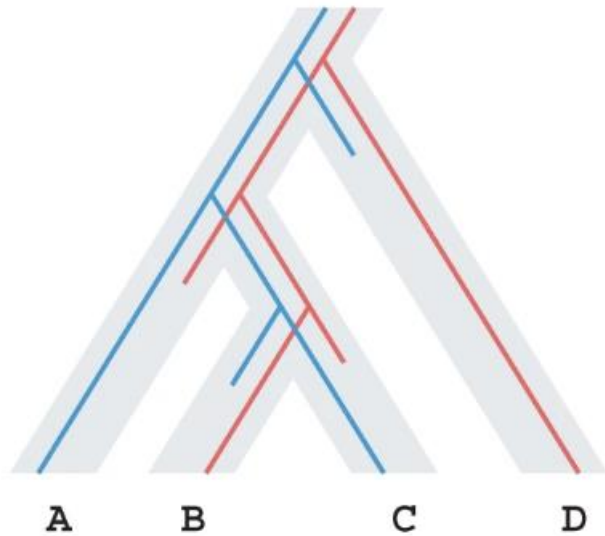


Character Based

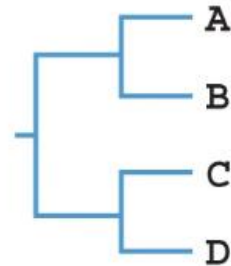
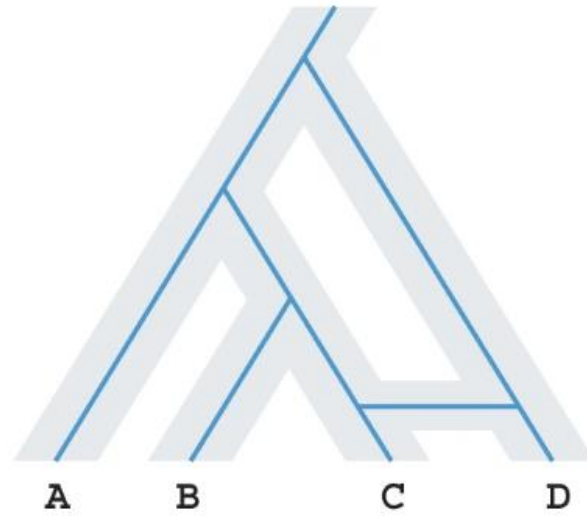


What can cause incorrect tree topologies?

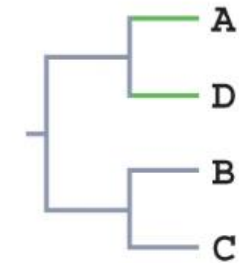
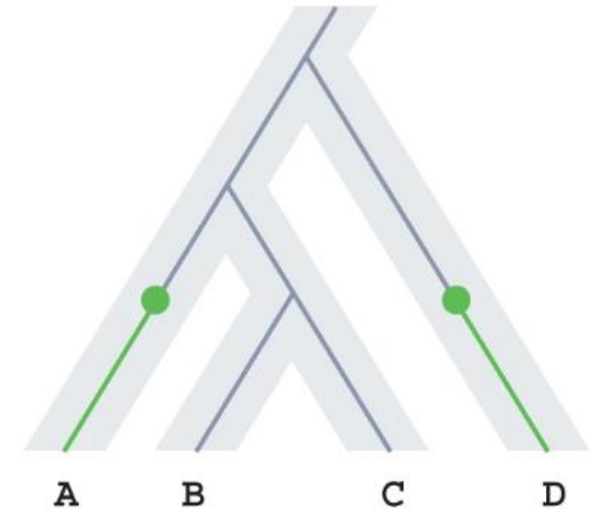
Duplication and loss



Horizontal Gene Transfer



Convergent evolution



What tests are used on phylogenetic trees?

Dataset

	0123456789
seqA	ACCGTTCGGT
seqB	ATGGTTCAGA
seqC	ATCGATCGGA



Replicate 1

	1562314951
seqA	CTCCGCTTTC
seqB	TTCGGTTATT
seqC	TTCCGTAATT

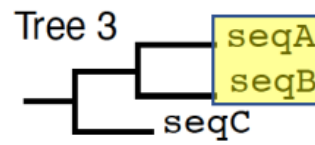
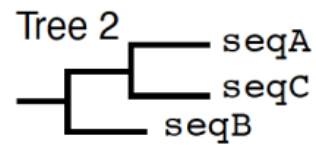
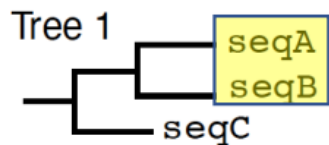
Replicate 2

	5234924418
seqA	TCGTTCTTCG
seqB	TGGTAGTTTG
seqC	TCGAACAATG

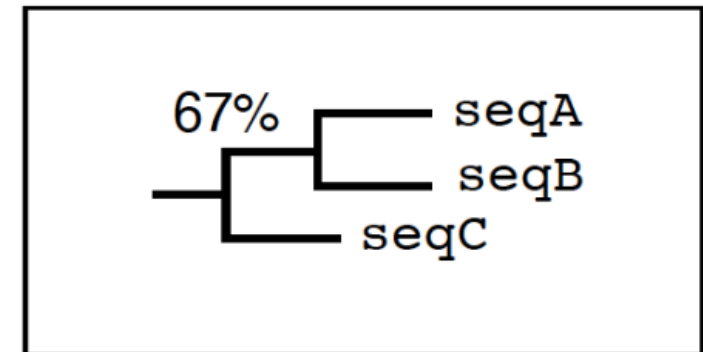
Replicate 3

	5607718907
seqA	TCAGGCGTAG
seqB	TCAAATGAAA
seqC	TCAGGTGAAG

etc



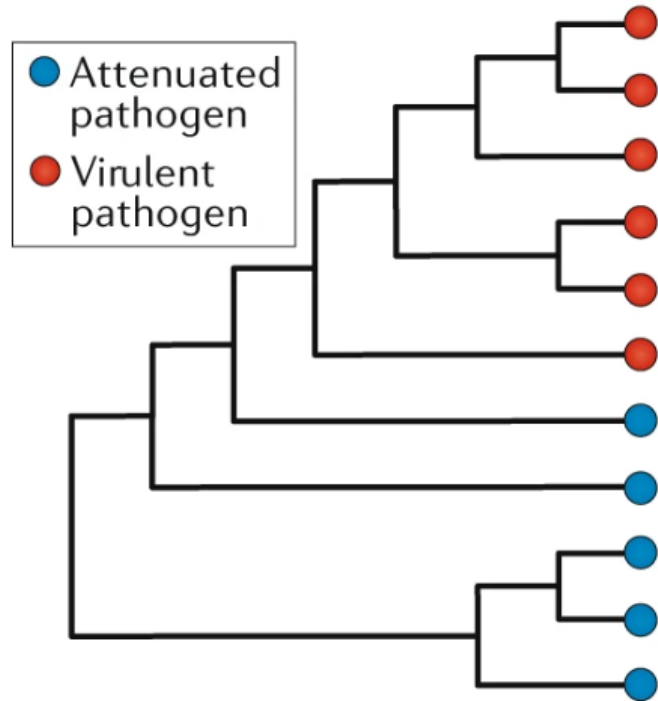
Values >70% are considered reliable



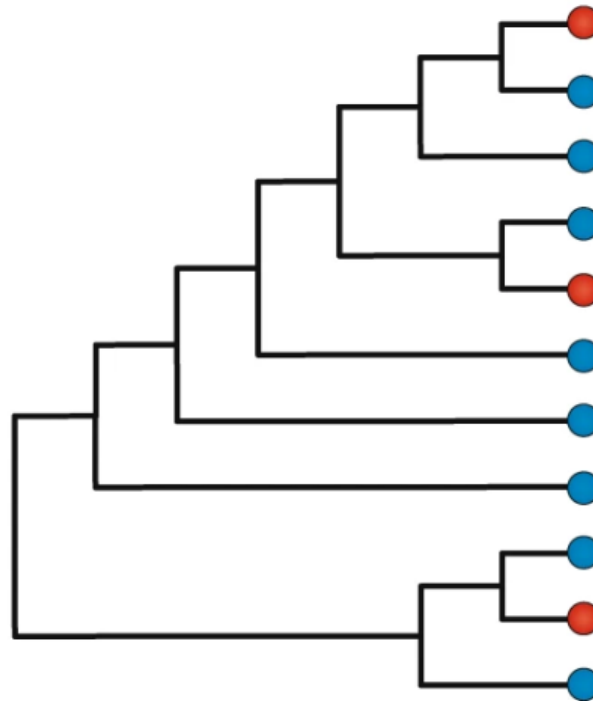
Bootstrapping can provide evidence for a phylogenetic tree

