

Virus sequence data during an outbreak

National Center for Biotechnology Information (NCBI)
at the National Library of Medicine (NLM), NIH

Eneida L. Hatcher, Ph.D.

March 10, 2021

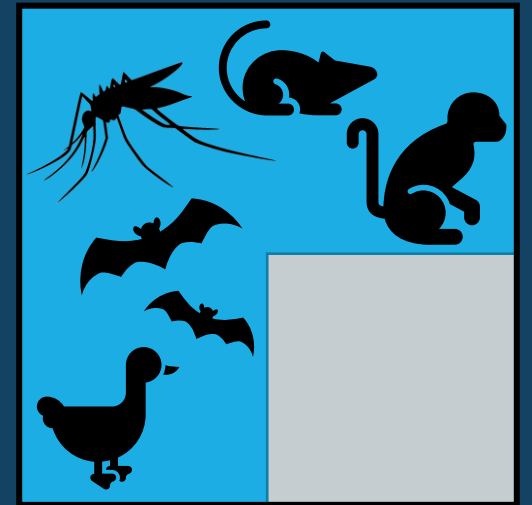


Agenda

- Viruses & hosts
- Genomics
- Surveillance & lineages

What are viruses?

- Intracellular parasites – cannot replicate outside of a host cell
- Inner nucleic acid core surrounded by an outer casing of proteins known as the capsid
 - Sometimes have a lipid envelope covering the capsid
- ~200 different viruses can cause a cold!
- We know a tiny fraction of the viruses in the world



3 out of 4 new
human
diseases come
from animals

Increasing risks of zoonoses

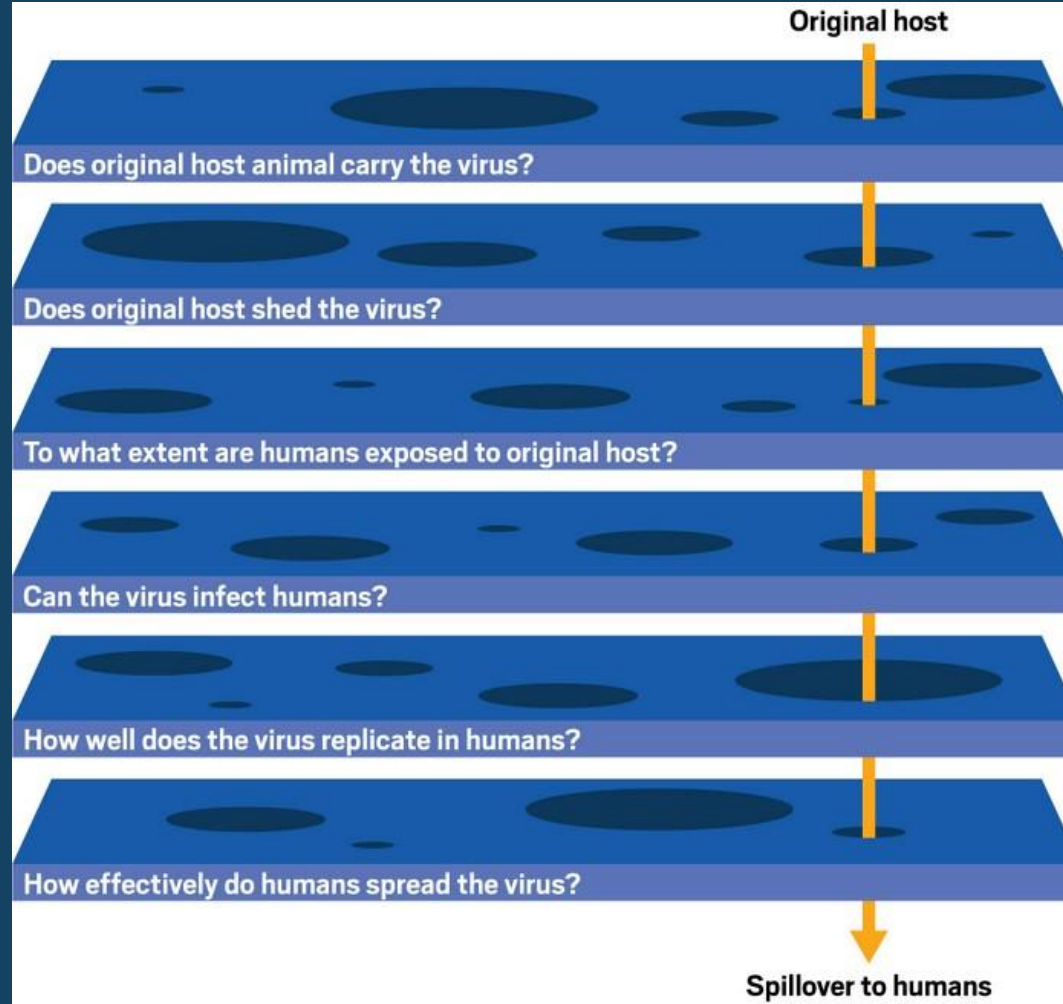
- Expanding housing and food needs
 - living or farming closer to wildlife habitats
 - more interactions between humans or livestock and wildlife
 - creates opportunities for spillover of previously unknown pathogens
- High mobility of new host
- Climate change
 - Insects which carry disease are expanding their range

