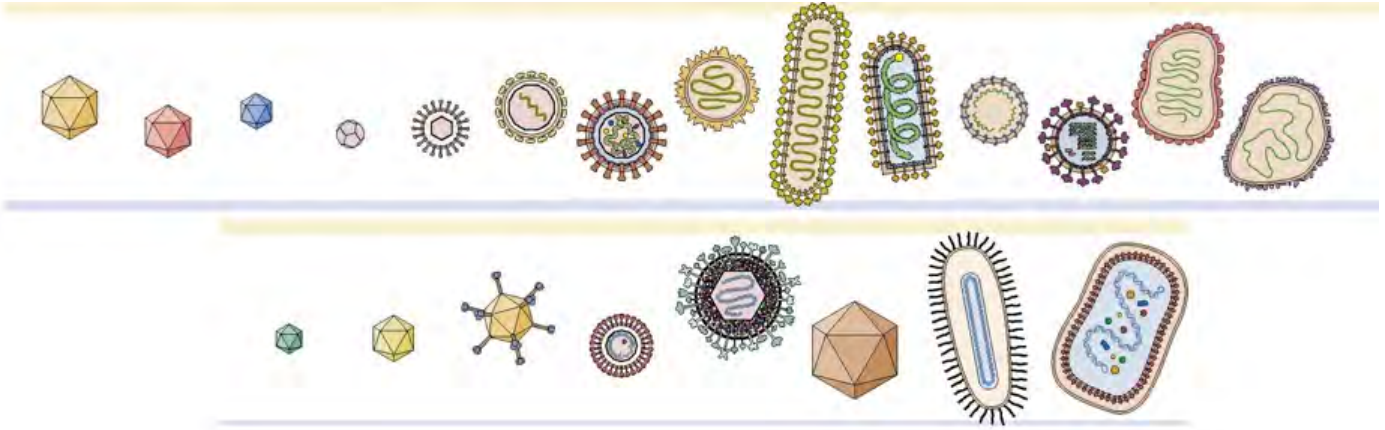
02/04/21

JOHNS HOPKINS
UNIVERSITY

Rapid Sequencing of SARS-CoV-2 to enable Epidemiologic Surveillance, Clinical Insight, and Pathobiology

Winston Timp
Johns Hopkins University
Department of Biomedical Engineering

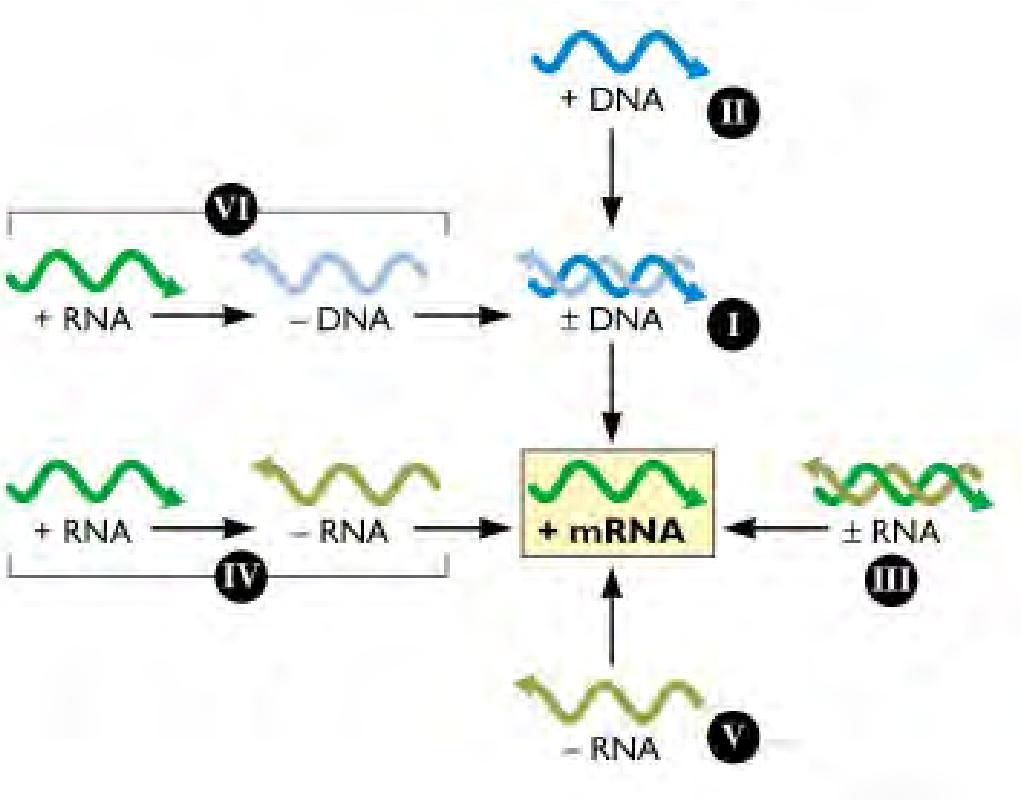
There are many kinds of viruses



And at least as many mechanisms of hijacking cells

Classification

- Baltimore Virus Classification System
- Based on the viral lifecycle and how nucleic acid is translated
- Influenza is a type V virus
- SARS-CoV-2 is a type IV virus
- Positive strand RNA viruses are especially concerning because – in principle, the viral genome itself could be translated to protein.



Coronaviruses

- Generally there are 7 coronaviruses that affect humans
- You haven't heard of these 4:
  - HCoV-OC43
  - HCoV-HKU1
  - HCoV-229E
  - HCoV-NL63
- They are mild and cause 15% of the "common cold"
- You have heard of these three:
  - SARS-CoV – AKA SARS (2003)
    - Infected 8000 with a CFR of 10%
  - MERS-CoV – AKA MERS (2012)
    - Total cases ~2500, CFR 37%
  - SARS-CoV-2 (COVID19)