The phylogenetics of evolving virus virulence

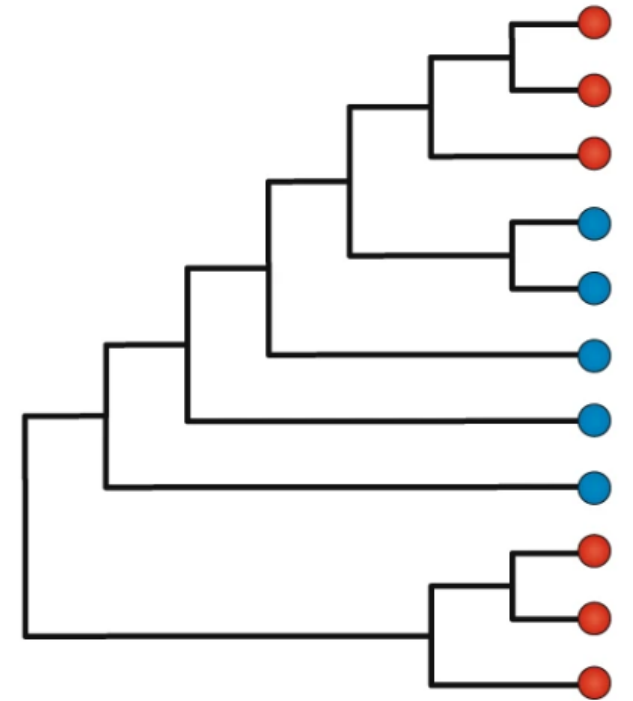
a Virulence determinant on deep node



b Virulence determinant on shallow node

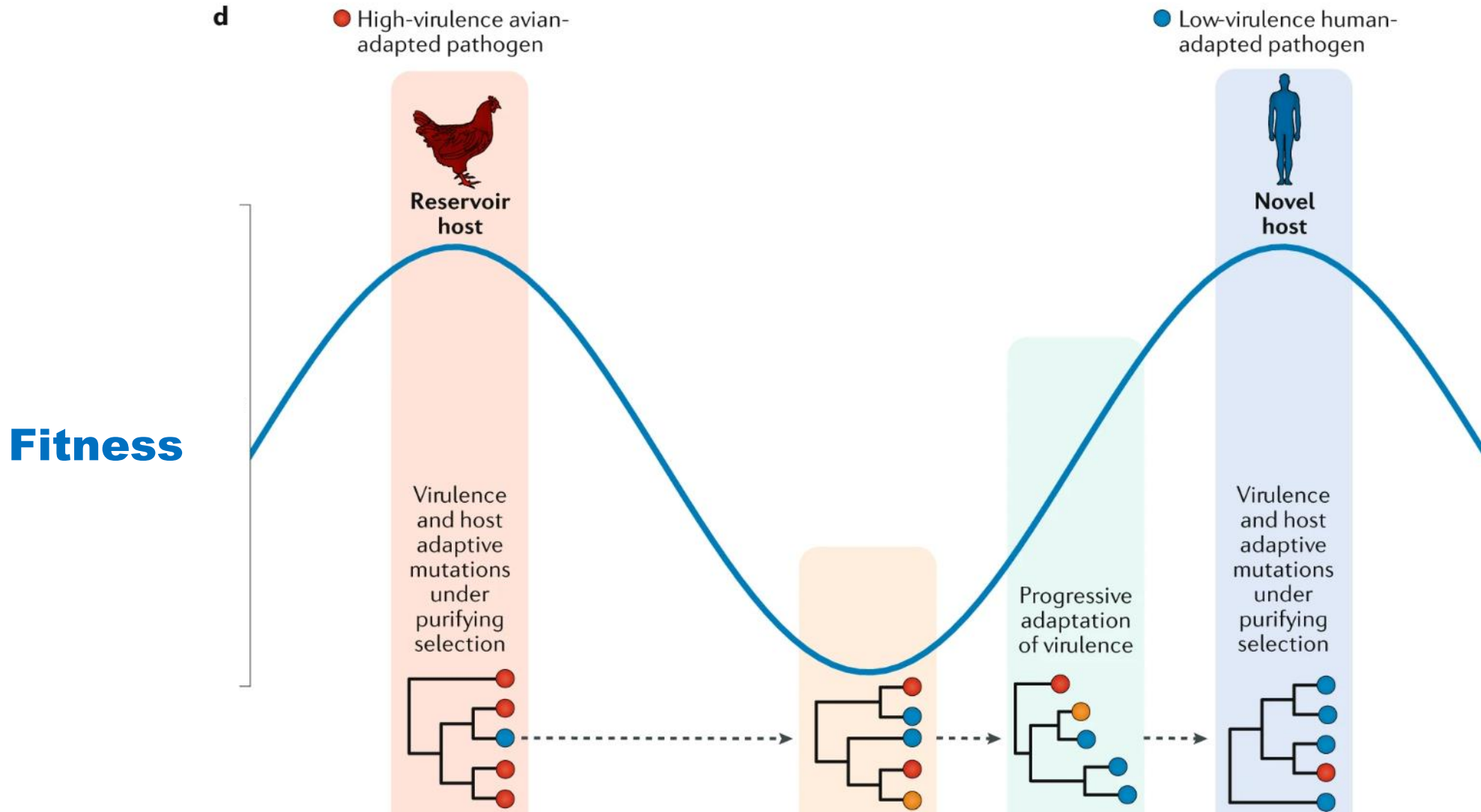


c Parallel or convergent evolution



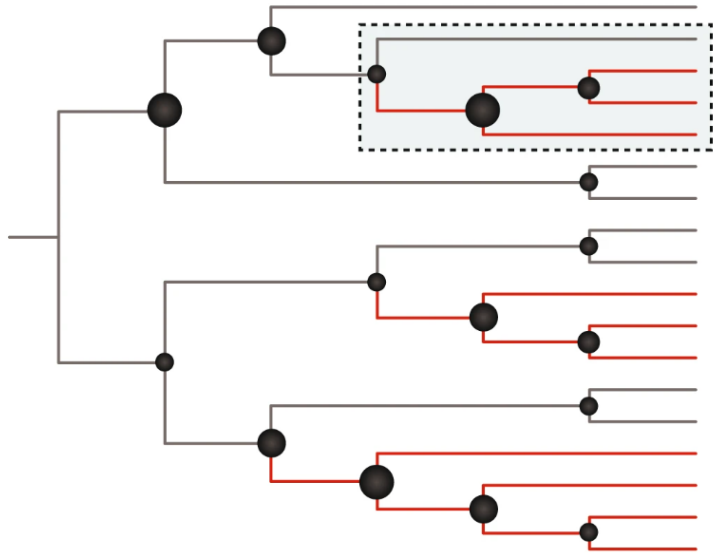
Phylogenetic analysis provides evolutionary information on virulence and transmissibility

How does virulence evolve in hosts?

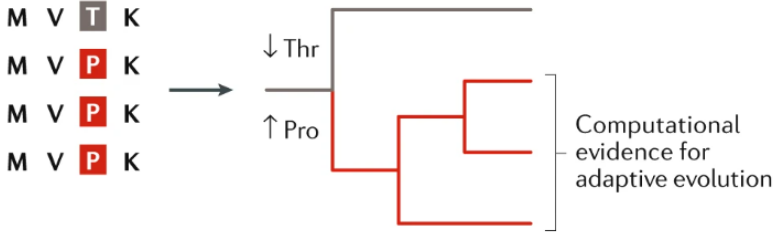


Why should we track virus phylogenies?

A Phylogenetic inference

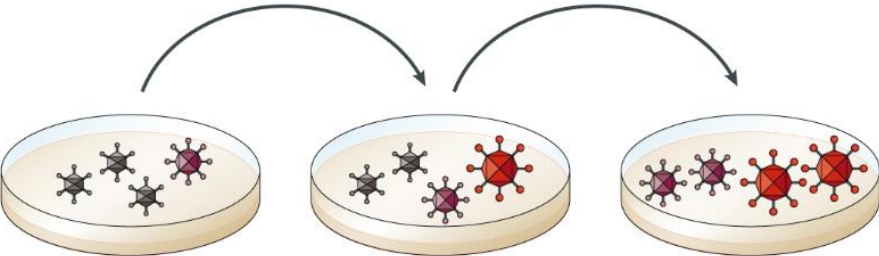


B Evolutionary analyses

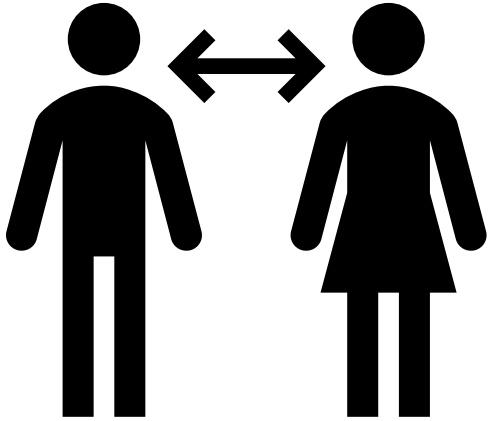
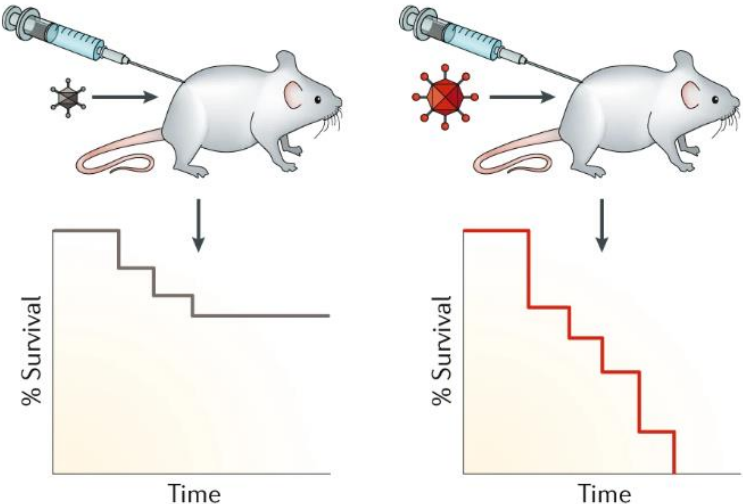


C Experimental assays

Ca In vitro



Cb In vivo

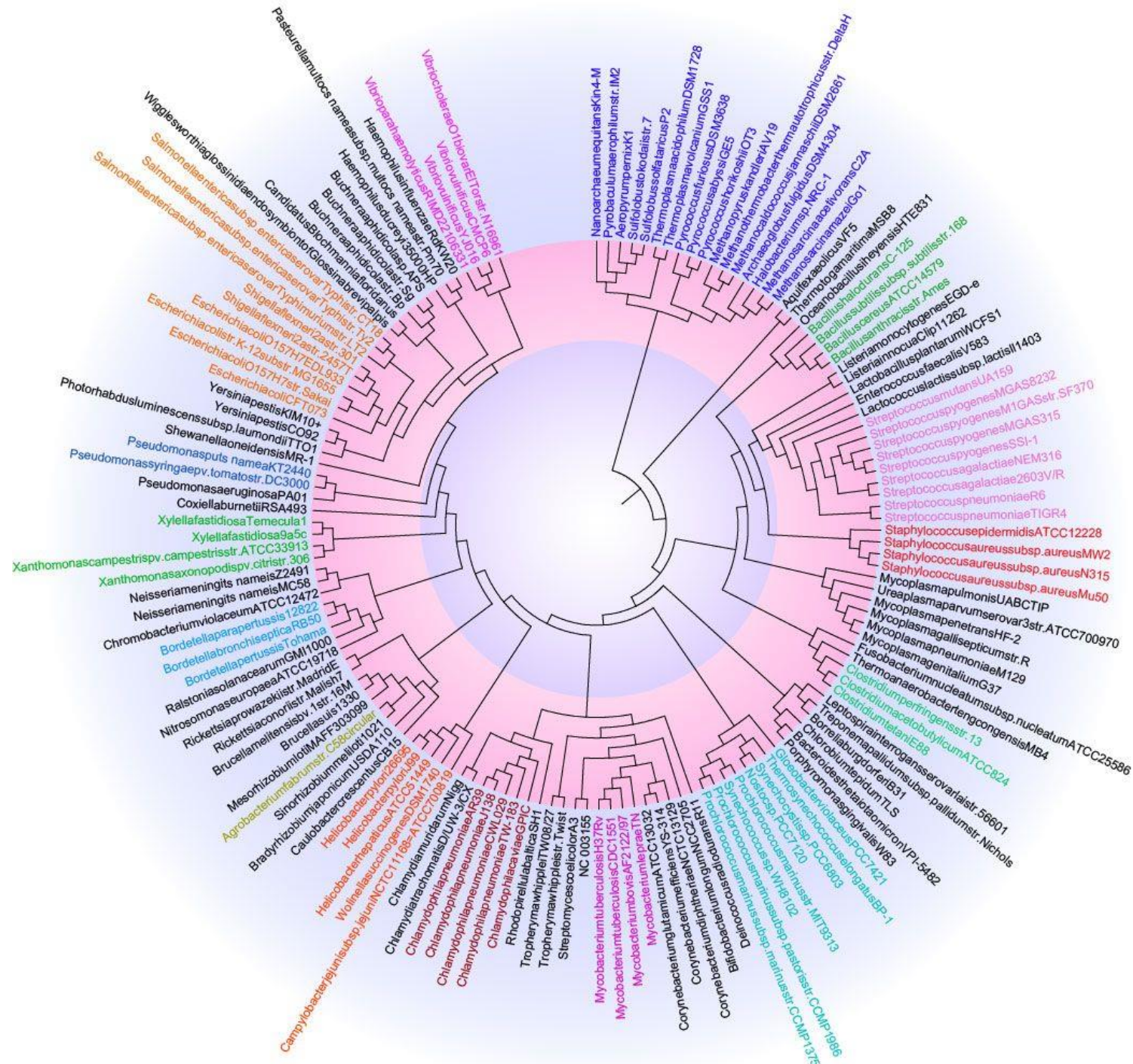


Summary

Phylogenetics helps us study the evolutionary information of biological entities

Phylogenetic trees are made using molecular data to determine relatedness

Virulence is a complex trait and studying its evolution reveals information about the virus



