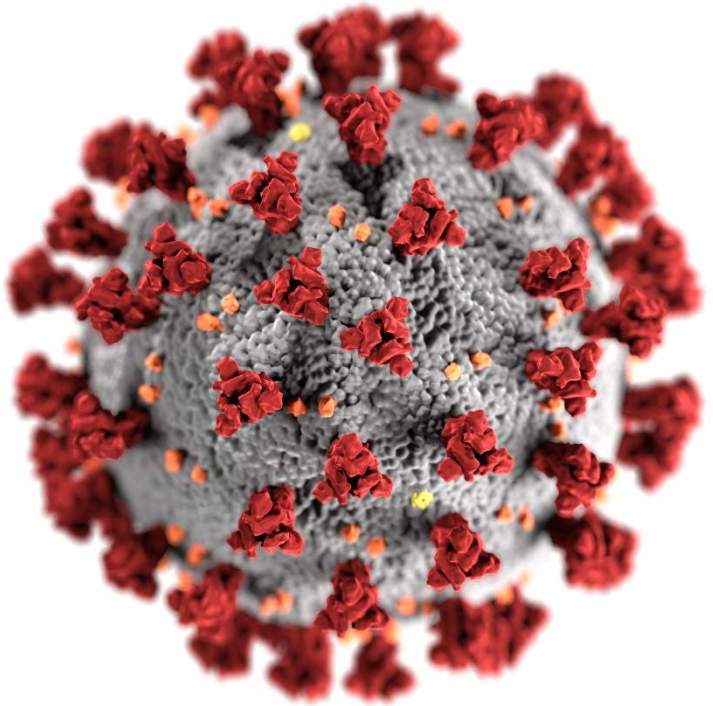
Emerging variants of SARS-CoV-2

COVID-19 Genomic Epidemiology Toolkit: Module 1.4



Michael Weigand, PhD
Bioinformatician
Centers for Disease Control and Prevention

cdc.gov/coronavirus

Toolkit map

Part 1: Introduction

1.1 What is genomic epidemiology?

1.2 The SARS-CoV-2 genome

1.3 How to read phylogenetic trees

1.4 Emerging variants of SARS-CoV-2

Part 2: Case Studies

2.1 SARS-CoV-2 sequencing in Arizona

2.2 Healthcare cluster transmission

2.3 Community transmission

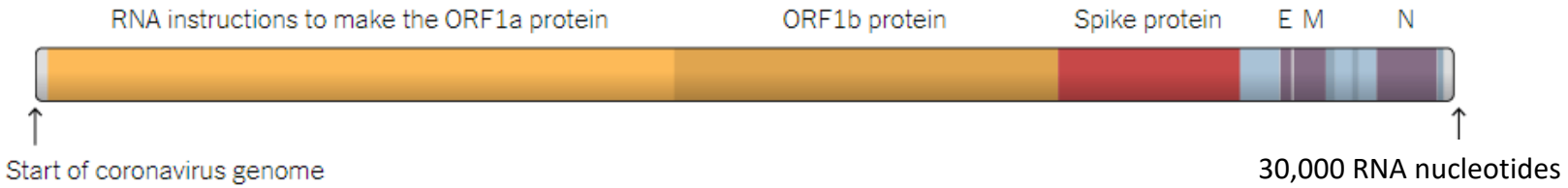
Part 3: Implementation

3.1 Getting started with Nextstrain

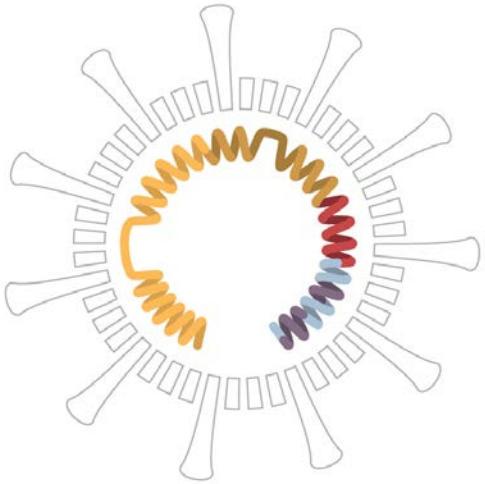
3.2 Getting started with MicrobeTrace

3.3 Phylogenetics with UShER

The SARS-CoV-2 genome



- RNA virus (single-stranded, positive-sense)
- Linear genome = ~30,000 nucleotides
- 11 coding-regions (genes)
- 12 potential gene products
  - e.g., Spike protein