• Jones BA, Grace D, Kock R, et al. Zoonosis emergence linked to agricultural intensification and environmental change. *Proc Natl Acad Sci U S A*. 2013;110(21):8399-8404. doi:10.1073/pnas.1208059110

• Peter Piot quote in Leslie Hook; Oct 2020; The next pandemic: where is it coming from and how do we stop it?; Financial Times Magazine; <https://www.ft.com/content/2a80e4a2-7fb9-4e2c-9769-bc0d98382a5c>

Can we predict zoonotic viruses?

- Identify variables associated with spillovers
- The variables change over space & time, making it harder to measure risk
- Characterize wildlife viruses to identify potential pathogens



How do you name a “new” virus?

- Viral disease: Assigned by the International Classification of Diseases (ICD), which is managed by World Health Organization
 - coronavirus disease 2019, COVID-19
- Virus species: Determined by the International Committee on Taxonomy of Viruses (ICTV)
 - Gives us a way to talk about objects which are (somehow) related to each other
 - Species: *Severe acute respiratory syndrome-related coronavirus*
 - Virus name: severe acute respiratory syndrome coronavirus 2, SARS-CoV-2
- Official names take time, NCBI works to identify synonyms to make data searches easier
- 2019-nCoV, Wuhan-Hu-1, SARS-CoV-2

Finding relatives

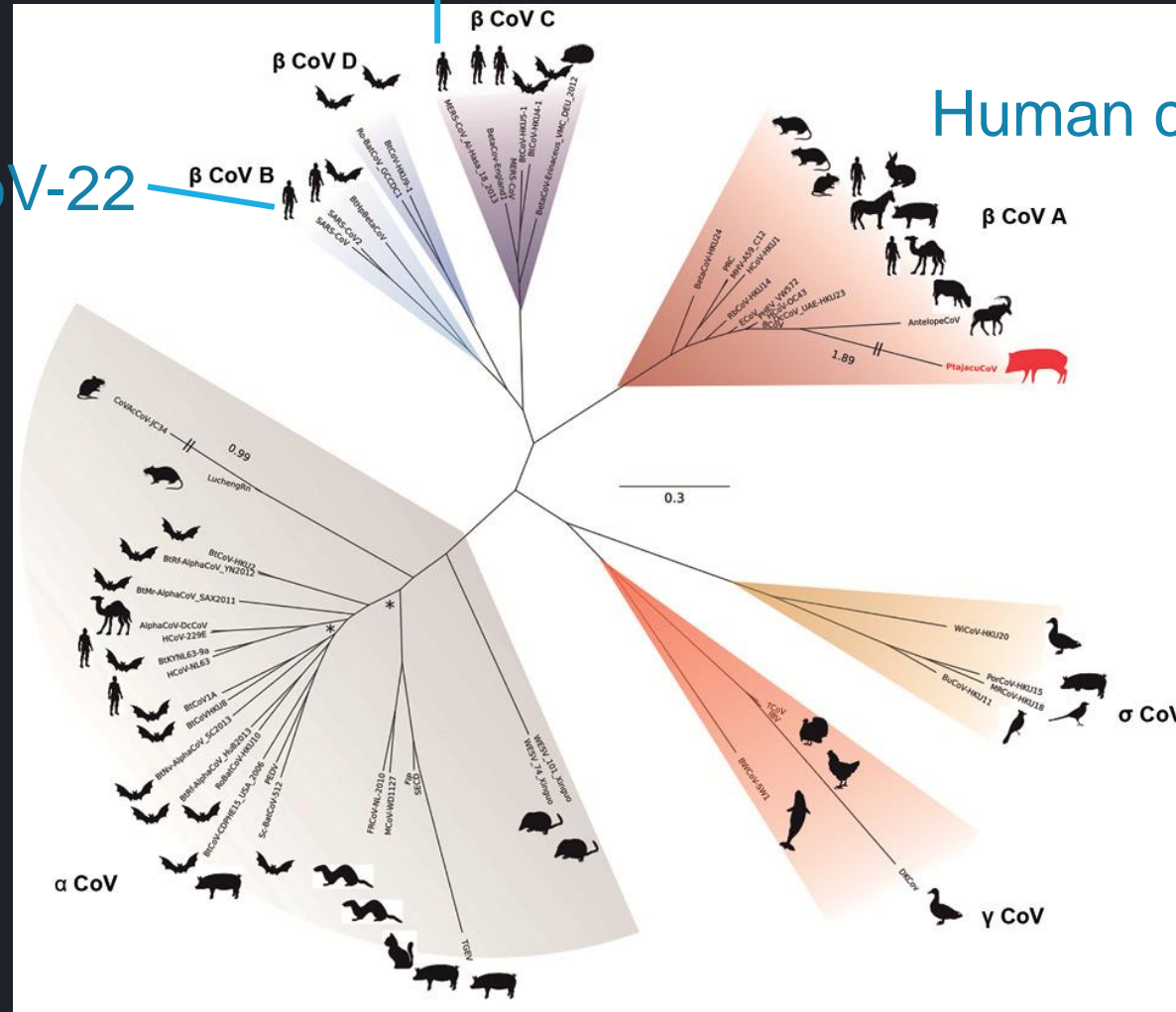
- When sequence of new virus is available, use searches like BLAST to find similar sequences
- Gives context to a new virus
- Identify genes
- Possible biology and replication mechanism
- Sometimes – inform possible treatments