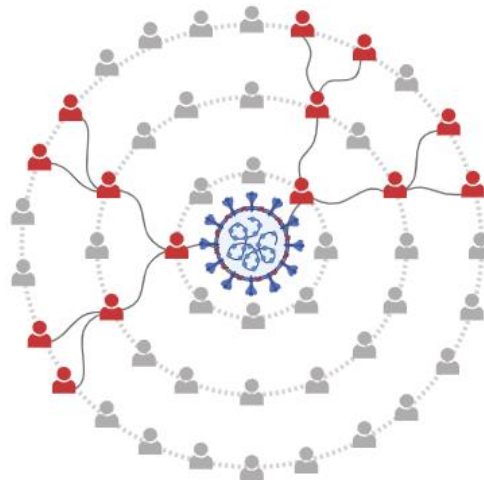
The Bedford Lab



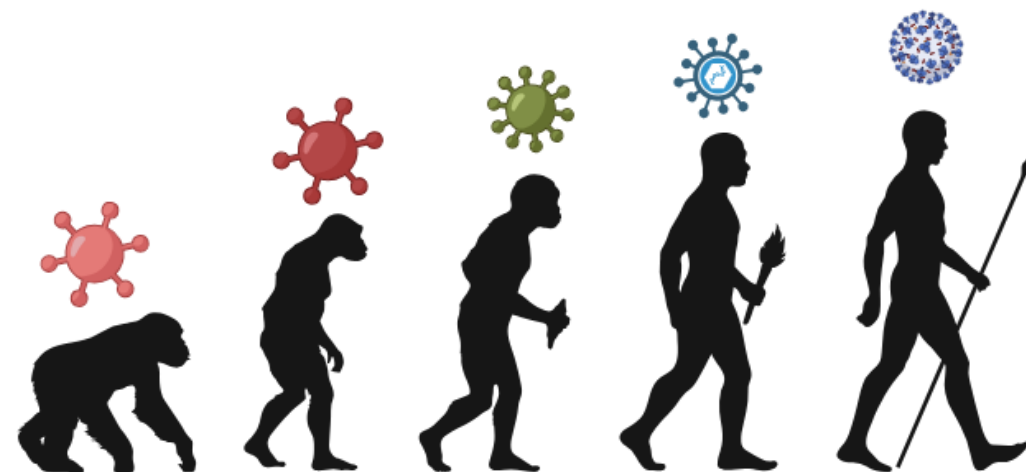
Fred Hutch
Cancer Center



Trevor Bedford
Principle Investigator



Katie Kistler
Postdoc



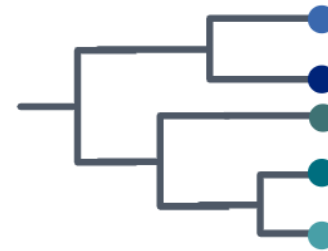
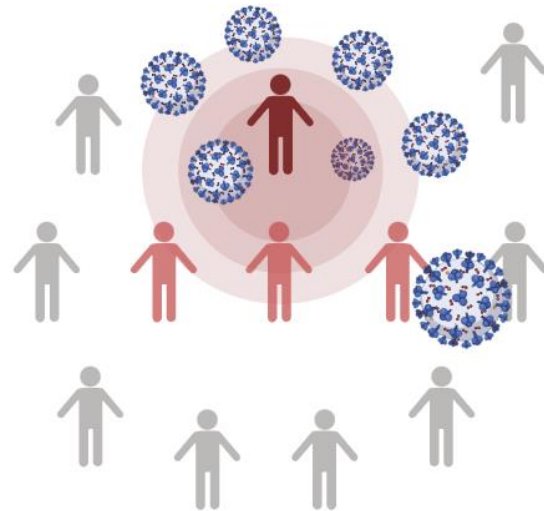
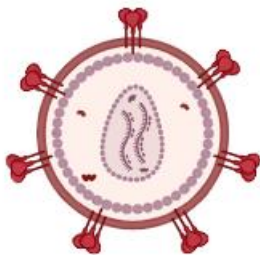
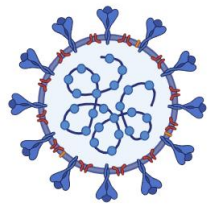
Meet the authors



**Michael
Worobey**
The University of
Arizona



Jonathan Pekar
PhD Student,
University of
California, San
Diego



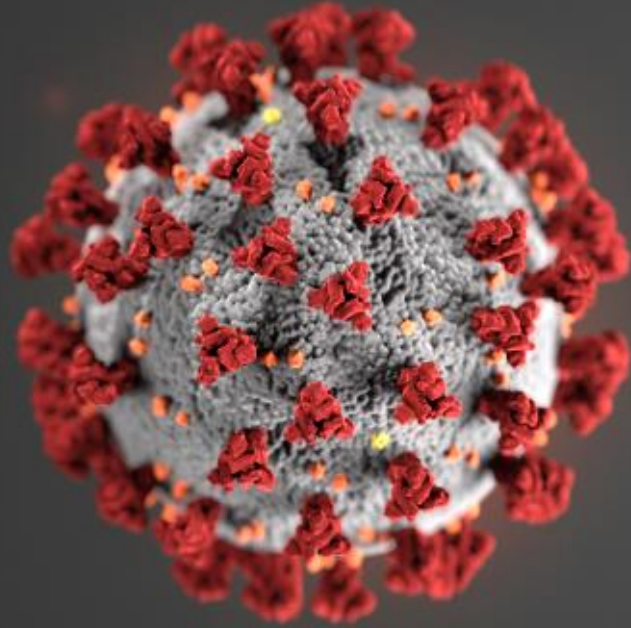


Questions?



The emergence of SARS- CoV-2 in Europe and North America

What is SARS-CoV-2?



Severe **A**cute **R**espiratory **S**yndrome **C**oronavirus **2**

What is COVID-19?

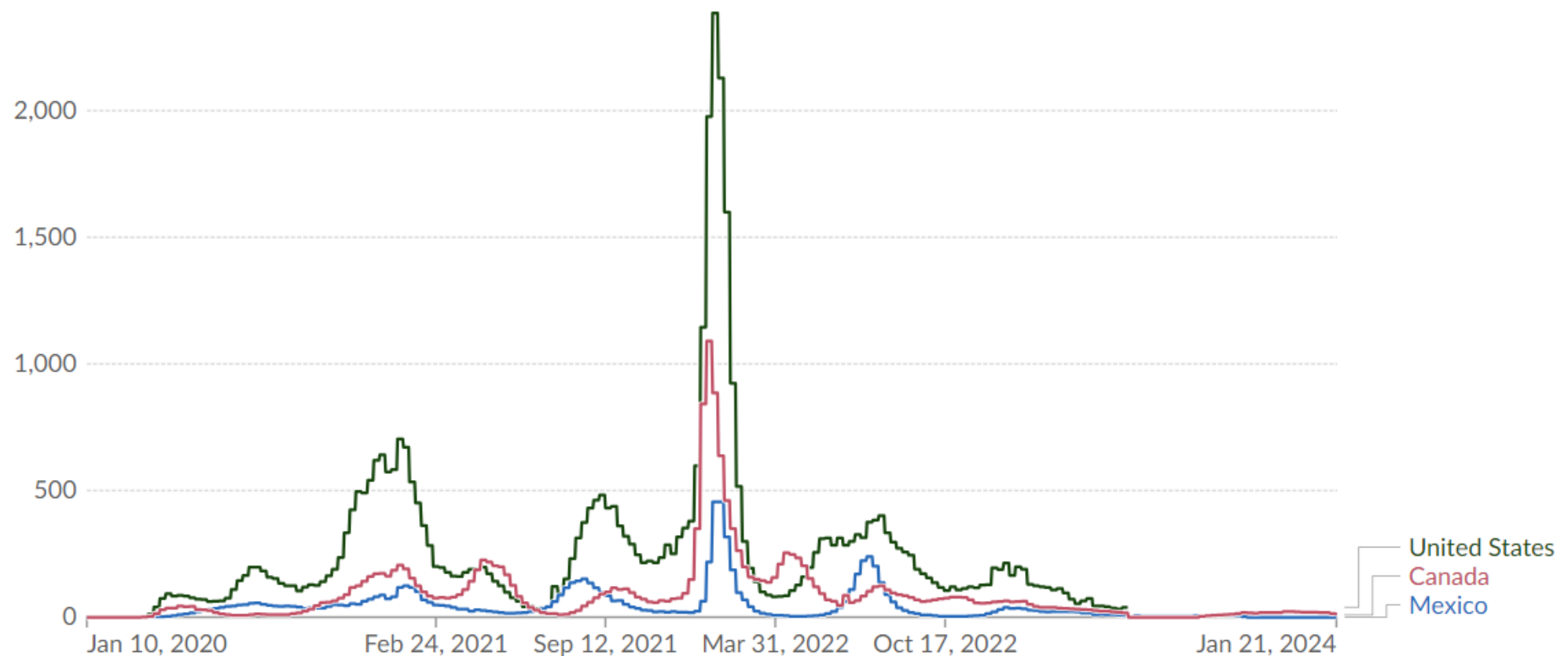
Daily new confirmed COVID-19 cases per million people

7-day rolling average. Due to limited testing, the number of confirmed cases is lower than the true number of infections.