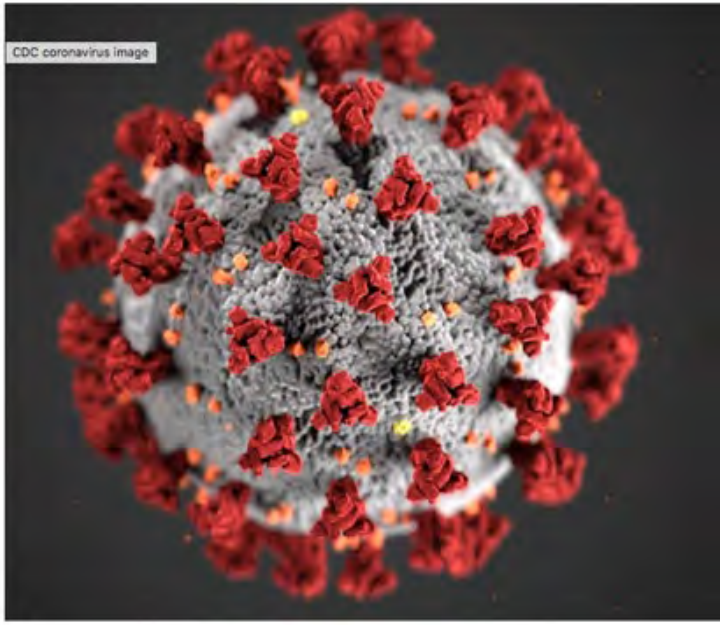
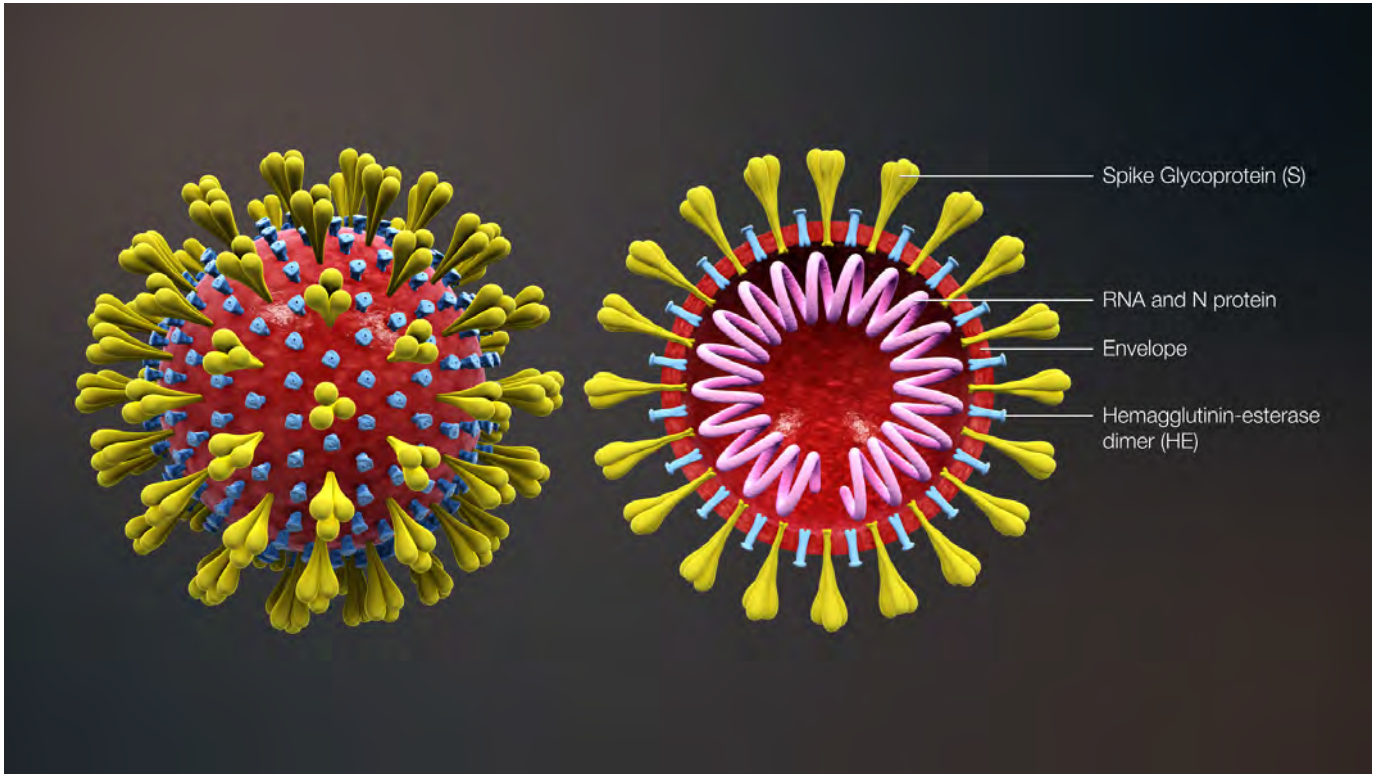


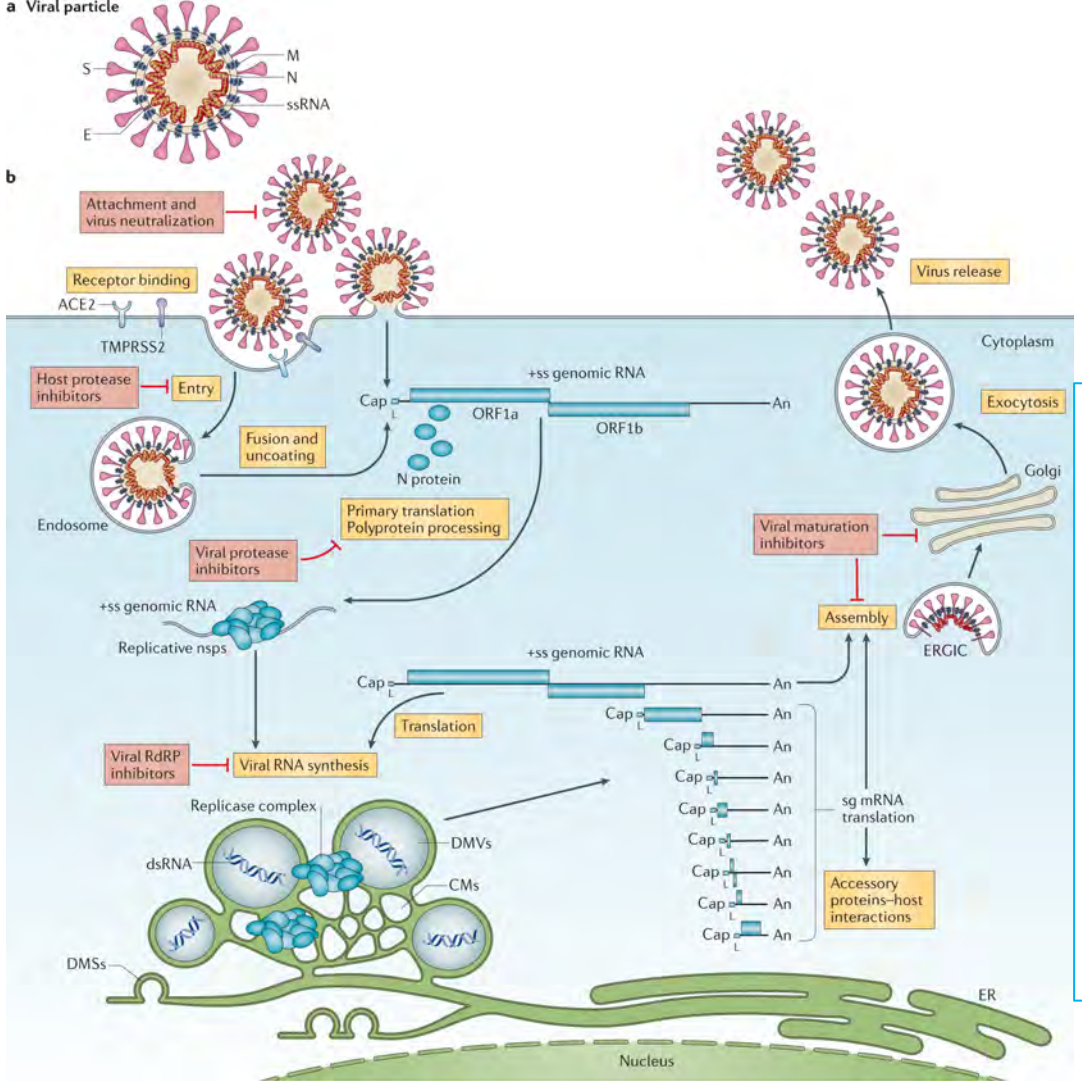
Image: CDC/Alissa Eckert, MS; Dan Higgins, MAMS

The Bad Guy


Spike Glycoprotein (S)
RNA and N protein
Envelope
Hemagglutinin-esterase dimer (HE)

CoV infection cycle

- Virus enters cell through either endocytosis or membrane fusion
- The viral RNA gets translated to make helicase and replicase-transcriptase complex (RdRp)to make more RNA, and the proteins from the virus
- Packaged up in Golgi apparatus, leaves the cell



V'kovski et al Nature Reviews Microbiology 2020

First genome

Novel 2019 coronavirus genome

Novel 2019 coronavirus

edward_holmes 6 Jan 10

> 10th January 2020
> This posting is communicated by Edward C. Holmes, University of Sydney on behalf of the consortium led by Professor Yong-Zhen Zhang, Fudan University, Shanghai

The Shanghai Public Health Clinical Center & School of Public Health, in collaboration with the Central Hospital of Wuhan, Huazhong University of Science and Technology, the Wuhan Center for Disease Control and Prevention, the National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control, and the University of Sydney, Sydney, Australia is releasing a coronavirus genome from a case of a respiratory disease from the Wuhan outbreak. The sequence has also been deposited on GenBank (accession MN908947 20.0k) and will be released as soon as possible.