MERS

SARS & SARS-CoV-22

Human colds

Human colds

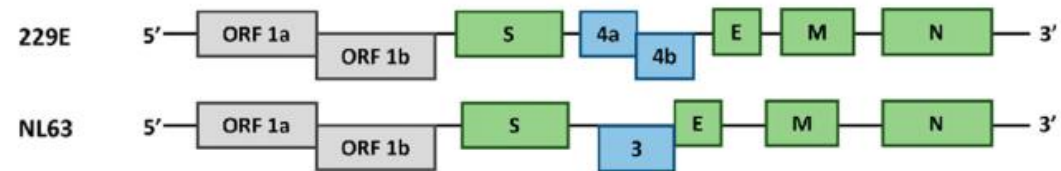


SARS-like
coronaviruses seem
to exist in swarm of
close cousins in the
wild

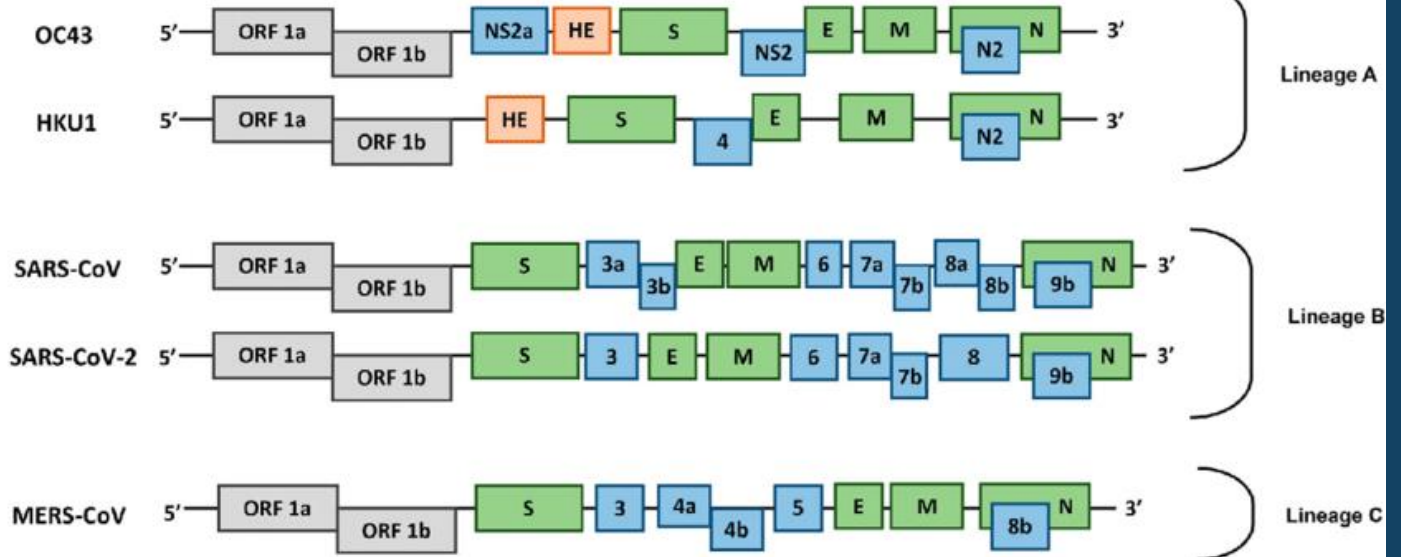
Coronavirus genomes



