



Data Seminar

03/05/2021



Outline

1. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK.
 - Motivation & Idea
 - Genomic sequence data
 - Structure of lineages
 - Dynamics of transmission lineages
 - Spatial distribution of the lineages
 - Dynamics of lineage importation
 - Summary
2. Whole-genome sequencing to track SARS-CoV-2 transmission in nosocomial outbreaks.
3. Genomic Diversity of SARS-CoV-2 During Early Introduction into the United States National Capital Region.



Outline

1. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK, Aanensen D M, Kraemer M U G, Rambaut A, et al. Science, 2021.
2. Whole-genome sequencing to track SARS-CoV-2 transmission in nosocomial outbreaks, Lucey M, Macori G, Mullane N, et al. Clinical Infectious Diseases: an Official Publication of the Infectious Diseases Society of America, 2020.
3. Genomic Diversity of SARS-CoV-2 During Early Introduction into the United States National Capital Region, Thielen P M, Wohl S, Mehoke T, et al. MedRxiv, 2020.



Motivation & Idea

Problem:

- How lineages vary in size, spatial distribution, and persistence?
- How key properties like epidemic size and duration influence each other?

Small-scale: Individual-level contact-tracing.

But large scale? - Pathogen genome sequencing.

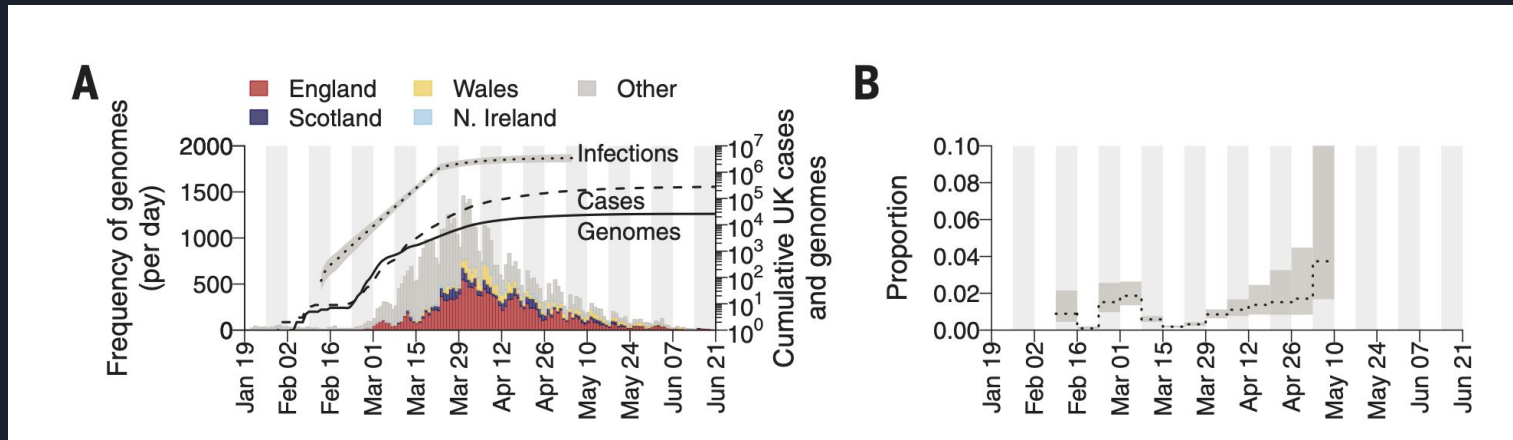
In this work, the author combines large genomic dataset with epidemiological and travel data to demonstrate the lineage dynamics of the UK COVID epidemic.

Genomic sequence data

50877 genomes analyzed before June 26, 2020. 26181 from UK.

Accounts for 9.29% of confirmed UK COVID cases.

0.66%(95% CI: 0.46% - 0.95%) of all inferred UK infections.