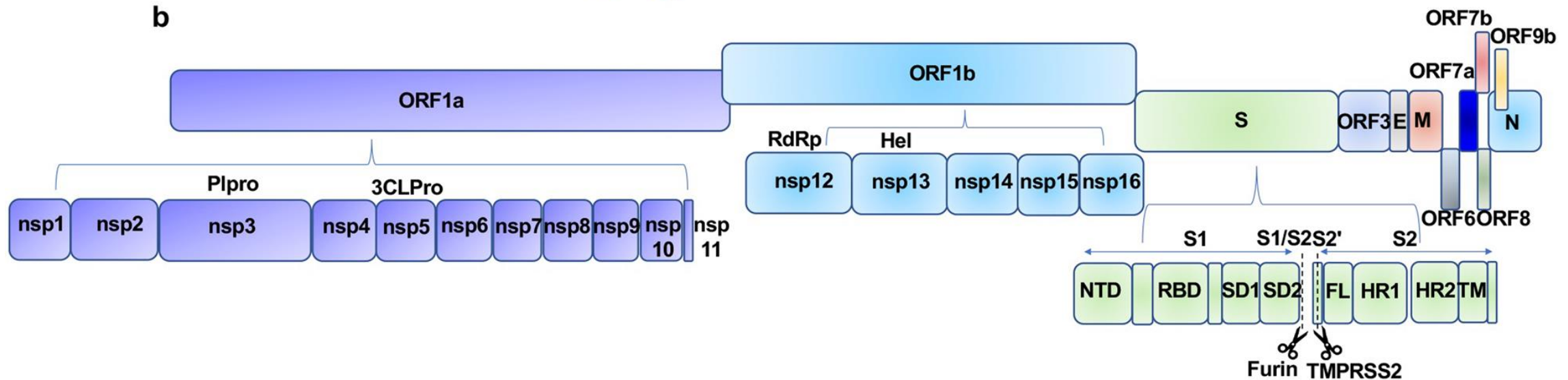
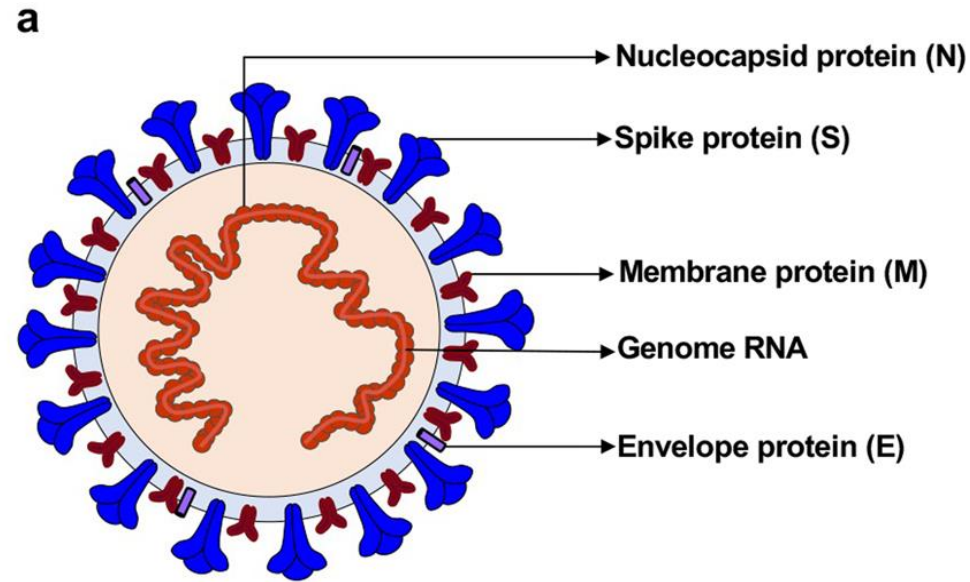
Our World
in Data

Table | Map | Chart

Settings



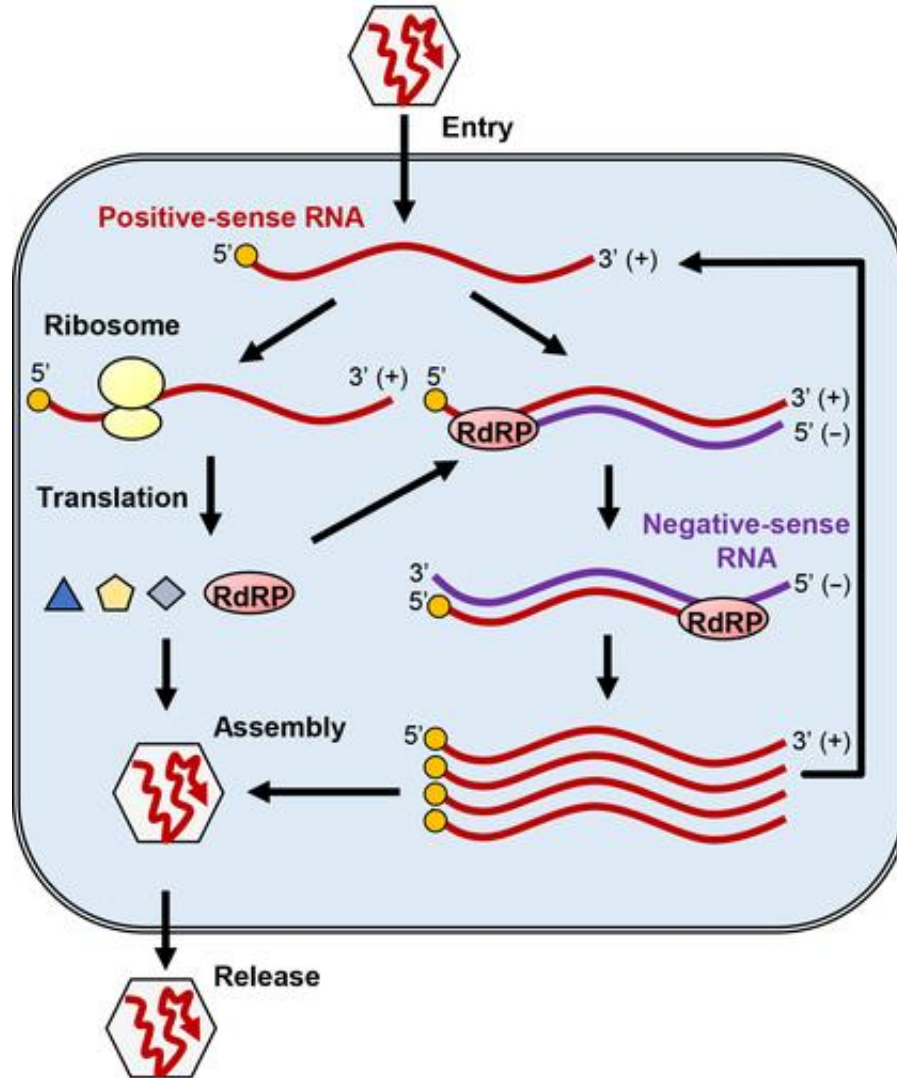
What are important characteristics of SARS-CoV-2?



What are positive-sense RNA viruses?

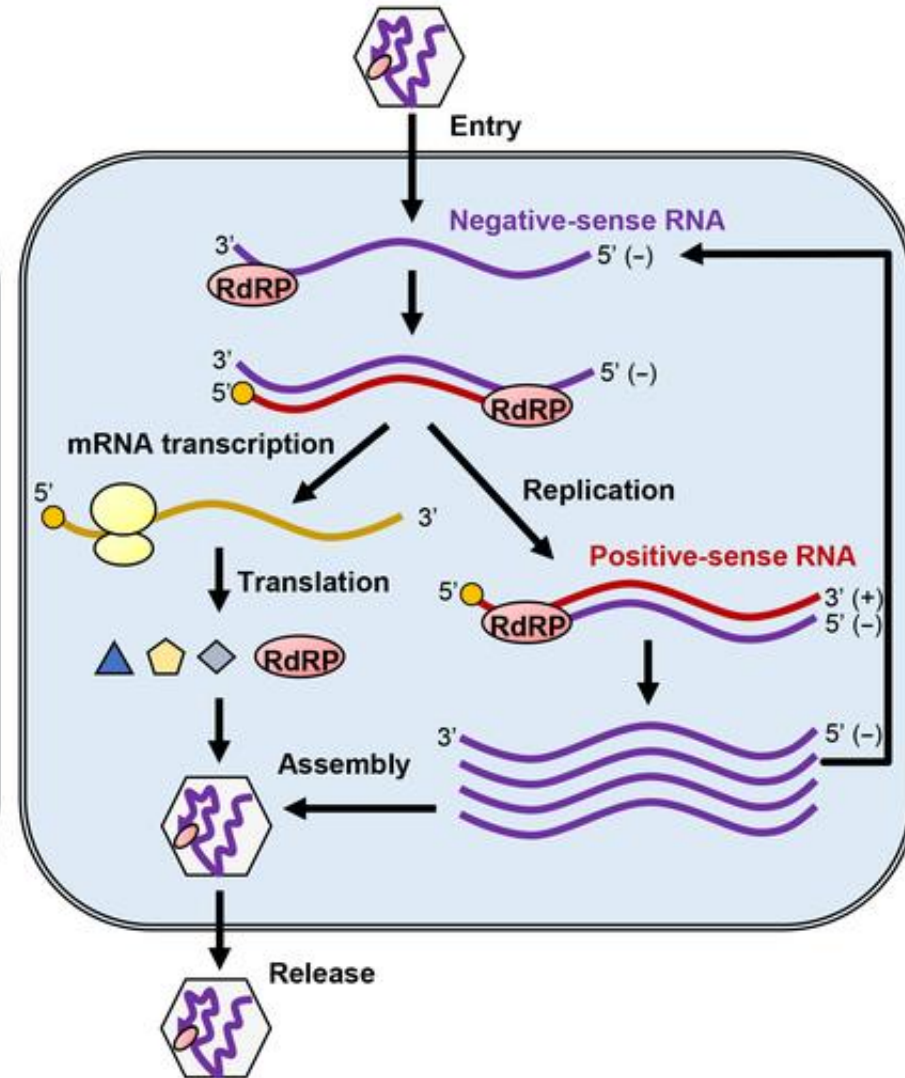
(A)