Alphacoronavirus



Betacoronavirus



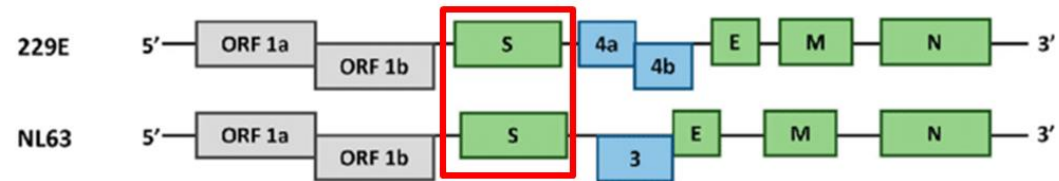
Coronavirus genomes



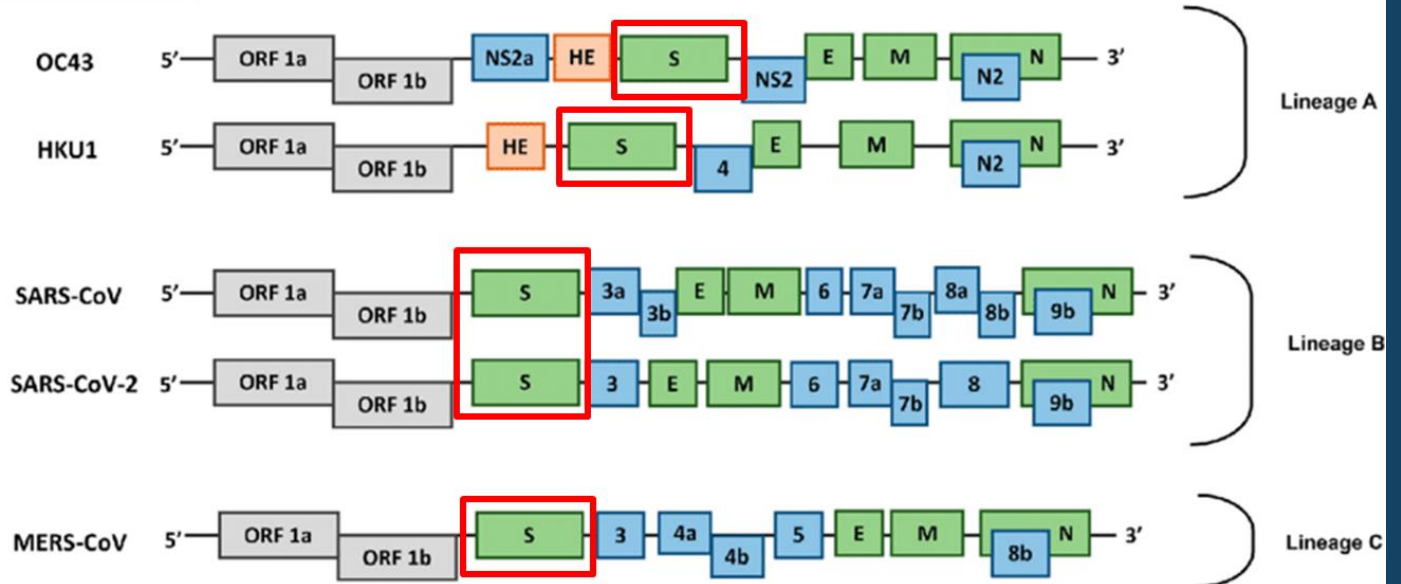
Spike proteins

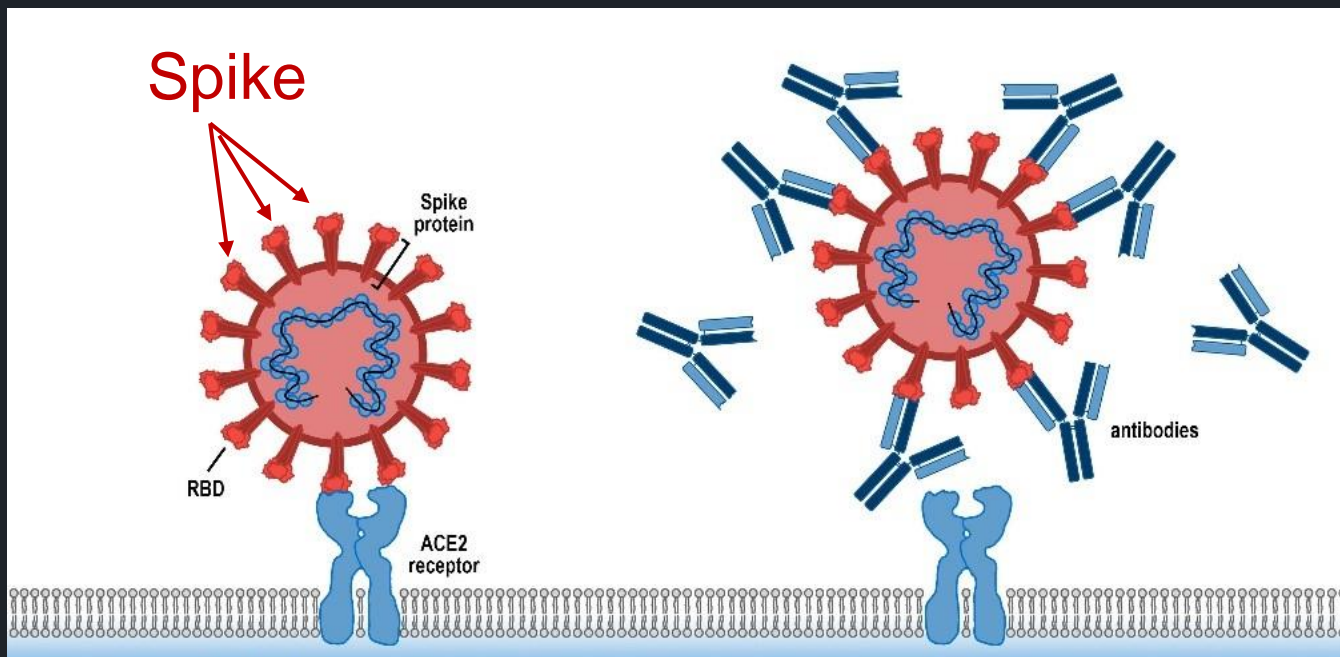


Alphacoronavirus