Naqvi et al. (2020) Insight into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approaches, PMID:32544429

Images from *The New York Times* "How Coronavirus Mutates and Spreads" www.nytimes.com/interactive/2020/04/30/science/coronavirus-mutations.html

Fingerprinting and phylogenetics

- Mutations in the genome produce a fingerprint that can be used to infer ancestral relationships (phylogeny), the topic of Module 1.3

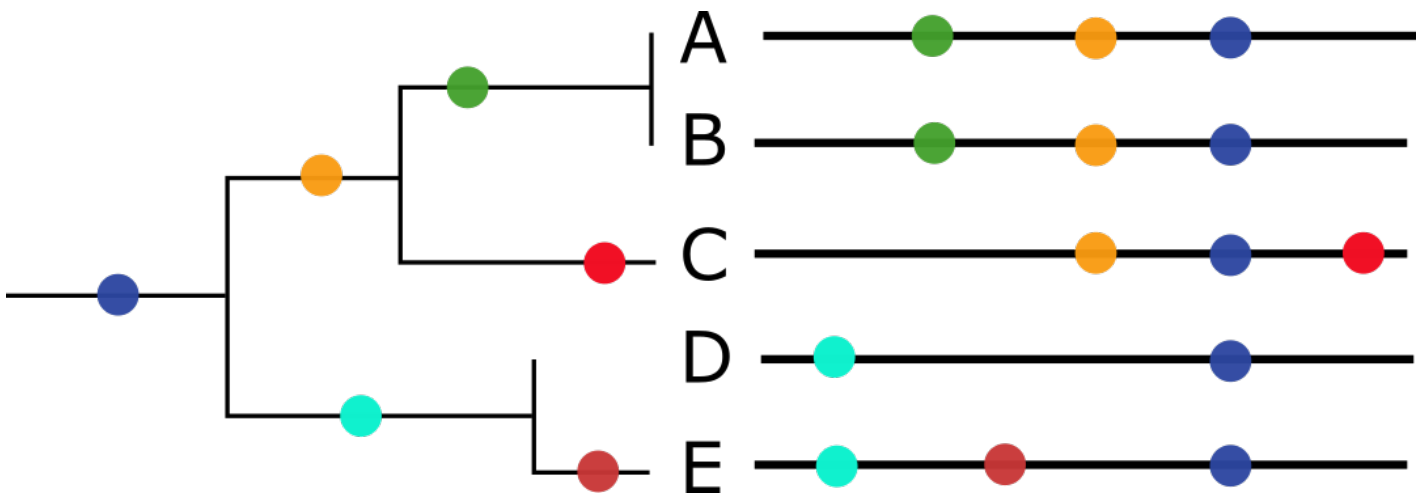


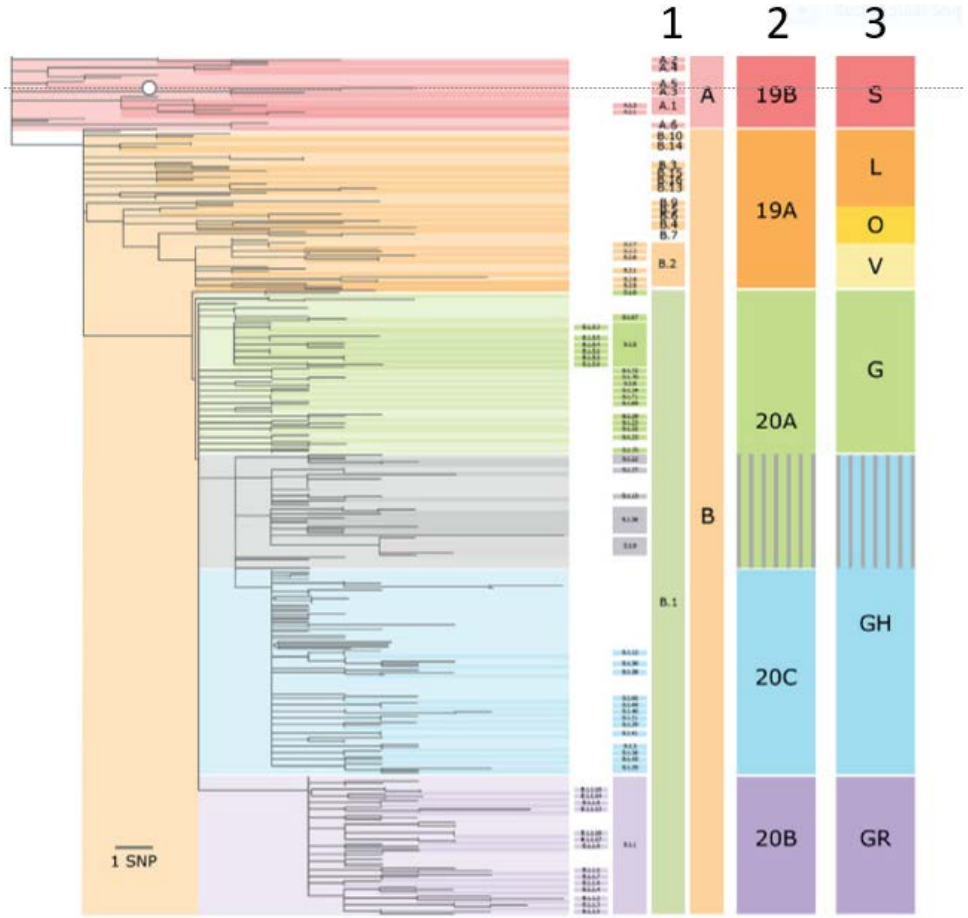
Image from Trevor Bedford Group: https://docs.nextstrain.org

SARS-CoV-2 clades:

Clade naming conventions:

1. PANGO lineages
   - cov-lineages.org
2. Clades by Nextstrain
   - nextstrain.org
3. Clades by GISAID
   - gisaid.org

Clades defined by specific collection of mutations



Adapted from Alm et al. 2020

Defining variants

- “Genetically distinct”
- Clades, lineages
  - B.1.1.7 / 20I
- Specific mutations, or combinations
  - Spike D614G
  - Spike deletion 69
  - Spike E484K, N501Y