positive-sense ssRNA virus

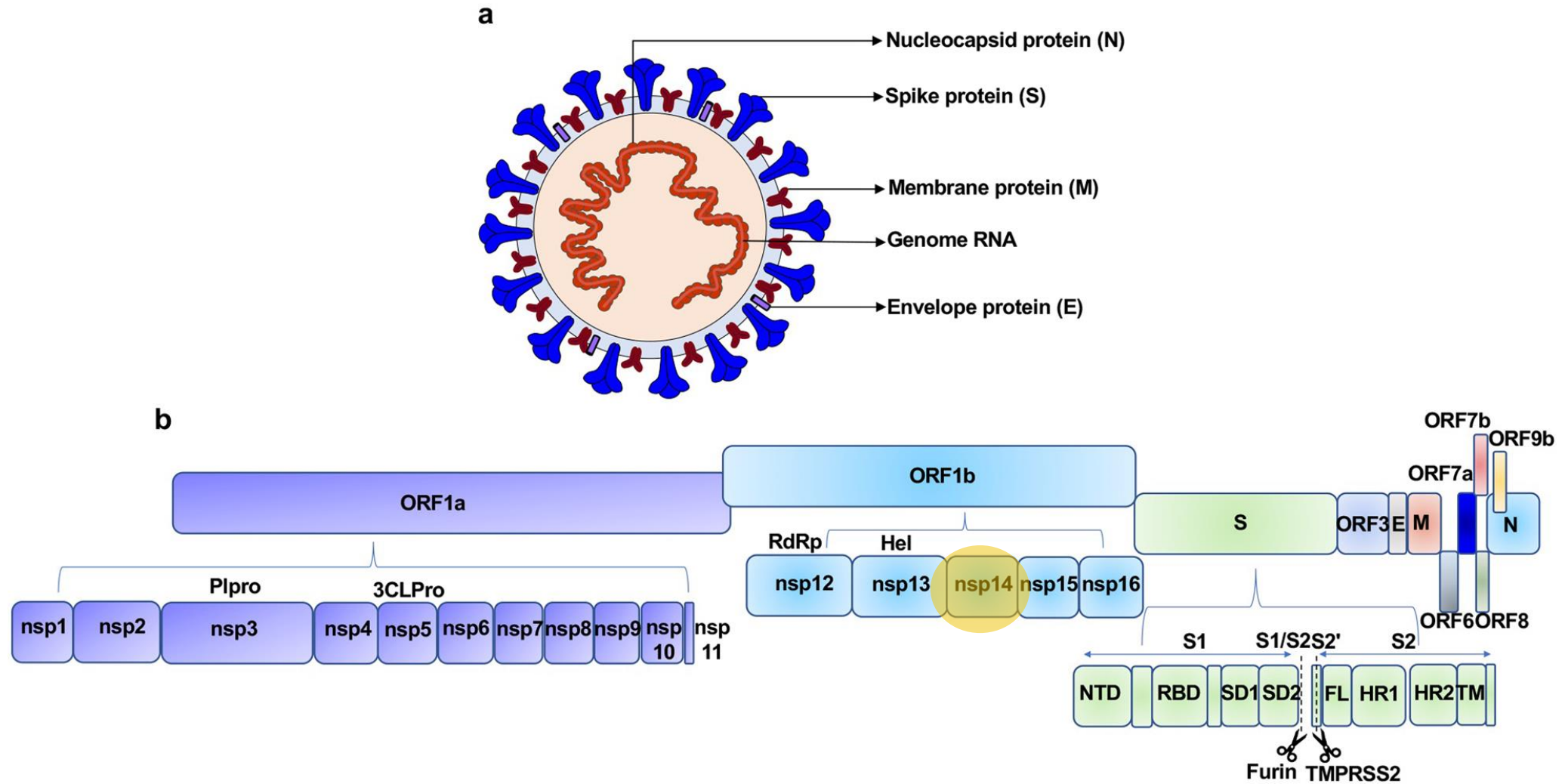


(B)

negative-sense ssRNA virus



How does proofreading activity of SARS-CoV-2 cause study limitations?

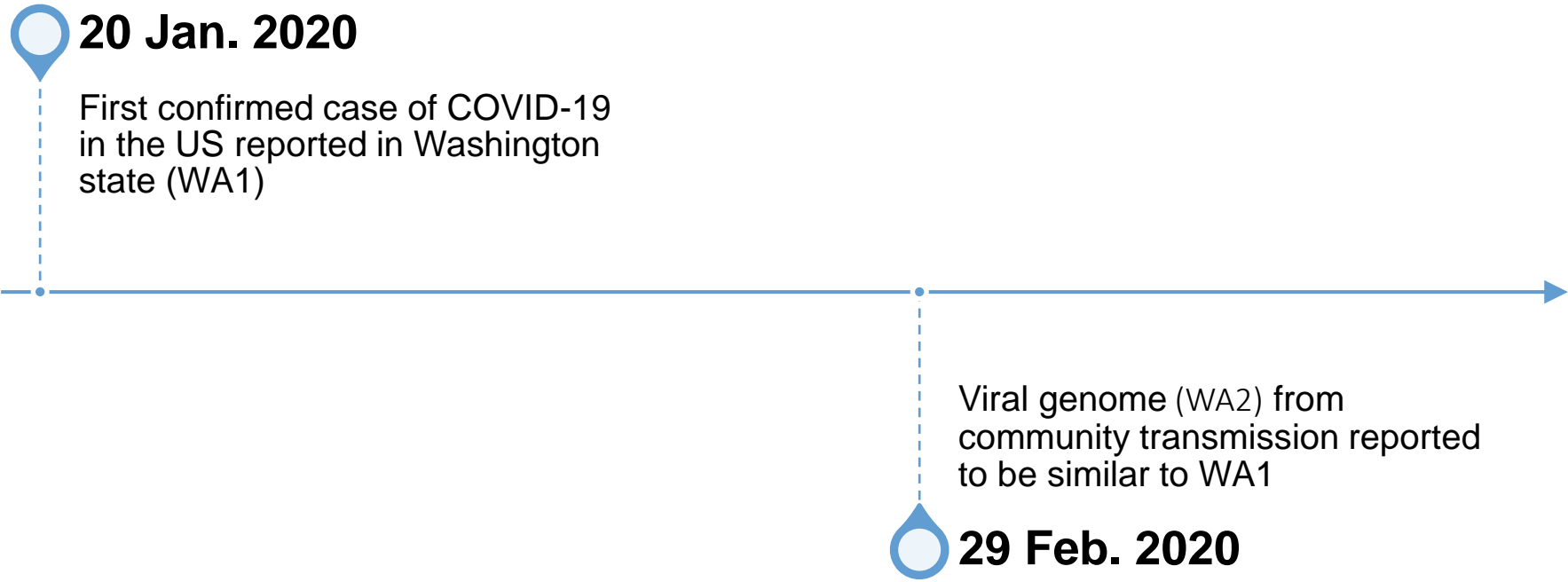


Proofreading by nsp14 leads to a relatively slow mutation rate

What was the main question of the study?

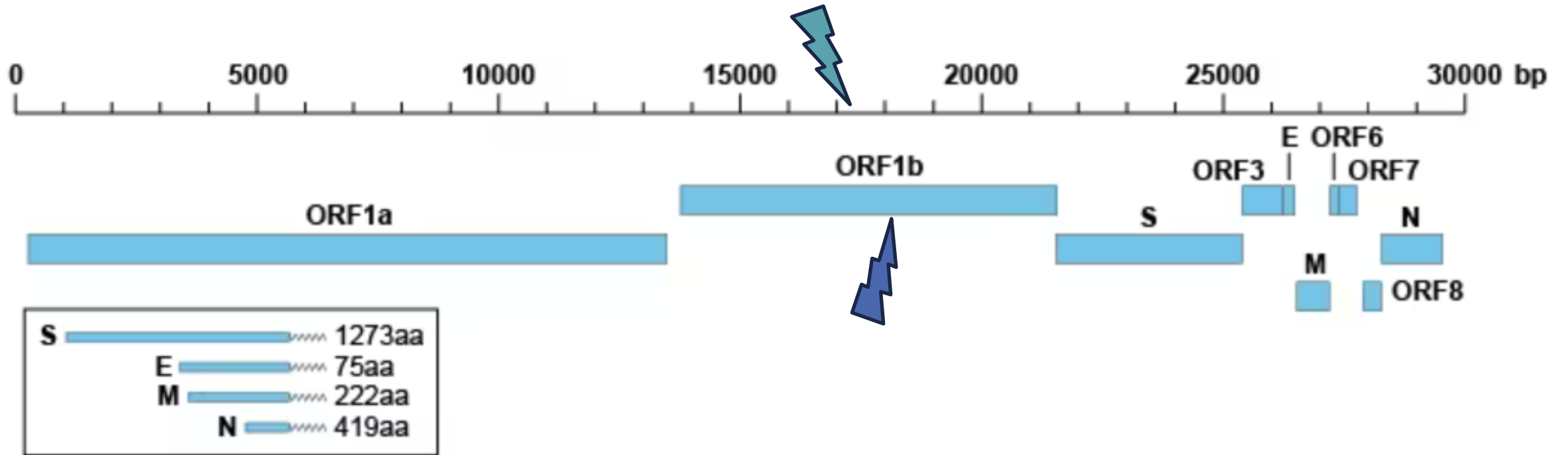


What does the timeline suggest about the introduction of SARS-CoV-2?



Hypothesis #1: WA1 established cryptic transmission in WA in mid-January

What is the Washington (WA) outbreak clade?

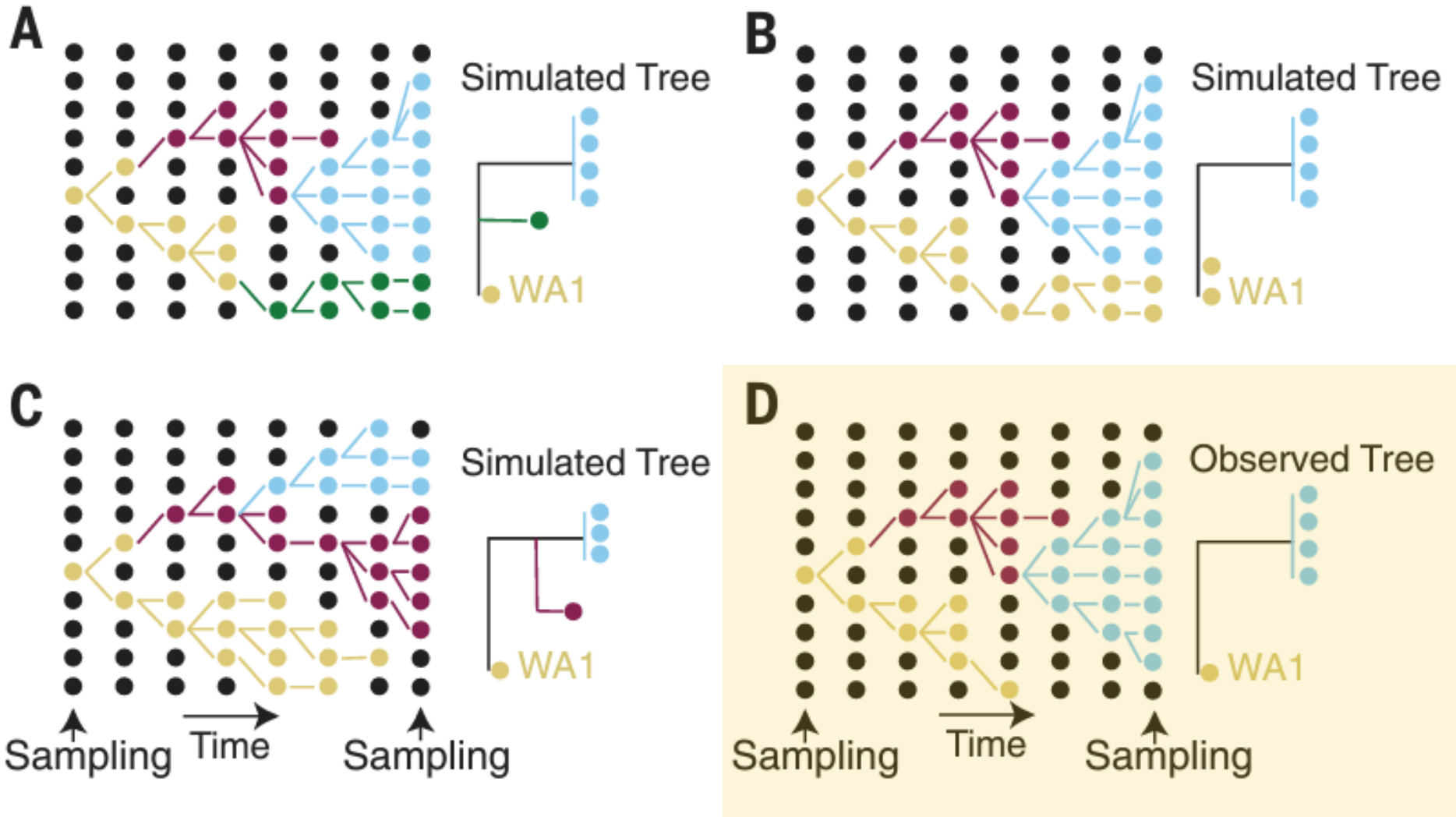


Mutation 1: C17747T

Mutation 2: A17858G

What are possible pathways given that WA1 is the parent strain of the WA outbreak clade?