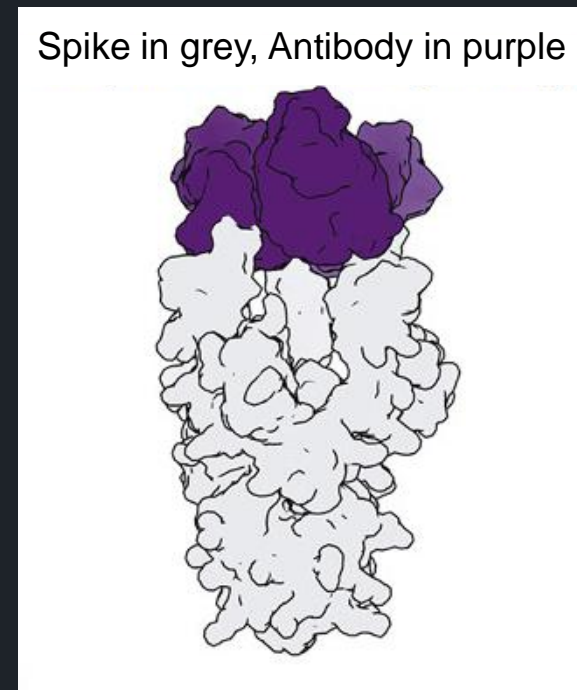
Betacoronavirus





No prior infection or vaccine = no immunity

Antibodies block contact



Some mutations can decrease binding to antibodies produced against another variant of the virus