Update: This genome is now available on GenBank and an updated version has been posted 20.0k.

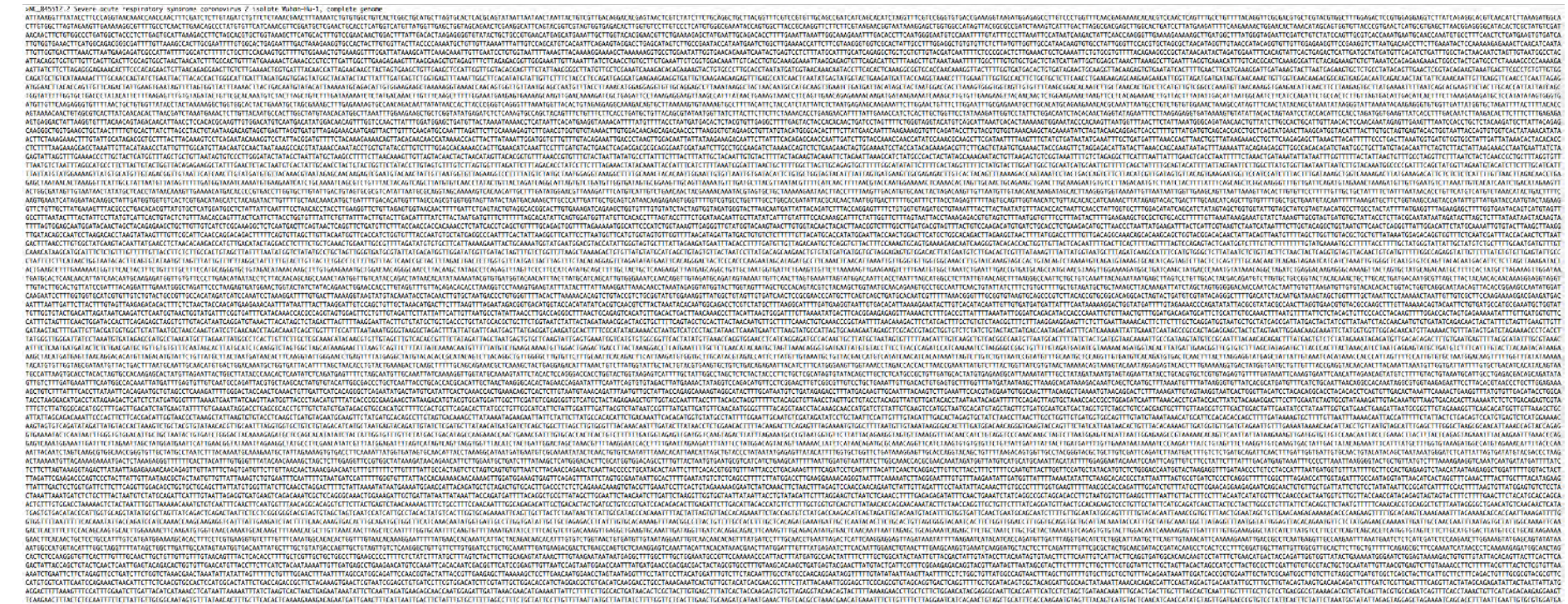
Disclaimer:
Please feel free to download, share, use, and analyze this data. We ask that you communicate with us if you wish to publish results that use these data in a journal. If you have any other questions –then please also contact us directly.

Professor Yong-Zhen Zhang,
Shanghai Public Health Clinical Center & School of Public Health,
Fudan University,
Shanghai, China.

email: zhangyongzhen@shphc.org.cn

1

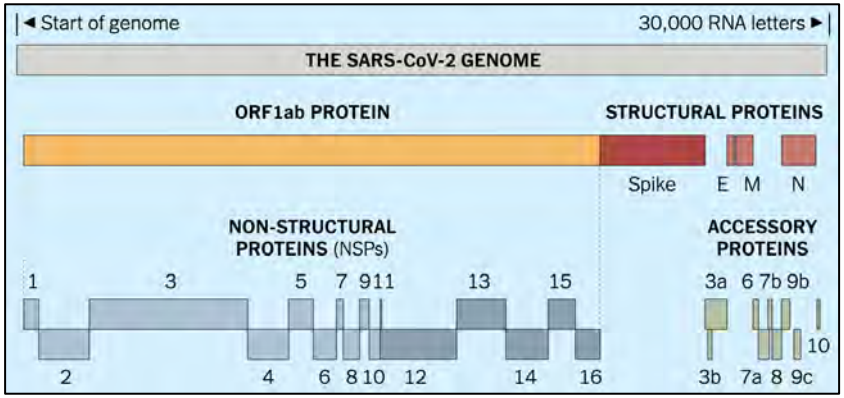
SARS-CoV-2 Genome



https://www.ncbi.nlm.nih.gov/nuccore/NC_045512

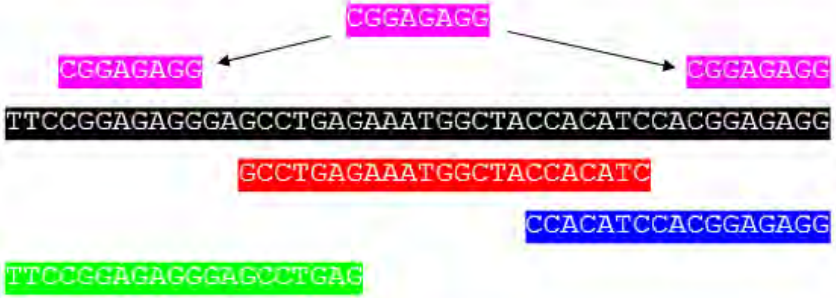
SARS-CoV-2 Viral Genome Characteristics

- **Genus**: Betacoronavirus
  - Lineage A (Embecovirus): hCoV-OC43, hCoV-HKU1
  - Lineage B (Sarbecovirus): SARS-CoV, **SARS-CoV-2**
  - Lineage C (Merbecovirus): MERS-CoV, bat HKU4 and HKU5
  - Lineage D (Nobecovirus): bat coronavirus HKU9
- **Genome structure**
  - Positive-sense RNA virus
  - Polyadenylated
  - 29,903 nucleotide, linear genome
  - 13 open reading frames
  - 29 potential gene products