● C17747/A17858 ● C17747/A17858G ● C17747T/A17858G ● Hypothetical mutation



What is FAVITES?

Table S1. Parameters for epidemic and sequence simulations.

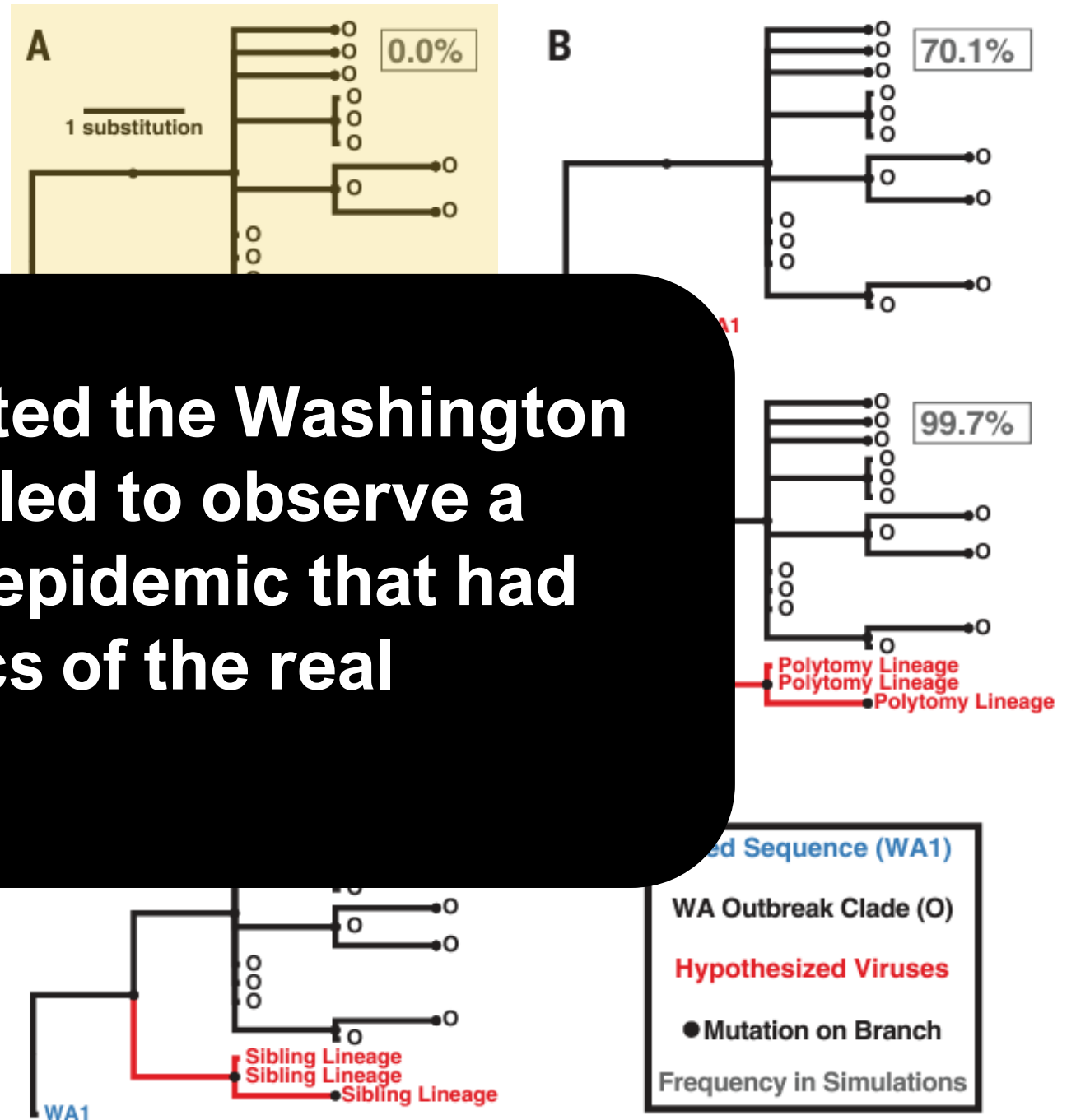
Parameter	Epidemic Simulation					
	WA (main text)	WA - slow spreading	WA - fast evolving	Germany/Italy (main text)	Germany/Italy - slow spreading	Germany/Italy - fast evolving
Days simulated	61	61	61	36	36	36
Number nodes in contact network	20000	5000	20000	5000	2500	5000
Average number of edges per node (degree)	16	16	16	16	16	16
Infectiousness	2.7	2.0-2.2	2.7	2.60-2.85	2.0-2.2	2.60-2.85
Scaled substitution rate (variant site subs/site/year)	0.0456	0.0456	0.0911	0.2647	0.2647	0.5293

¹Additional parameters can be found in the example JSON files at https://github.com/Worobeylab/SC2_outbreak.

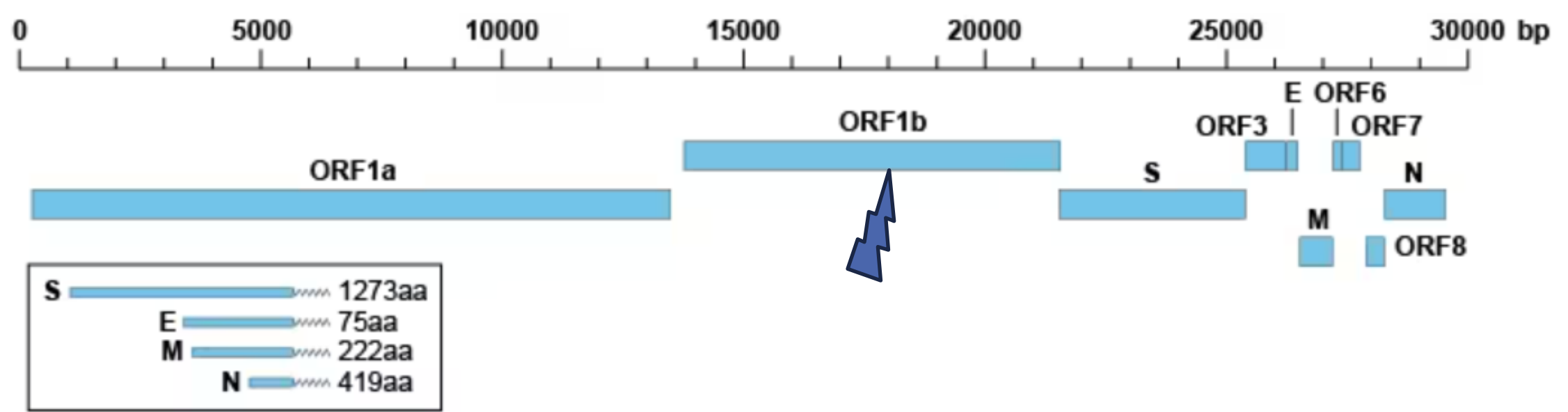
FrAmework for **VI**ral **T**ransmission and **E**volution **S**imulation

How likely are each of the possible scenarios?

"when we simulated the Washington outbreak....we failed to observe a single simulated epidemic that had the characteristics of the real phylogeny."



What is notable about SARS-CoV-2 genomes sampled from British Columbia, Canada?



Mutation 1: C17747T
Mutation 2: A17858G

16/27 genomes

What is another possible epicenter for the WA clade outbreak?



Hypothesis #2: British Columbia (BC) was the epicenter of the outbreak in North America

How does the size of the outbreak compare between WA and BC?