- Viola Denninger, Sebastian Fiedler, Alison Ilsley, Heike Fiegler and Sean Devenish. October, 2020. Accurate solution phase affinity profiling of a SARS-CoV-2 antibody in serum <https://www.labonline.com.au/content/consumables/sponsored/accurate-solution-phase-affinity-profiling-of-a-sars-cov-2-antibody-in-serum-343712551>
- Antibody Uses Mimicry to Block SARS Coronavirus; April 2019; <https://als.lbl.gov/antibody-uses-mimicry-to-block-sars-coronavirus/>

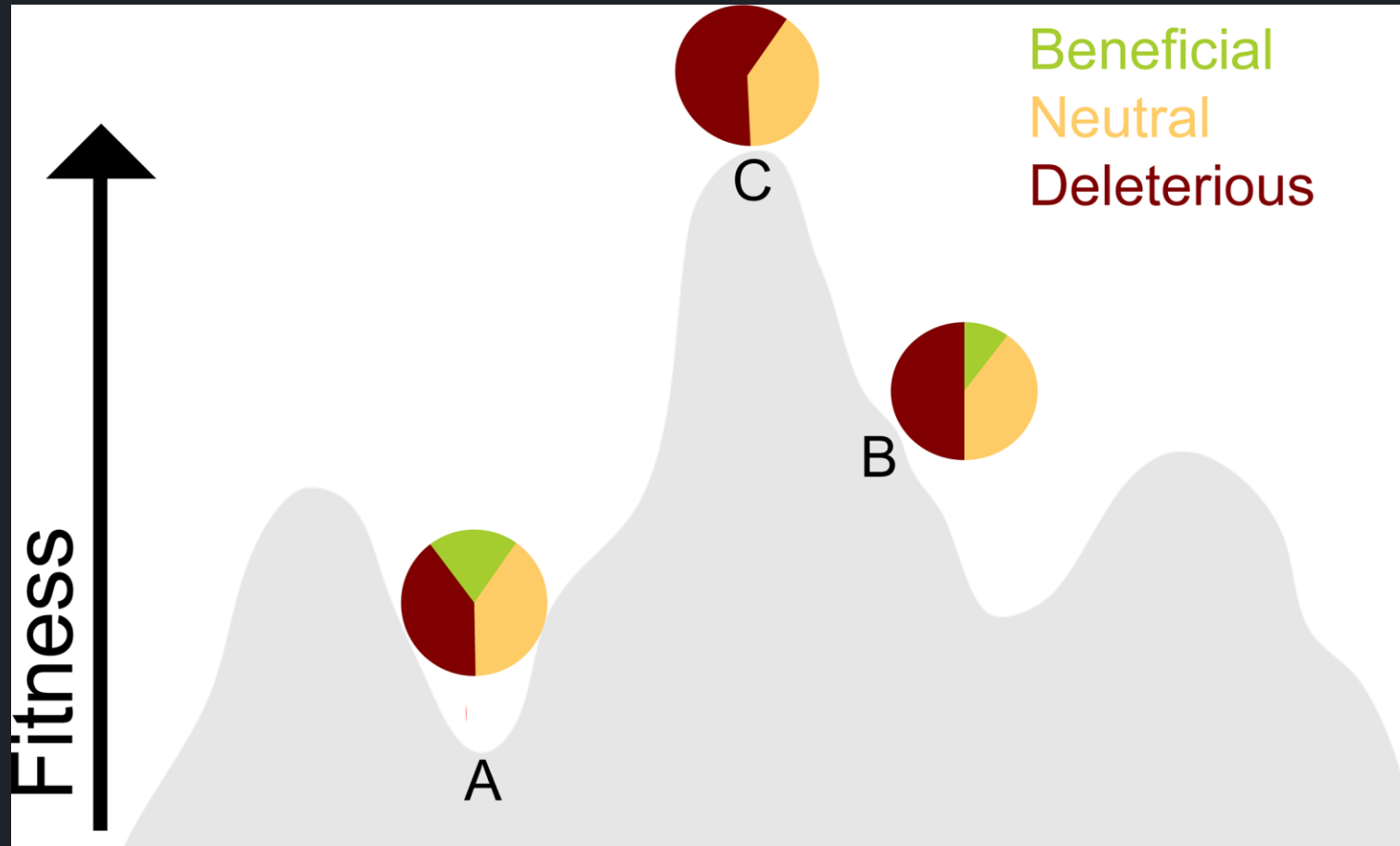
Mutation rate



- More infections = more random mutations = increased chances of a mutation set that will greatly benefit the virus