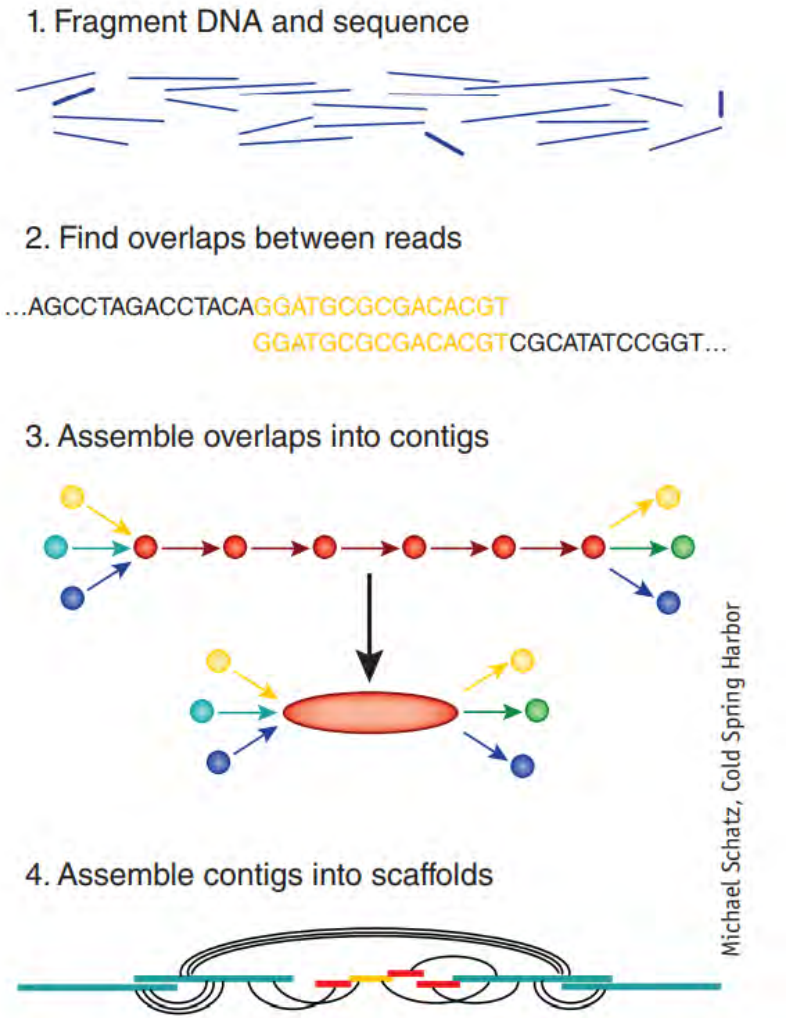


https://www.nytimes.com/interactive/2020/04/03/science/coronavirus-genome-bad-news-wrapped-in-protein.html

Genome Assembly



- To assemble from reads, you look for overlaps and try to stitch it back together
- Like a jigsaw puzzle
- Bigger the pieces, the easier the puzzle
- Many computational tools exist to do this: e.g. canu, spades, megahit



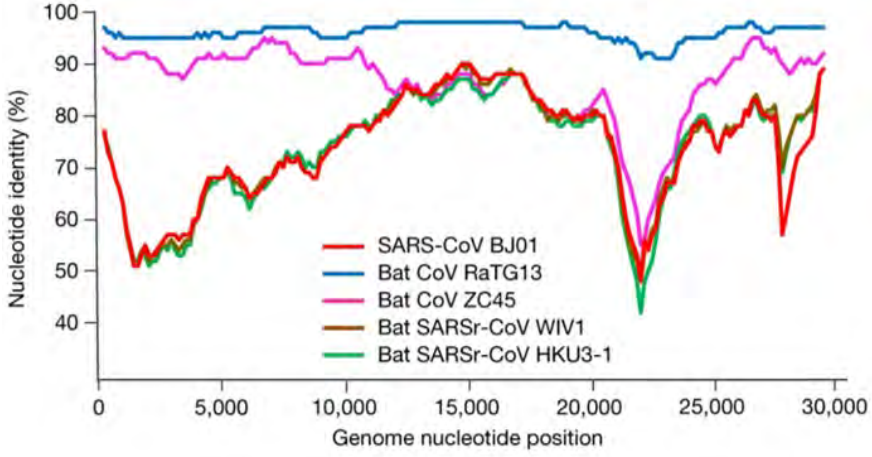
Family roots of SARS-CoV-2

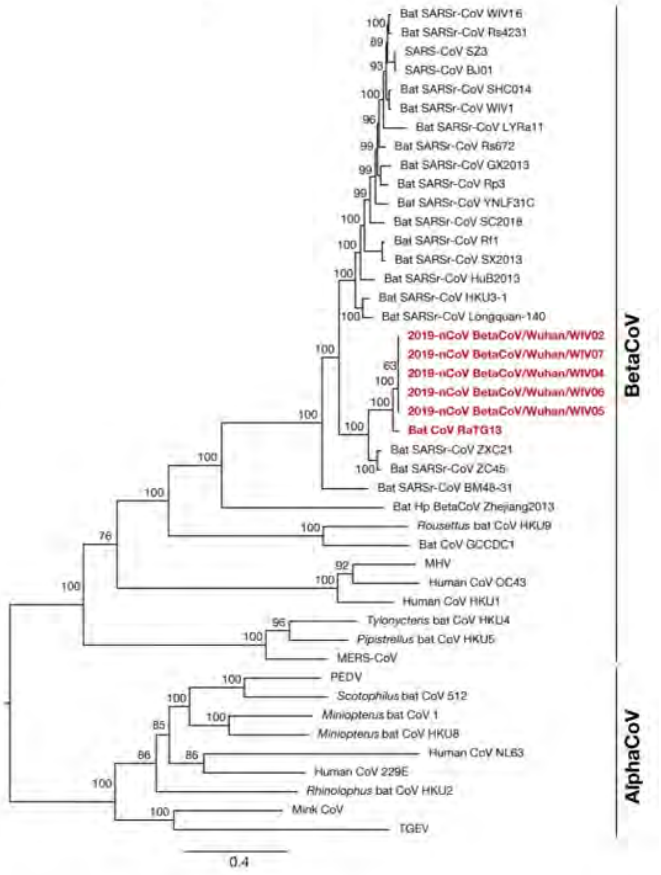
Article | Open Access | Published: 03 February 2020

A pneumonia outbreak associated with a new coronavirus of probable bat origin

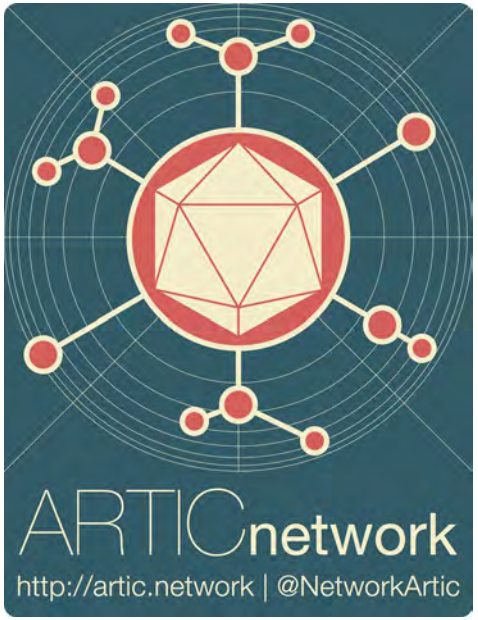
Peng Zhou, Xing-Lou Yang, [...] Zheng-Li Shi

Nature **579**, 270–273(2020) | Cite this article



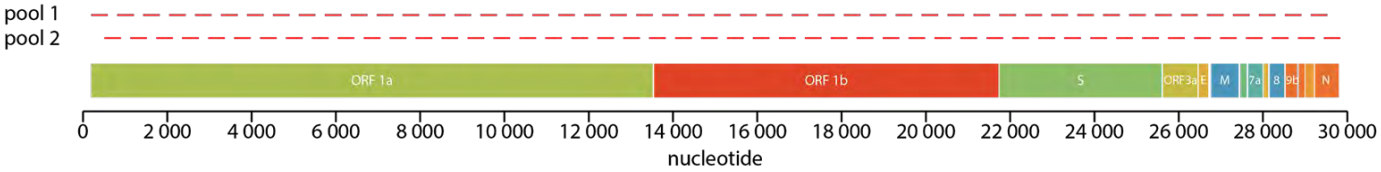


ARTIC network



Leveraging work by the ARTIC network (Rambaut, Goodfellow, Loman, Rowe) - using two pools of tiled amplicons.