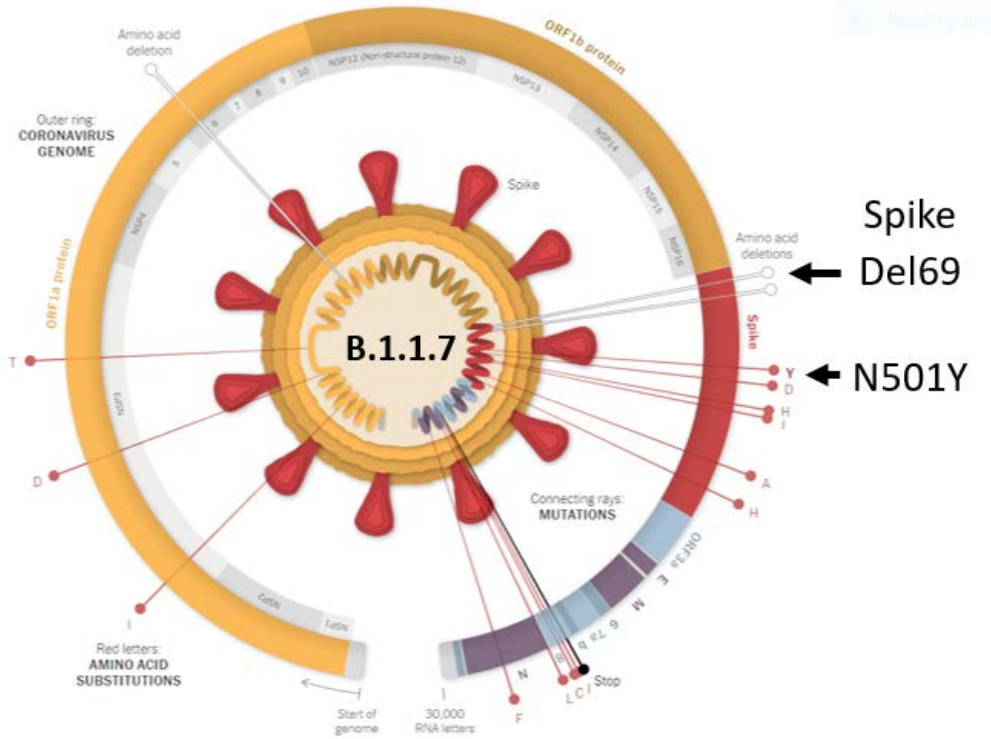


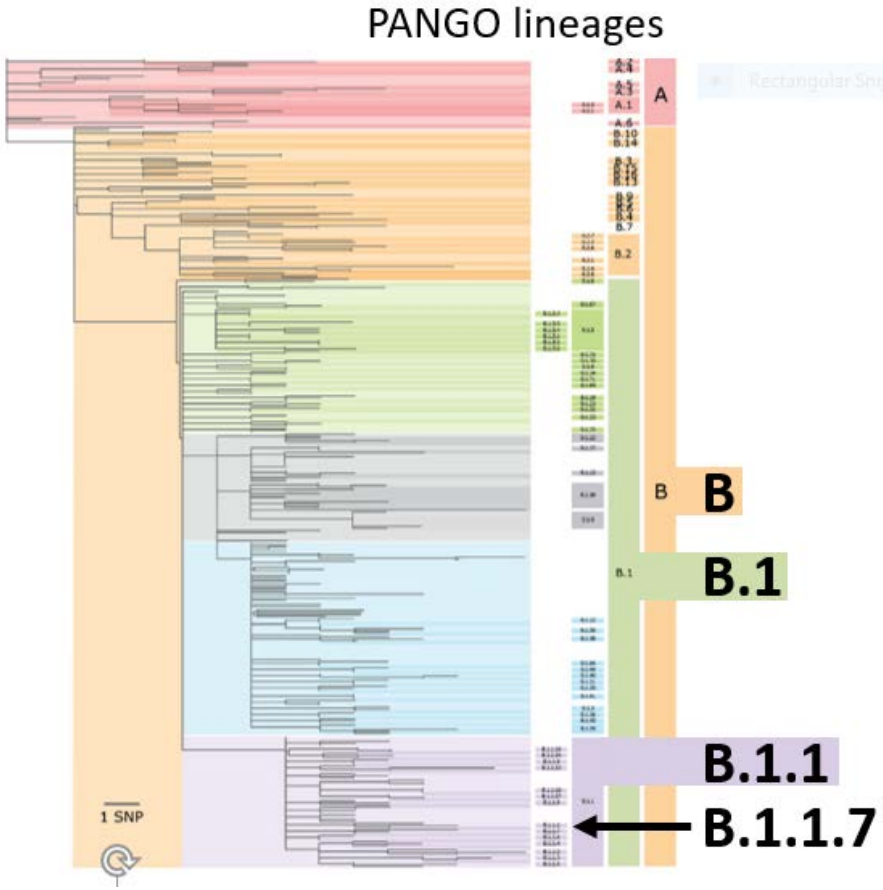
Image from *The New York Times* “Inside the B.1.1.7 Coronavirus Variant” https://www.nytimes.com/interactive/2021/health/coronavirus-mutations-B117-variant.html
Amino acid abbreviations: https://www.ncbi.nlm.nih.gov/Class/MLACourse/Modules/MolBioReview/iupac_aa_abbreviations.html

What's in a name?