Washington State

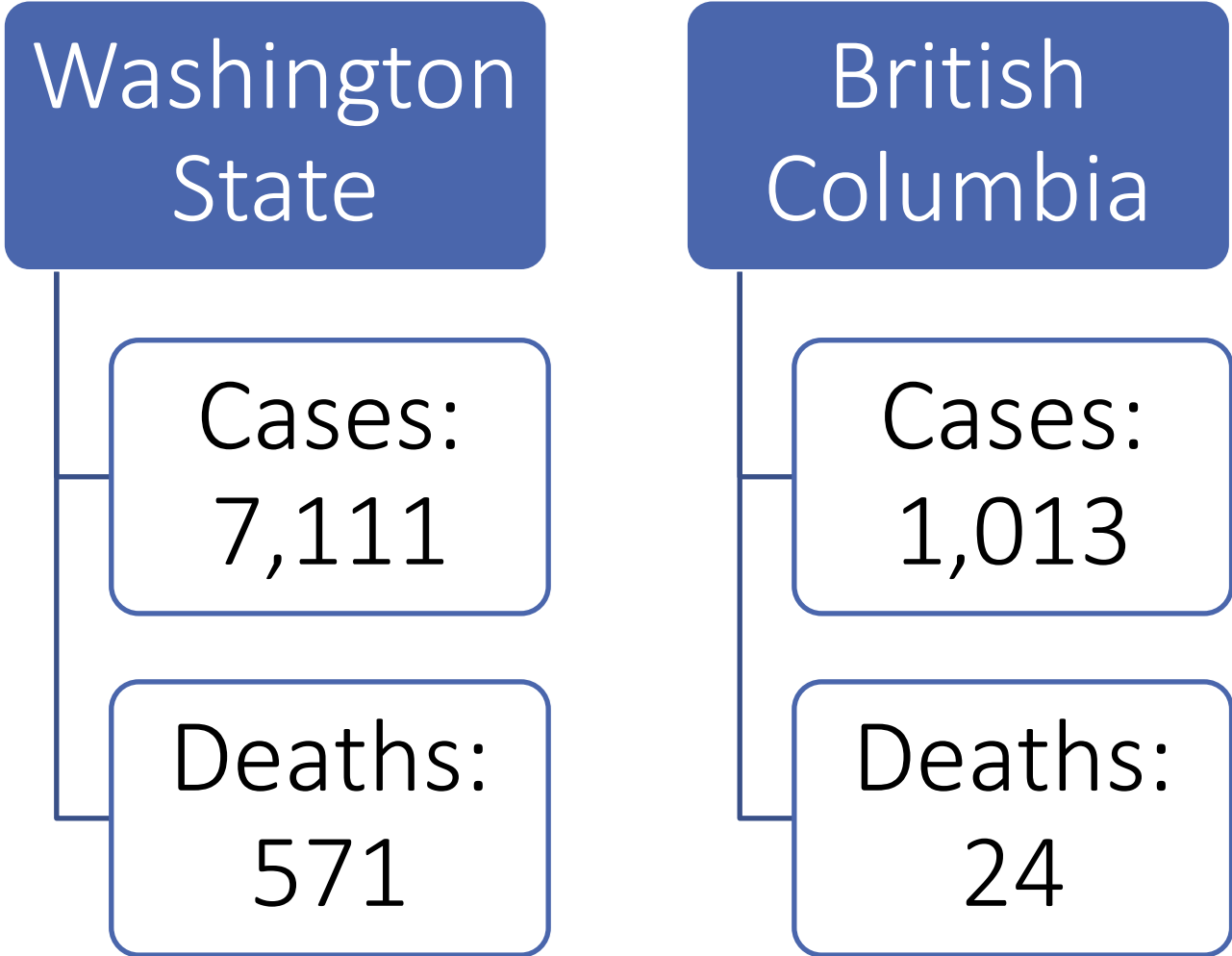
Cases:
7,111

Deaths:
571

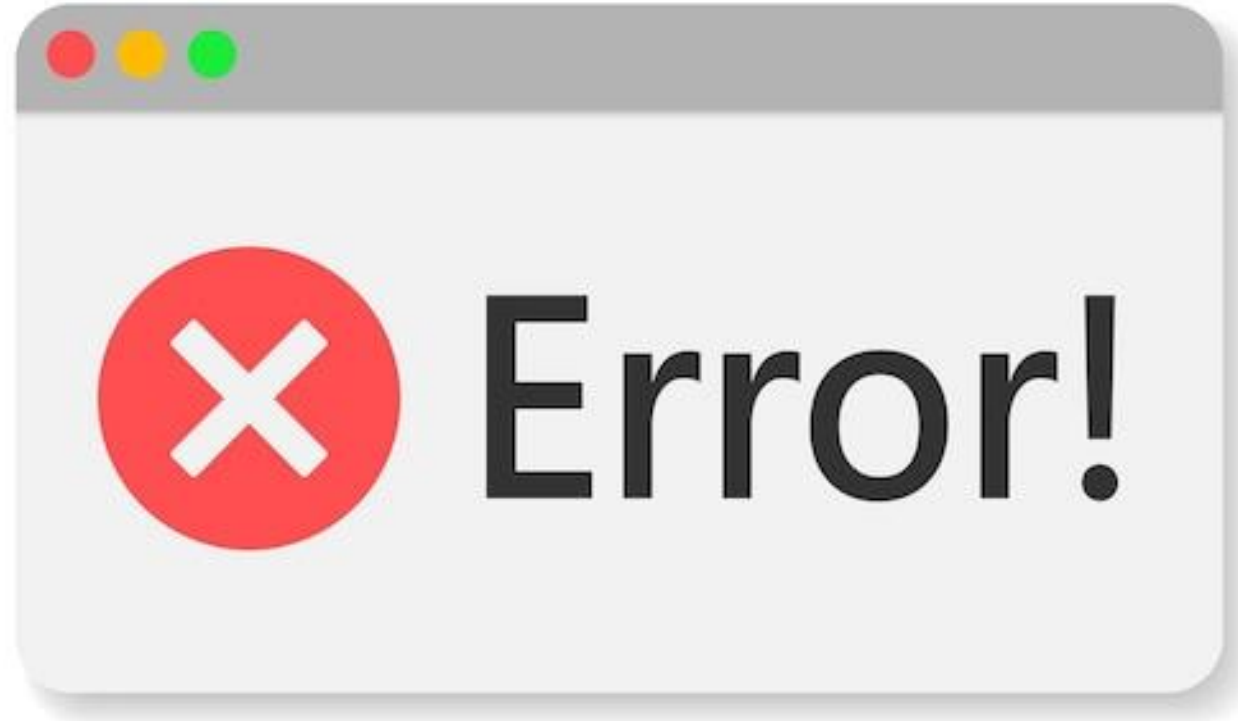
British Columbia

Cases:
1,013

Deaths:
24

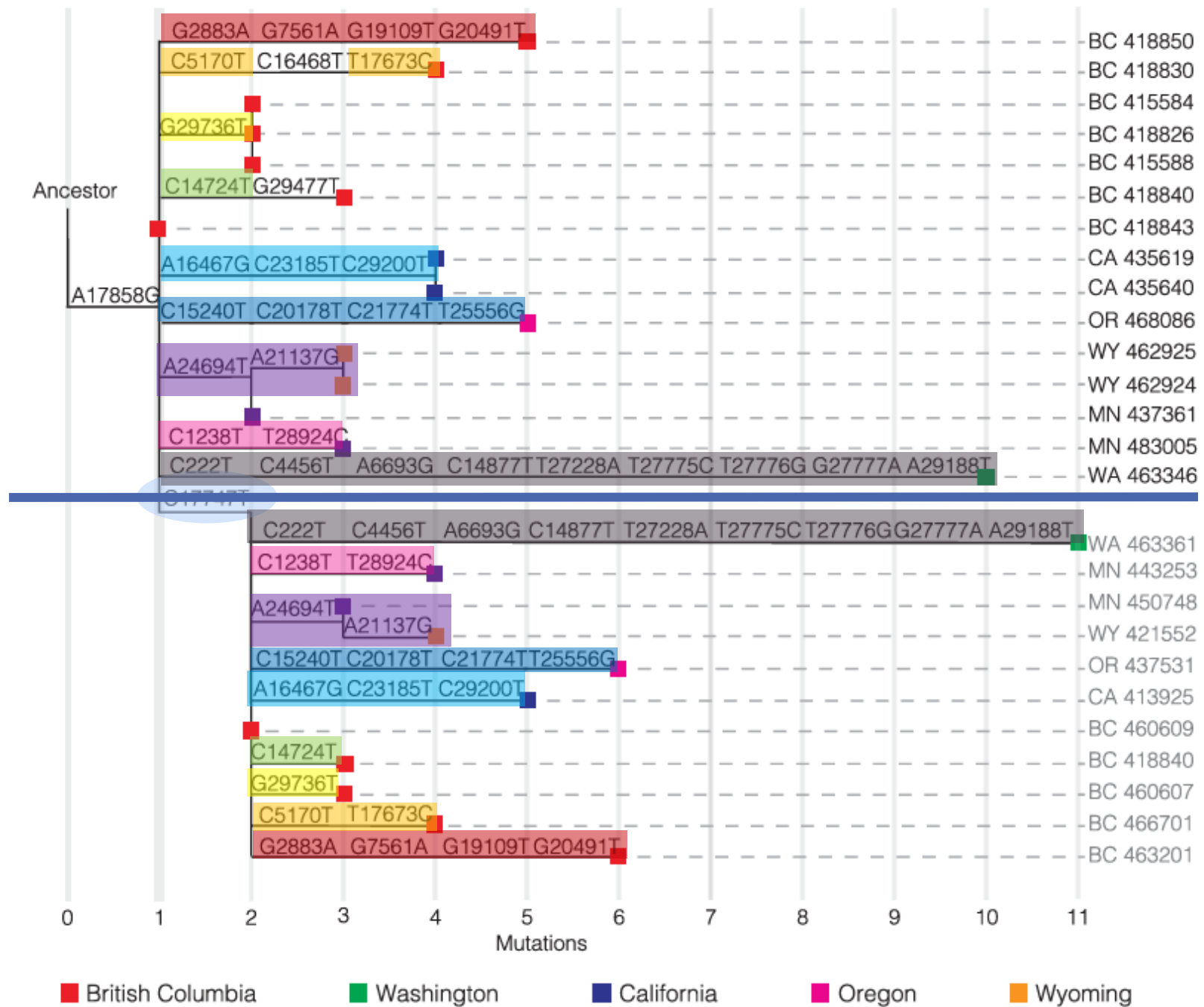


Why do the BC genomes have the ancestral C17747 allele?

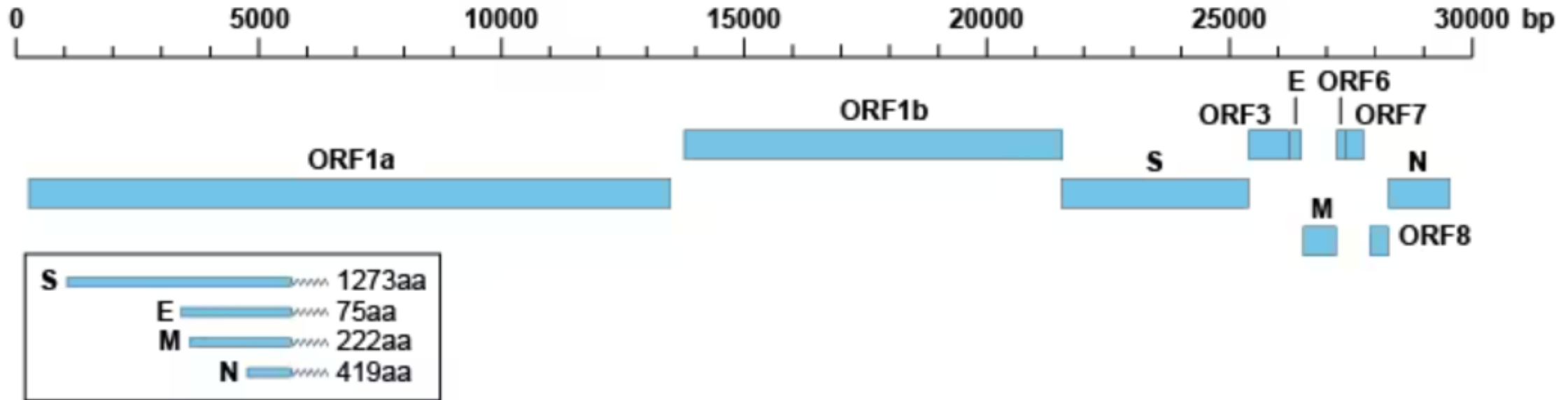


Hypothesis #3: there was a sequencing error and BC genomes really possess the C17747T mutation

How do the mutations of the genomes with the C17747T mutation compare to those without?