SARS-CoV-2 genome (30kb) tiled using Primal Scheme (Primer3 wrapper) with 98 primers in 2 pools:



https://artic.network/ncov-2019

Quick et al Nature Protocols 2017

Initiating Sequencing in the JHH Diagnostics Laboratory

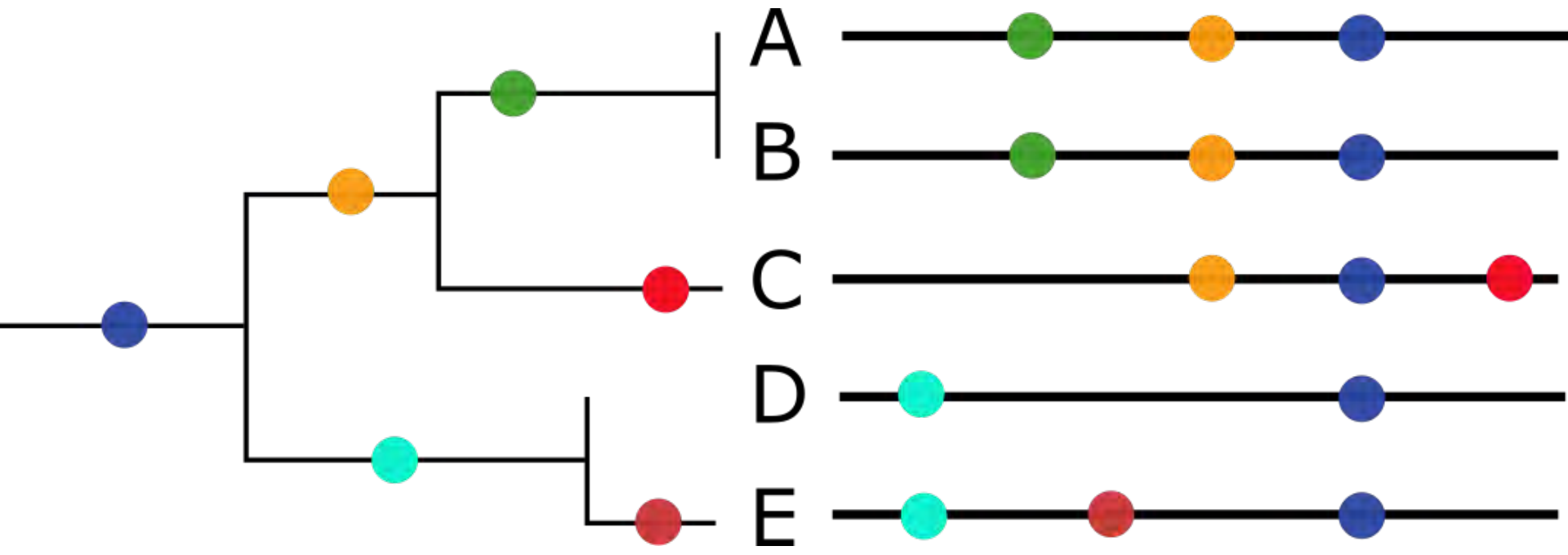
Initial sequencing capacity was established 3/5/20

- Sequenced positive control for diagnostic PCR
- 100% concordance with published genome
- Limited sample throughput (ARTIC V1, 1/flowcell)

Expanded capacity initiated 4/1/20

- Optimized data flow and informatics
- Revised sample preparation approach (ARTIC v3)
- Increased throughput (10+ per flowcell)
- Large batch capabilities (80+ samples / plate)

Comparing Sequences – Phylogenetic Trees



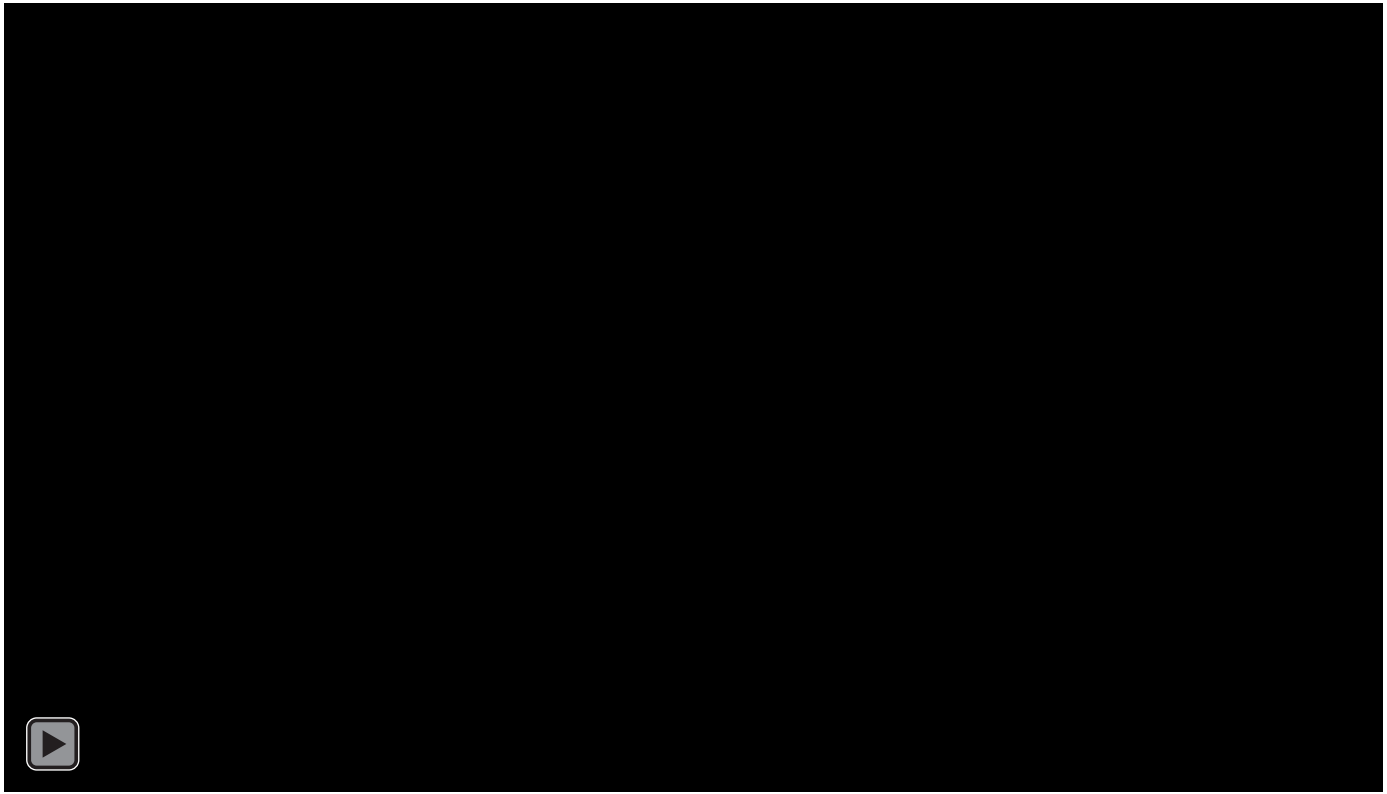
https://nextstrain.org/narratives/trees-background/?n=3

Reconstructing Viral Transmission Networks

Spatial transmission networks can be reconstructed by tracking:

- **Mutations** in viral genomes
- **Date** of collection
- **Location** of collection