- SARS-CoV-2 variant names can reflect lineages or mutations.

- Analogous:
  - B.1.1.7 (PANGO lineage)
  - 501Y.V1 (Spike mutation)
  - 20I (Nextstrain clade)

- L452R
  - Spike mutation at position 452
  - Leucine (L) -> Arginine (R)

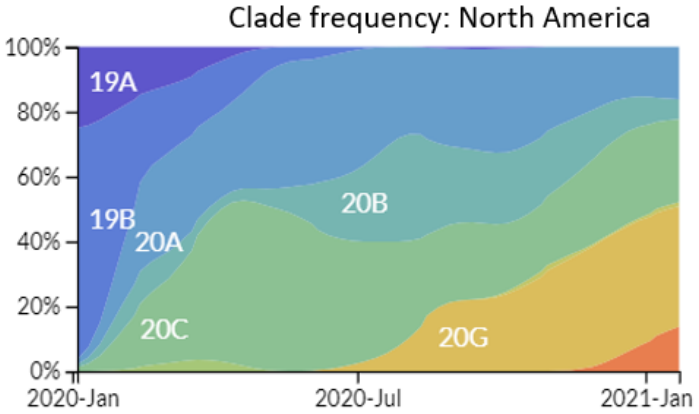


Adapted from Alm et al. 2020

Variant frequency

- All variants start rare – product of random mutation
- Some increase in abundance through transmission