After a host-jump, virus likely to have low fitness in new host, but many possible mutations which would increase fitness

When viruses have spent a lot of evolutionary time infecting a host, more likely to have high fitness, and few options for beneficial mutations



More data, more problems?

- At the time you are seeing this, NCBI probably has ~100,000 SARS-CoV-2 sequences
- Becomes difficult to parse the data
- Common processes like alignments and phylogenetic trees get computationally daunting
- How do you find data for viruses that have certain characteristics?
 - From a geographic location, or collected recently
 - With mutations in important regions

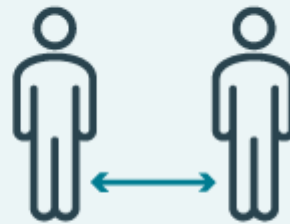
Lineages

- Most SARS-CoV-2 variants have multiple mutations
 - Lineages defined by constellations of mutations
 - Humans like definitions, nature does not
- Popular lineage classifications
 - Pangolin: B.1.1.7, B.1.135
 - NextClade: 20A, 20C
 - 501Y.v1, 501Y.v2
- Better: ORF3:N120Y, S:N501Y, S:D614G, S:GLY843Δ

ACT NOW!



WEAR A MASK



STAY 6 FEET APART

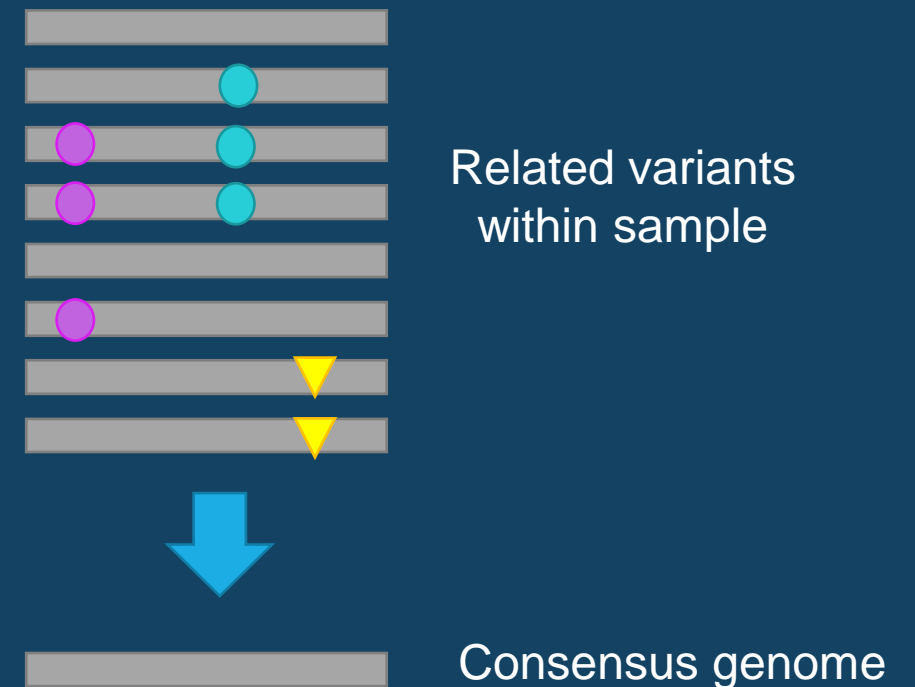


AVOID CROWDS

Applies to all lineages

But wait, there's more...

- Virus mutates while in an individual
- The sample may include multiple variants
- The “genome” is an average
- The raw data is a better representation of the actual infection





Questions?



Eneida.Hatcher@NIH.gov

Visit our new NCBI Virus SARS-CoV-2 Data Hub
www.ncbi.nlm.nih.gov/labs/virus/vssi/#/sars-cov-2

(Art by David S. Goodsell, <http://pdb101.rcsb.org/sci-art/goodsell-gallery>)