This was demonstrated in real-time during the 2014-15 Ebola outbreak (right)



http://virological.org/t/phylogeography-of-2014-2015-ebola-virus-epidemic/36

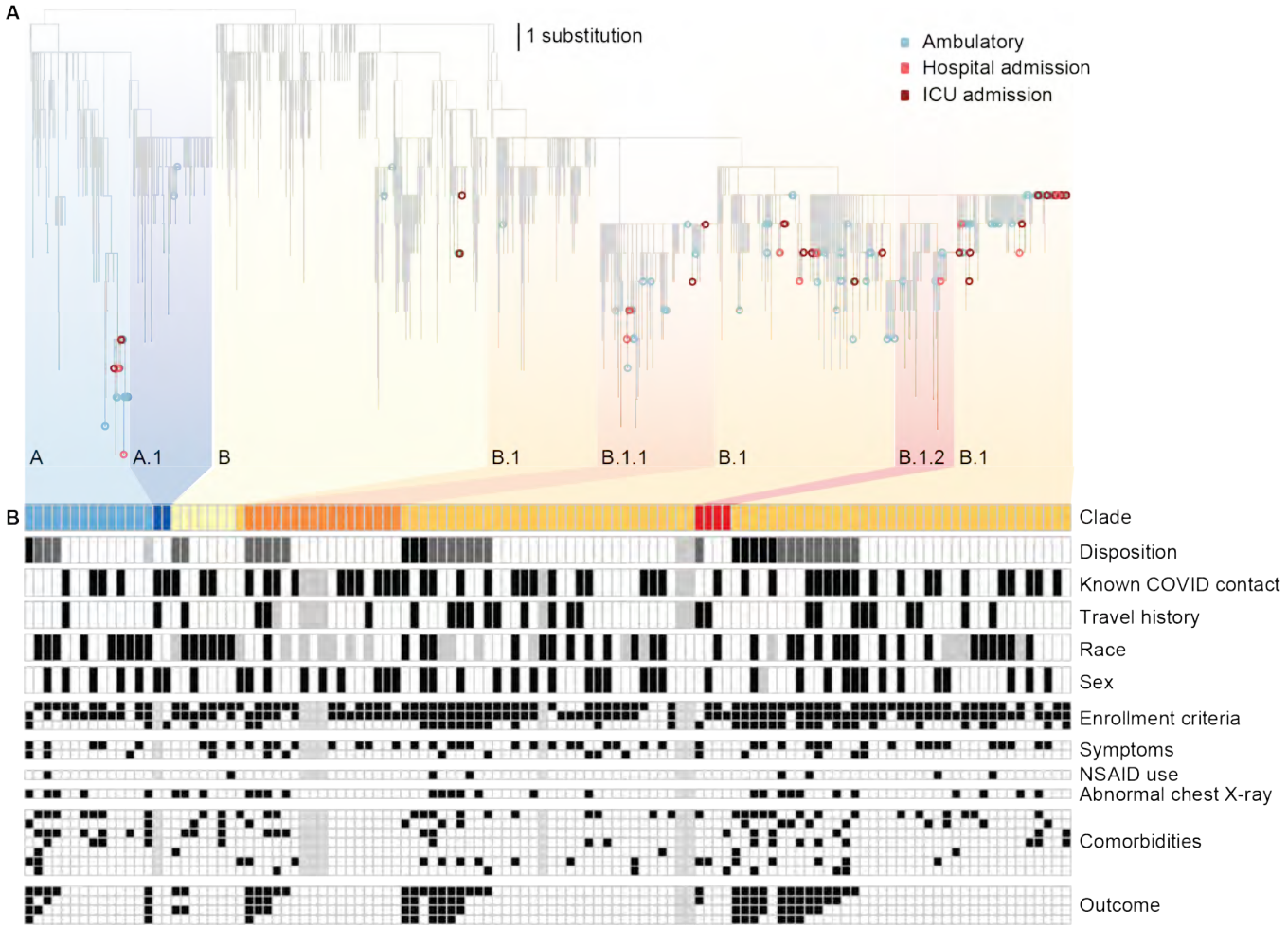
SARS-Cov-2 International Transmission Networks

Hadfield et al *Bioinformatics* 2018

Video generated from Nextstrain.org on 4/18/20

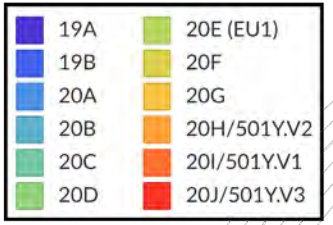
SARS-CoV-2

- We worked over the summer to correlate with initial patient phentotype/co-morbidities and patient outcomes
- We don't have the power to say for sure, but nothing obvious pops up, in agreement with other global data



Sars-CoV-2

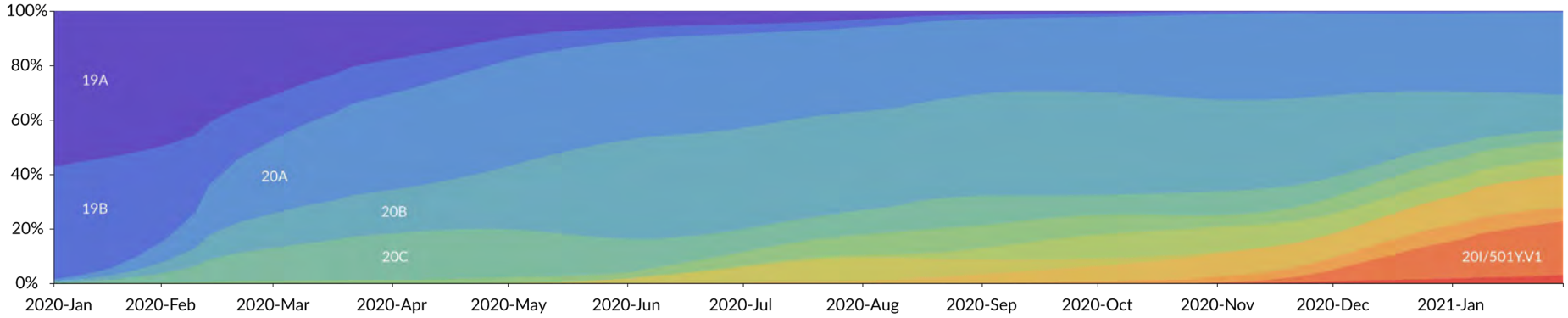
- Current phylogeny shows diversity and evolution of the virus
- Three current variants of concern:



	B.1.1.7	B.1.351	P.1
Alternate name	501Y.V1	501Y.V2	501Y.V3
Country identified	United Kingdom	South Africa	Brazil
Mutations	23	21	17
Spike mutations	8	9	10
Key RBD, spike mutations beyond N501Y in all	E69/70 deletion, P681H 144Y deletion, A570D	E484K, K417N, orf1b deletion	E484K, K417T, orf1b deletion
Other mutations, including N-terminal	T7161, S982A, D1118H	L18F, D80A, D215G, Δ242-244, R264I, A701V	L18F, T20N, P26S, D138Y, R190S, H655Y, T10271
Transmissibility Δ	>50% increased	Not established	Not established
Lethality Δ	Not resolved	?	?
Immune escape/ Vaccine efficacy reduction	Not established Partial in Novavax trial (96->86%)	Yes Partial resistance In 2 vaccine trials	Likely Partial resistance
Countries reported (not = to local transmission)	73	31	9
US States reported	32	2	1 (travel)