1. Changing epidemiology:
   - Spread among specific population
   - Superspreading events
2. Altered viral property (‘phenotype’)
   - Increased transmissibility or infectivity
   - Immune escape

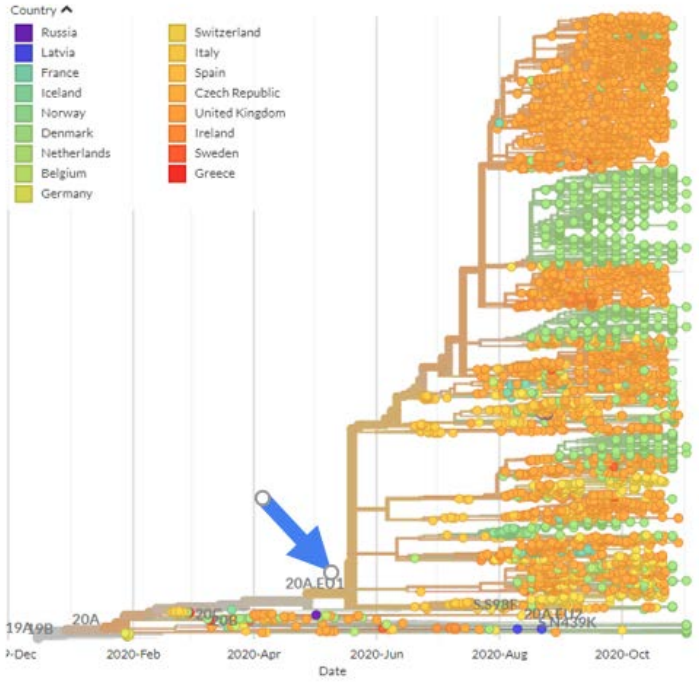


https://nextstrain.org/

High frequency due to epidemiology

- **<u>Founder Effect</u>**: *the loss of genetic variation that occurs when a new population is established by a very small number of individuals*

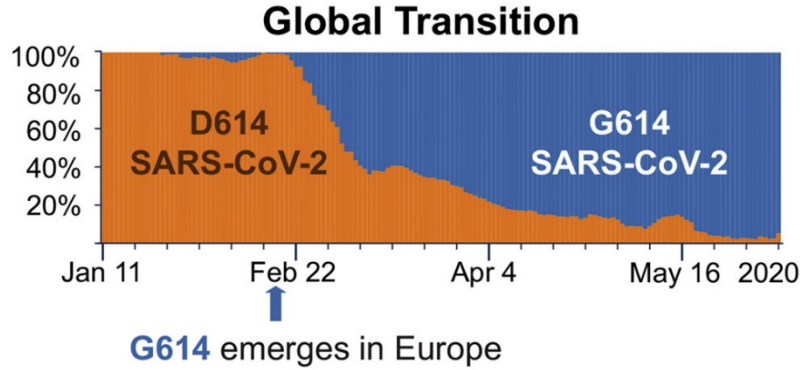
- Example: 20E (20A.EU1)
  - Spread across Europe fueled by summer vacation travel
  - Multiple introductions to countries
  - Predominant European variant in autumn
  - Hodcroft *et al.* MedRxiv. 2020

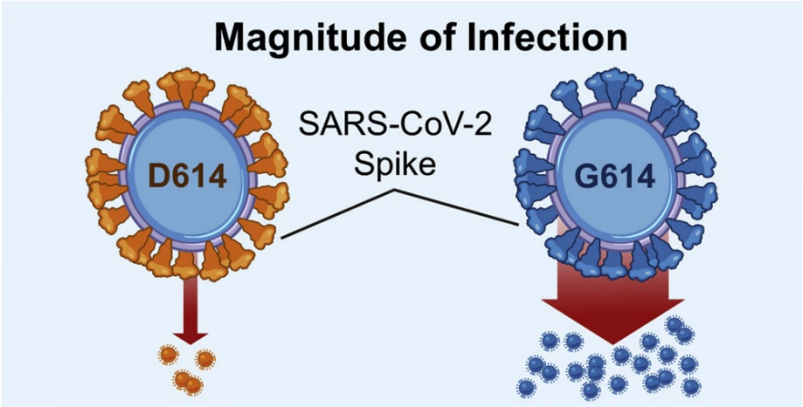


https://nextstrain.org/groups/neherlab/ncov/20A.EU1/2020-11-10?c=country&f_region=Europe

Spike D614G, the original variant

- Emerged early in pandemic, as SARS-CoV-2 spread across Europe and North America, become globally dominant
- Aspartic acid (D) -> Glycine (G)
- Suspected to alter vial properties
  - Increased viral load
  - Higher transmissibility?