What is WA-S566?



Mutation 1: ~~C17747T~~

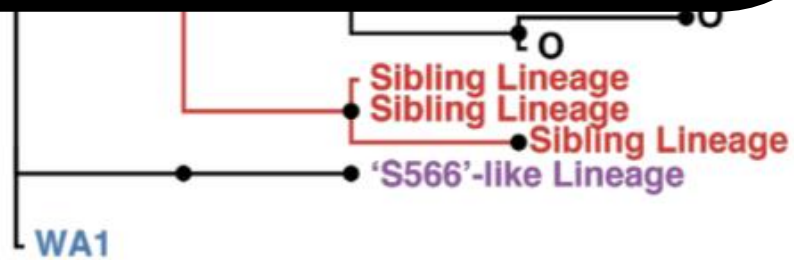
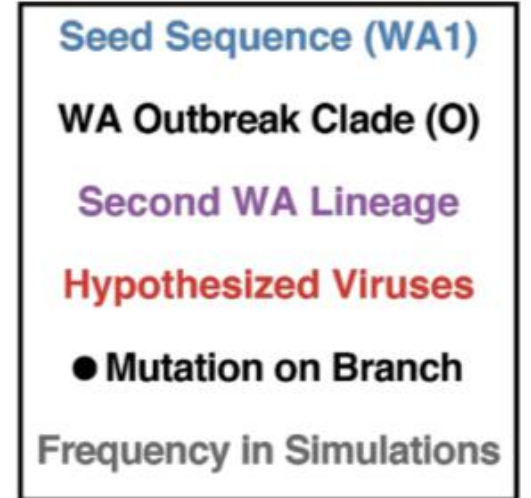
Mutation 2: ~~A17858G~~

Differs from WA1 at 7 other sites

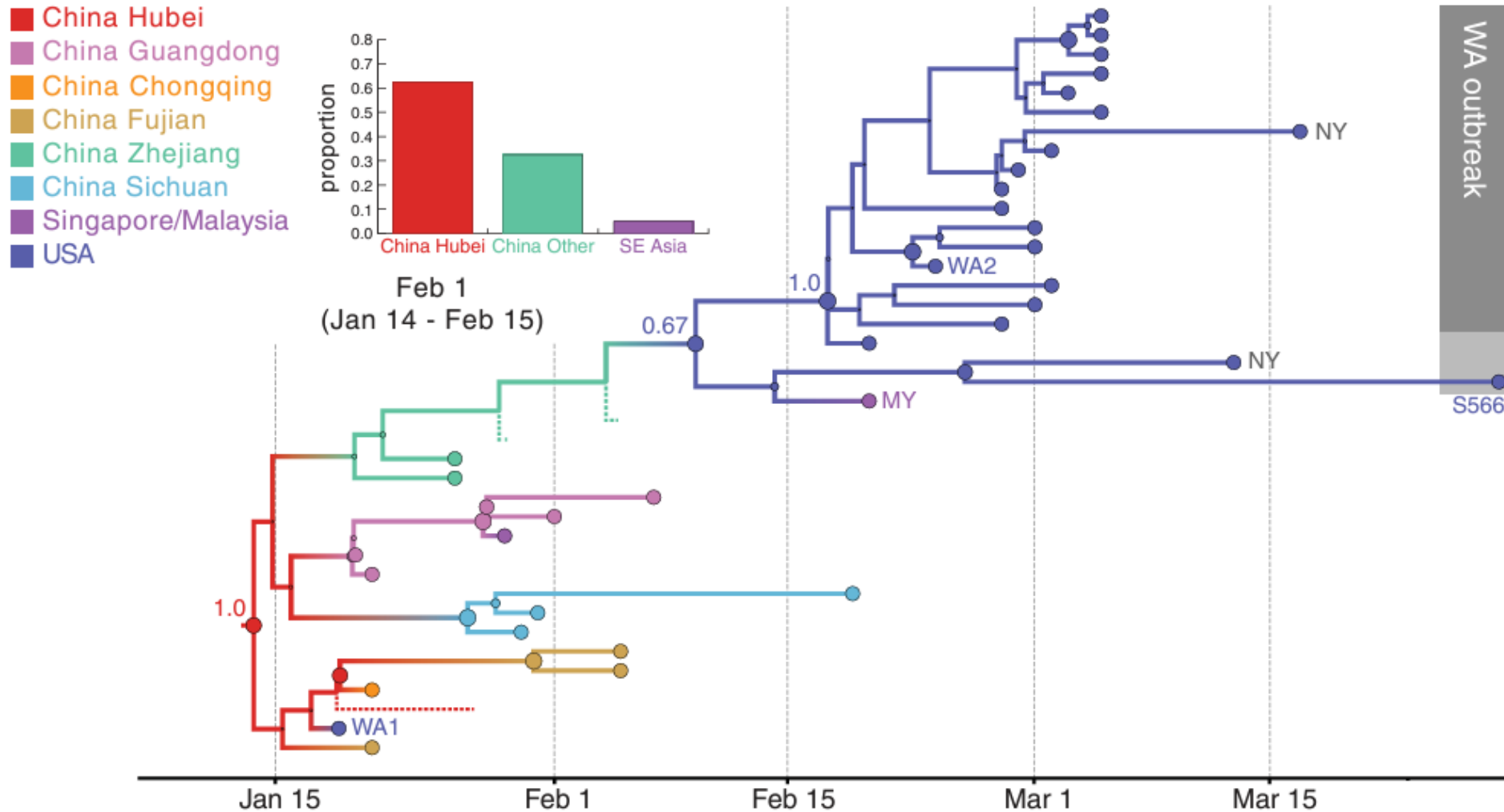
How many divergent lineages were observed before the rise of the 2 characteristic WA clade mutations?

(N = 1000)

“...it is unlikely that, had it been the ancestral virus, WA1 would have given rise to only the S566 lineage and the WA outbreak clade.”



How was COVID-19 introduced to North America?



What does the phylogenetic tree of all SARS-CoV-2 strains look like?



Summary



The WA1 strain was not the ancestral strain of the WA outbreak clade.



There was an error in sequencing of some BC strains that resulted in a critical C→T mutation being missed.



The outbreak of the pandemic in North America was a result of multiple introductions of the SARS-CoV-2 virus.

References

Slide 2: https://evogeneao.s3.amazonaws.com/images/tree_of_life/tree-of-life_1000.png

Slide 3: <https://images.theconversation.com/files/220682/original/file-20180529-80620-175vclv.png?ixlib=rb-1.1.0&q=15&auto=format&w=754&h=1281&fit=crop&dpr=3>

Slide 3: <https://www.mun.ca/biology/scarr/139417.jpg>

Slide 4: [Penn software helps to identify course of cancer metastasis, tumor 'evolution' \(medicalxpress.com\)](#)

https://medlineplus.gov/images/PX000060_PRESENTATION.jpeg

[Understanding Nextstrain | Fred Hutchinson Cancer Center](#)

Slide 9: <https://images.schoolmouv.fr/svt-1ere-cours-3-img1-02.png>

Slide 10: https://assets-global.website-files.com/621e95f9ac30687a56e4297e/64a8d74bcf720ca09f0f023c_V2_1676212488941_acecc7dd-4fb7-40ba-b8b9-968ab7adaeb7.png

Slide 11: [Phylogeny for the faint of heart: a tutorial – ScienceDirect](#)

Slide 12:

https://th.bing.com/th/id/R.8ca994cab98fe2bad94ec08521666994?rik=FOlrneO05YoXRg&riu=http%3a%2f%2fwww.clustal.org%2fimages%2fclustalo_big.png&ehk=96u3azm9aqXcMkSjyGw9rH91sedGJ%2b9MFWJ9URRo6l%3d&risl=&pid=ImgRaw&r=0&sres=1&sresct=1

<https://th.bing.com/th/id/R.25bfda022c5fe894010df205da99955f?rik=EGerAhLcNbfKQA&riu=http%3a%2f%2ftcoffee.crg.cat%2fapps%2ftcoffee%2fpublic%2fcircles%2ftcoffee-msa.png&ehk=CiuPmG%2bEeAnPwCJd7Skp%2bcyeNW%2fi12pnxpuXc6z0Ayw%3d&risl=&pid=ImgRaw&r=0>

Slide 13: [Nucleotide Databases- Definition, Types, Examples, Uses \(microbenotes.com\)](#)

https://h3abionet.org/images/2019/workshop_call_images/ensembl_logo.png

Slide 14: [Distance methods of phylogenetic tree construction involve converting... | Download Scientific Diagram \(researchgate.net\)](#)

Slide 17: [Phylogenetic tree building in the genomic age | Nature Reviews Genetics](#)

Overall: [Phylogeny for the faint of heart: a tutorial – ScienceDirect](#)

[Phylogenetics Algorithms and Applications - PMC \(nih.gov\)](#)

Slide 14: [Fig. 2: Major biological factors that contribute to incongruence. | Nature Reviews Genetics](#)

[Phylogenetic tree building in the genomic age | Nature Reviews Genetics](#)

Slide 22: https://govsite-assets.s3.amazonaws.com/9FsL81tShOeNiCGaLapB_covid-19-1330px.jpg

Slide 23:

Slide 24: https://media.springernature.com/full/springer-static/image/art%3A10.1038%2Fs41392-021-00653-w/MediaObjects/41392_2021_653_Fig1_HTML.png

Slide 25: https://www.researchgate.net/figure/Life-cycle-of-positive-and-negative-sense-single-stranded-RNA-ssRNA-viruses-A-B_fig1_343955700

Slide 26: https://media.springernature.com/full/springer-static/image/art%3A10.1038%2Fs41392-021-00653-w/MediaObjects/41392_2021_653_Fig1_HTML.png

Slide 28: <https://www.novusbio.com/support/sars-cov-research-resources>

Slide 32: <https://www.novusbio.com/support/sars-cov-research-resources>

Slide 34: <http://www.bccdc.ca/health-info/diseases-conditions/covid-19/archived-b-c-covid-19-data>

Slide 34: <https://doh.wa.gov/data-and-statistical-reports/diseases-and-chronic-conditions/communicable-disease-surveillance-data/respiratory-illness-data-dashboard>

Slide 35: https://img.freepik.com/premium-vector/window-operating-system-error-warning-dialog-window-popup-message-with-system-failure-flat-design_812892-54.jpg

Slide 37: <https://www.novusbio.com/support/sars-cov-research-resources>



https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.fredhutch.org%2Fen%2Fnews%2Fcenter-news%2F2021%2F09%2Ftrevor-bedford-macarthur-fellowship.html&psig=AOvVaw0vne19Vg4ljqr_SihigmOt&ust=1707362373216000&source=images&cd=vfe&opi=89978449&ved=OCBMQjRxqFwoTcliH5M-imIQDFQAAAAAdAAAAABAQ