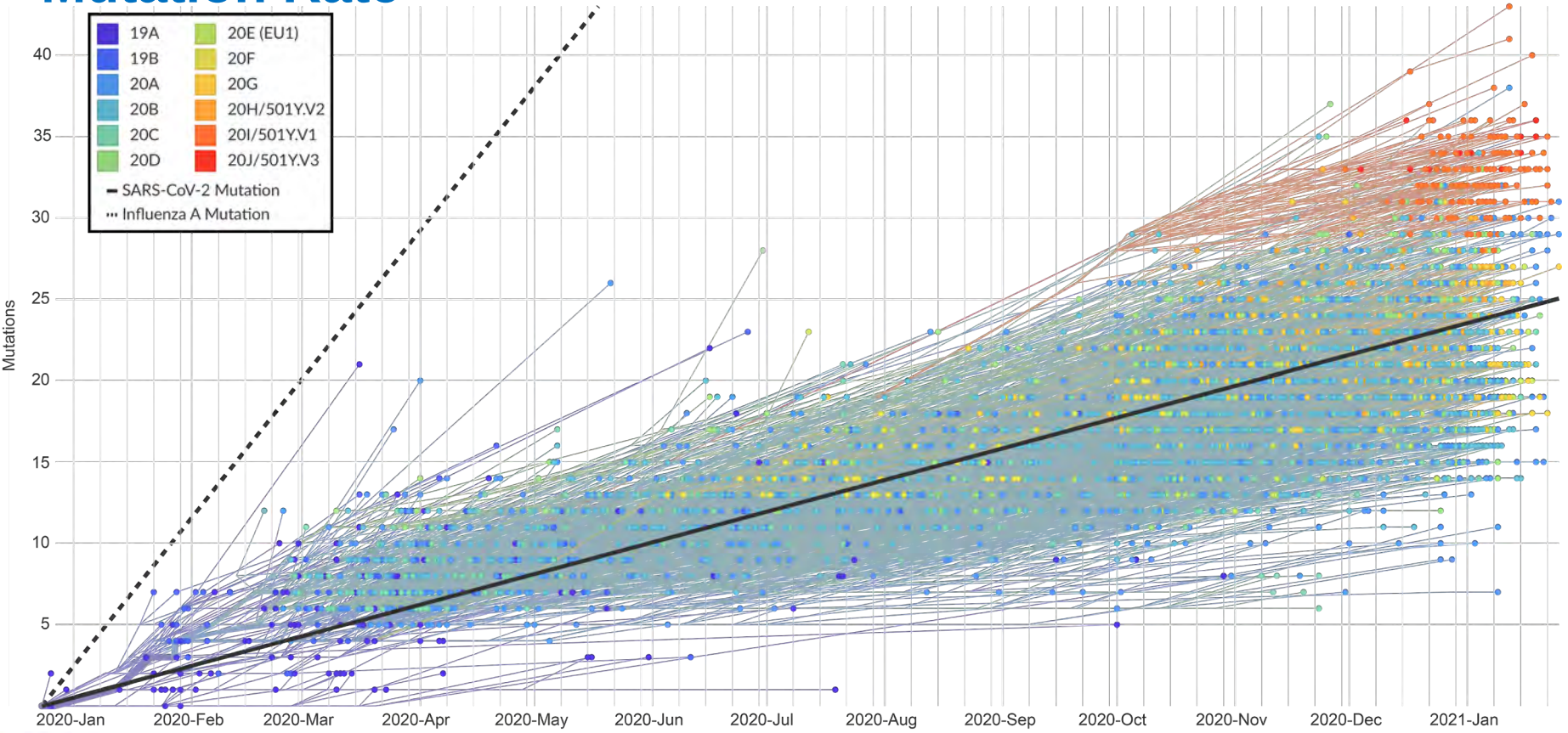
https://twitter.com/EricTopol/status/1356741894718386179?s=20 (2/2/21)

Clades over time



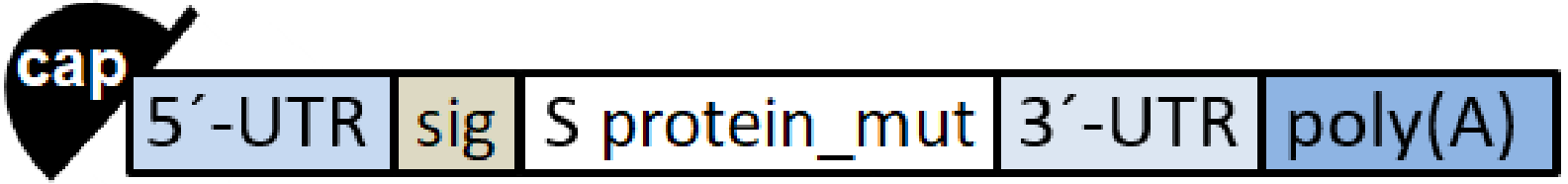
- The original clades have been depleted in favor of later clades – but it seems that this was largely due to sampling effects and different efficicnecies of NPI rather than mutations
- B.1.1.7/20I/501Y.V1 may be different, it seems to have higher transmissibility
- The community largely expects that with vaccine based selective pressure on the virus, we'll start to see more changes in the clade distributions

Mutation Rate



- Mutation Rate for SARS-CoV-2 is estimated as 23 substutions per site per year.
- In contrast, H3N2 influenza A is ~108 substituions per site per year

Vaccines!



- Using the sequence information we have from the virus, we can **rapidly** design vaccines
- mRNA Vaccine was actually designed in essentially a weekend back in March – testing efficacy and safety took longer
- Spike sequence used as the sequence – whole spike for Pfizer, just the RBD for Moderna
- Mutations introduced (K986P, V987P) to stabilize the protein
- Pseudouridine (ψ) instead of Uridine (U) to prevent innate immune uptake and improve translation

```
GAGAAΨAAAC ΨAGΨAΨΨCΨΨ CΨGGΨCCCCA CAGACΨCAGA GAGAACCCGC   50
CACCAΨGΨΨC GΨGΨΨCCΨGG ΨGCΨGCΨGCC ΨCΨGGΨGΨCC AGCCAGΨGΨG  100
ΨGAACCΨGAC CACCAGAACA CAGCΨGCCΨC CAGCCΨACAC CAACAGCΨΨΨ  150
ACCAGAGGCG ΨGΨACΨACCC CGACAAGGΨG ΨΨCAGAΨCCA GCGΨGCΨGCA  200
CΨCΨACCCAG GACCΨGΨΨCC ΨGCCΨΨΨCΨΨ CAGCAACGΨG ACCΨGGΨΨCC  250
ACGCCAΨCCA CGΨGΨCCGGC ACCAAΨGGCA CCAAGAGAΨΨ CGACAACCCC  300
GΨGCΨGCCCΨ ΨCAACGACGG GGΨGΨACΨΨΨ GCCAGCACCG AGAAGΨCCAA  350
CAΨCAΨCAGA GGCΨGGAΨCΨ ΨCGGCACCAC ACΨGGACAGC AAGACCCAGA  400
GCCΨGCΨGAΨ CGΨGAACAAC GCCACCAACG ΨGGΨCAΨCAA AGΨGΨGCGAG  450
ΨΨCCAGΨΨCΨ GCAACGACCC CΨΨCCΨGGGC GΨCΨACΨACC ACAAGAACAA  500
CAAGAGCΨGG AΨGGAAAGCG AGΨΨCCGGGΨ GΨACAGCAGC GCCAACAACΨ  550
GCACCΨΨCGA GΨACGΨGΨCC CAGCCΨΨΨCC ΨGAΨGGACCΨ GGAAGGCAAG  600
CAGGGCAACΨ ΨCAAGAACCΨ GCGCGAGΨΨC GΨGΨΨΨAAGA ACAΨCGACGG  650
CΨACΨΨCAAG AΨCΨACAGCA AGCACACCCC ΨAΨCAACCΨC GΨGCGGGAΨC  700
ΨGCCΨCAGGG CΨΨCΨCΨGCΨ CΨGGAACCCC ΨGGΨGGAΨCΨ GCCCAΨCGGC  750
AΨCAACAΨCA CCCGGΨΨΨCA GACACΨGCΨG GCCCΨGCACA GAAGCΨACCΨ  800
GACACCΨGGC GAΨAGCAGCA GCGGAΨGGAC AGCΨGGΨGCC GCCGCΨΨACΨ  850
AΨGΨGGGCΨA CCΨGCAGCCΨ AGAACCΨΨCC ΨGCΨGAAGΨA CAACGAGAAC  900
GGCACCAΨCA CCGACGCCGΨ GGAΨΨGΨGCΨ CΨGGAΨCCΨC ΨGAGCGAGAC  950
AAAGΨGCACC CΨGAAGΨCCΨ ΨCACCGΨGGA AAAGGGCAΨC ΨACCAGACCA 1000
GCAACΨΨCCG GGΨGCAGCCC ACCGAAΨCCA ΨCGΨGCGGΨΨ CCCCAAΨAΨC 1050
ACCAAΨCΨGΨ GCCCCΨΨCGG CGAGGΨGΨΨC AAΨGCCACCA GAΨΨCGCCΨC 1100
ΨGΨGΨACGCC ΨGGAACCGGA AGCGGAΨCAG CAAΨΨGCGΨG GCCGACΨACΨ 1150
CCGΨGCΨGΨA CAACΨCCGCC AGCΨΨCAGCA CCΨΨCAAGΨG CΨACGGCGΨG 1200
ΨCCCCΨACCA AGCΨGAACGA CCΨGΨGCΨΨC ACAAACGΨGΨ ACGCCGACAG 1250
CΨΨCGΨGAΨC CGGGGAGAΨG AAGΨGCGGCA GAΨΨGCCCCΨ GGACAGACAG 1300
GCAAGAΨCGC CGACΨACAAC ΨACAAGCΨGC CCGACGACΨΨ CACCGGCΨGΨ 1350
```