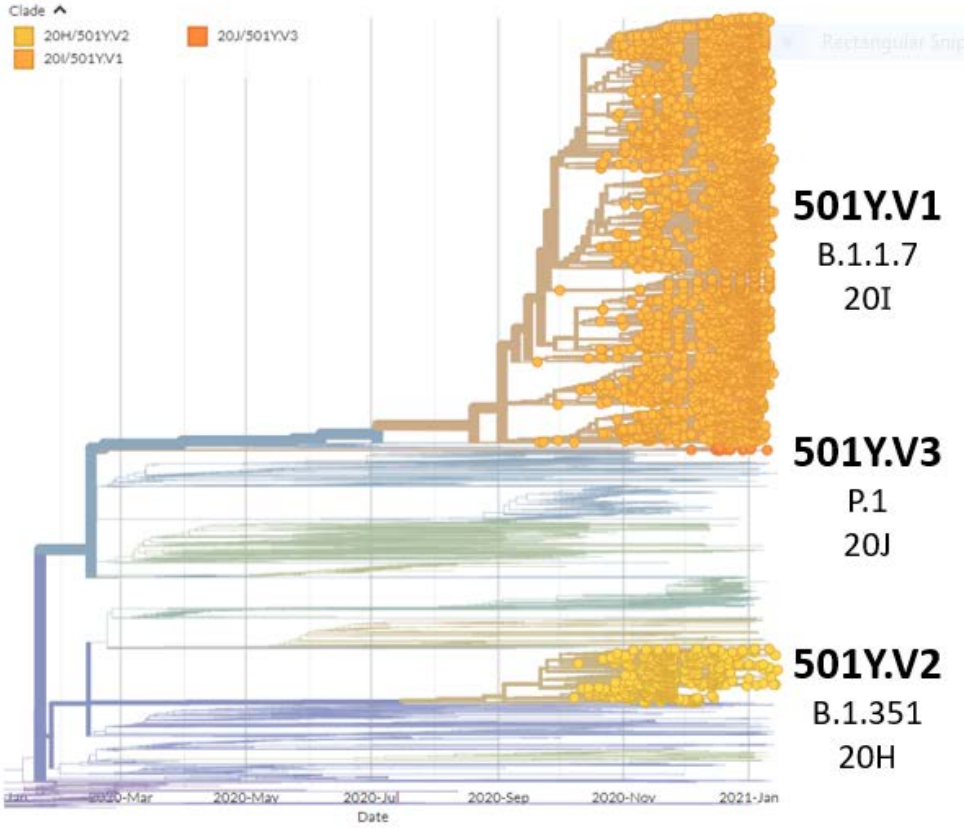


Korber *et al.* (2020) Cell, PMID: 32697968

Recurrent mutation

- **Homoplasy**: *repeat occurrence of mutation in unrelated branches of the phylogeny*
  - Convergent evolution

- Example:
  - Spike N501Y variants
  - Multiple clades/lineages



https://nextstrain.org/groups/neherlab/ncov/S.N501

What makes a “Variant of interest”

- When does an emerging variant become a “variant of interest” (VOI)?
  - or “variant under investigation” (VUI)
  - or “variant of concern” (VOC)

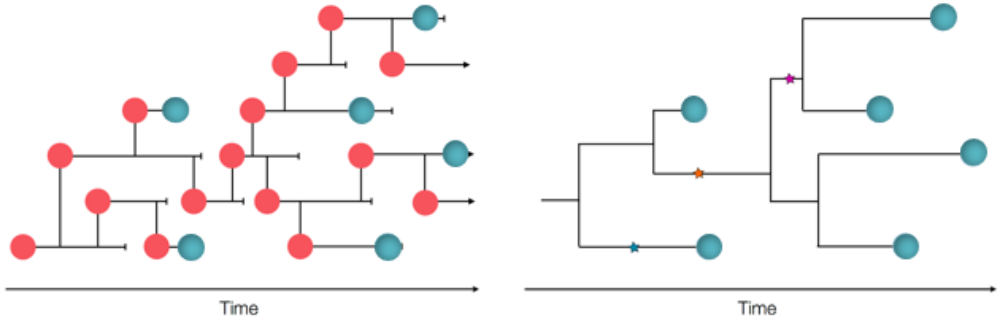
- Concerning epidemiological, immunological, or pathogenic characteristics

- Mutations with theoretical or empirical evidence of altered properties
  - Alterations to spike receptor binding domain (RBD)
  - Recurrent mutations in unrelated parts of the phylogeny

What does this mean for genomic epidemiology?

- Genetic fingerprinting to investigate transmission:
  - Probably very little; *“Phylogeny (still) approximates epidemiology”*

- BUT some mutations may alter viral properties, and thus epidemiology!
  - Variant identification can inform investigation and response



Images from Trevor Bedford Group: https://docs.nextstrain.org

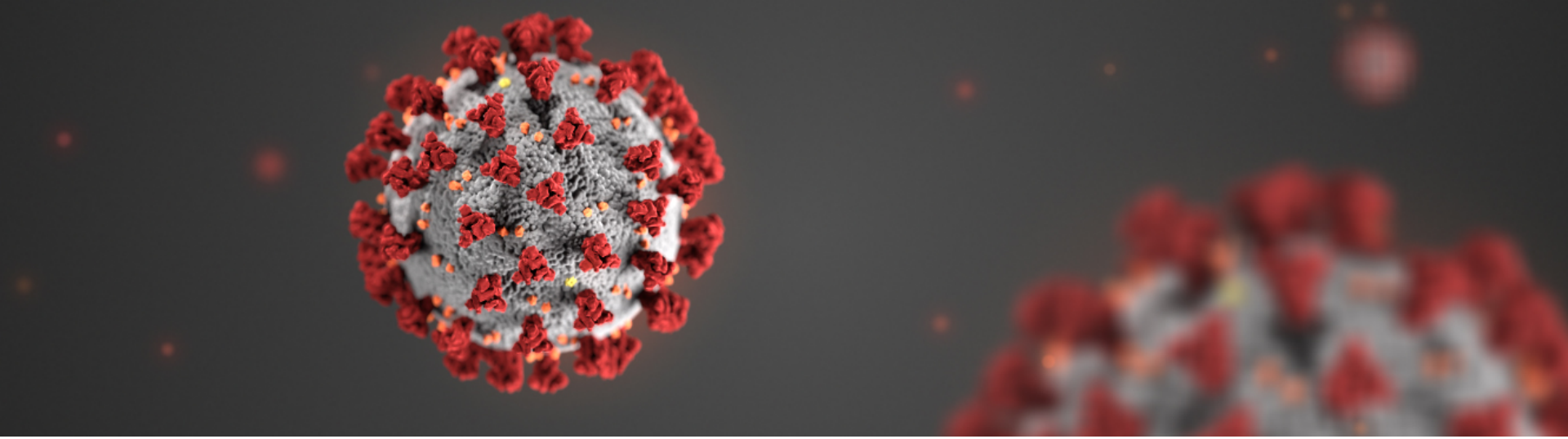
Summary

- Continued transmission, and thus viral replication, naturally leads to new variants
- Genomic surveillance is critical for detecting emerging variants that may:
  - Spread more quickly, or cause altered disease severity
  - Evade diagnostic detection, or vaccine-induced immunity
- This is a rapidly-changing area for investigation—and nomenclature for SARS-CoV-2 remains fluid and often confusing
- Quickly determine variant, clade names:
  - Pangolin - https://pangolin.cog-uk.io/
  - Nextclade - https://clades.nextstrain.org/

Learn more

- Other introduction modules
  - The SARS-CoV-2 genome – Module 1.2
  - How to read phylogenetic trees – Module 1.3

- COVID-19 Genomic Epidemiology Toolkit
  - Find further reading
  - Subscribe to receive updates on new modules as they are released go.usa.gov/xAbMw



For more information, contact CDC
1-800-CDC-INFO (232-4636)
TTY: 1-888-232-6348 www.cdc.gov

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.