Acknowledgments

JHU SoM/DoM

Heba Mostafa
Stuart Ray
Oluwaseun Falade-Nwulia
Lauren Sauer
Paul Morris
Victoria Gniazdowski

JHU/APL

Peter Thielen
Thomas Mehoke
Jared Evans
Craig Howser
Brian Merritt
Amanda Ernlund

JHU WSE

Winston Timp
Michael Schatz
Steven Salzberg
Srividya Ramakrishnan
Melanie Kirsche
Norah Sadowski
Yunfan Fan
Sam Kovaka
Ariel Gershman
Alaina Shumate
Ales Varabyou
Alex Szalay
Gerard Lemson
Dmitry Medvedev

JHBSPH

Justin Lessler
Shirlee Wohl

VDR/VPR

Julie Messersmith

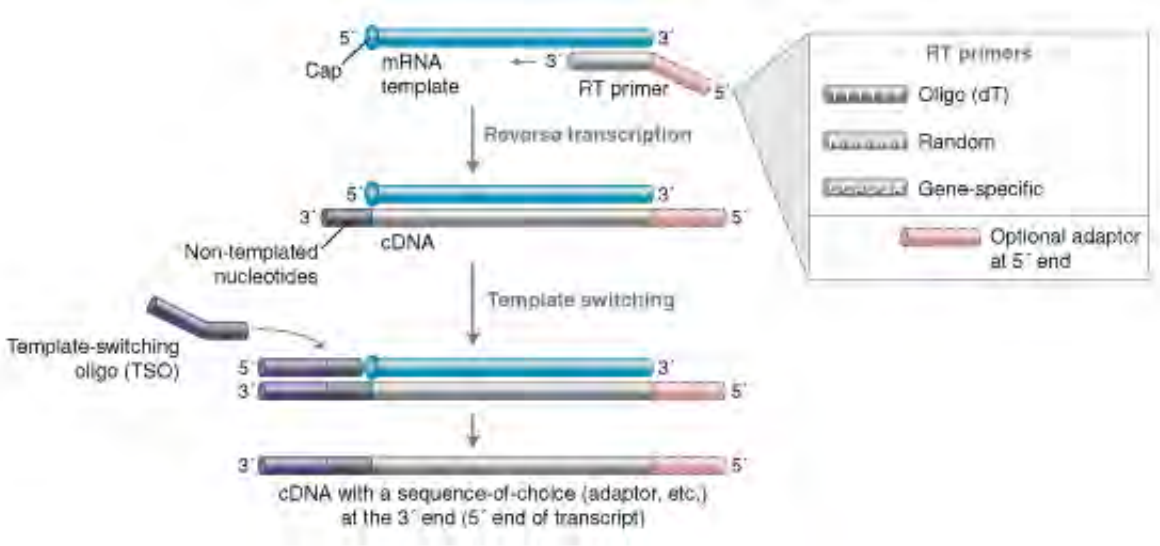
Support from the JHU COVID-19 Research Response Program

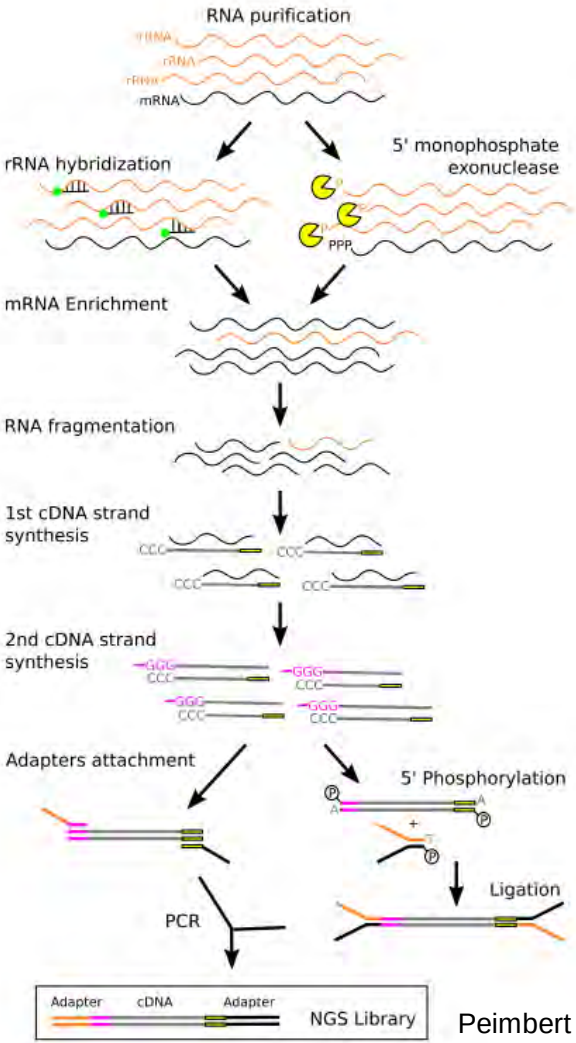
Image from US Consumer Product Safety Commission
https://twitter.com/uscpsc/status/1246558080290217985?lang=en

Extra Slides

RNA -> DNA: Reverse Transcriptase

- Reverse transcriptases originally isolated from viruses – by Baltimore and Temin independently (they shared the Nobel for this in 1975)
- Eukaryotic organisms also have reverse transcriptases – for mobile elements of the genome like retro transposons and for replication at the ends of chromosomes (telomeres)
- Can be primed with poly-dT (against the polyA at the end of mRNA), random, or gene specific esequendces.





Sequencing Methods

- Starts from purified RNA
- Vast majority of RNA is ribosomal, but can be depleted either through digestion at 5' monophosphate or through hybridization and digestion
- Then fragmentation and random priming to make "complentary DNA" or cDNA using a "reverse transcriptase"
- Then make a sequencing library as "normal"

Peimbert et al 2016