Structure of lineages

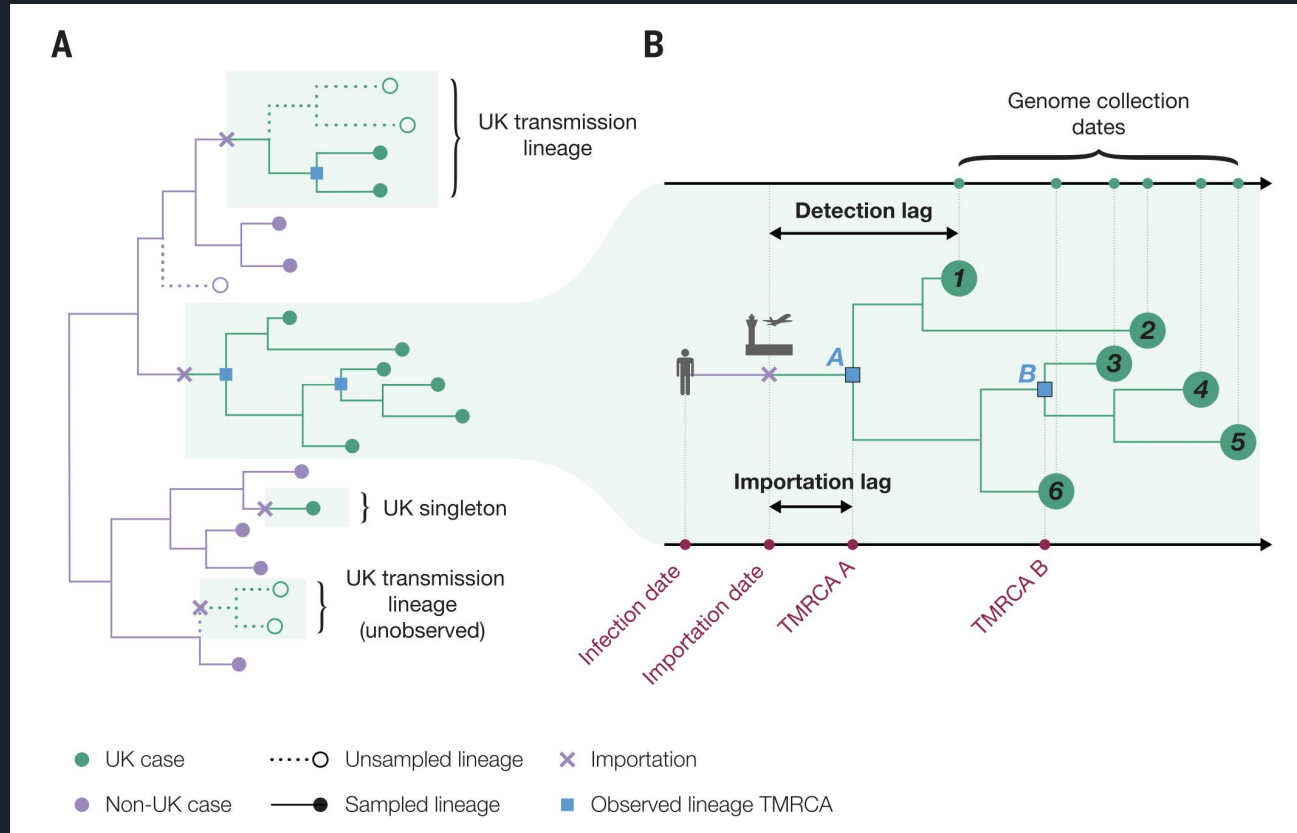
Lineage:

- (1) 2+ UK genomes.
- (2) Ancestral out of UK.

TMRCAs:

Time of the most recent common ancestor.

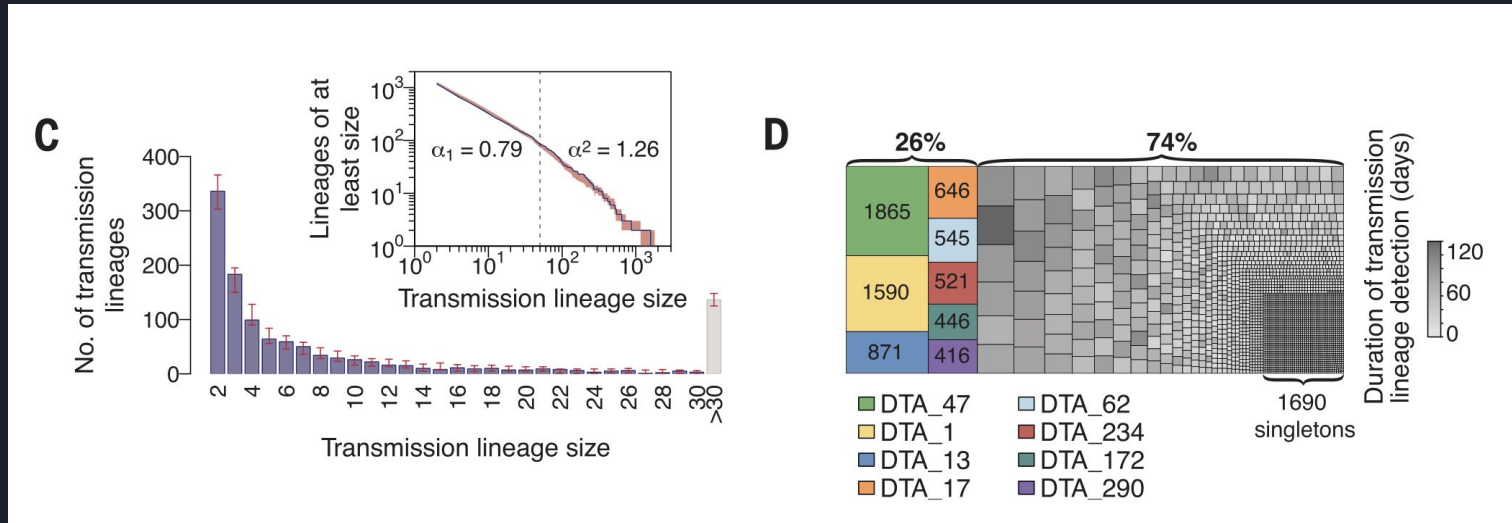
(Inferred transmission event)



Distribution of lineage sizes

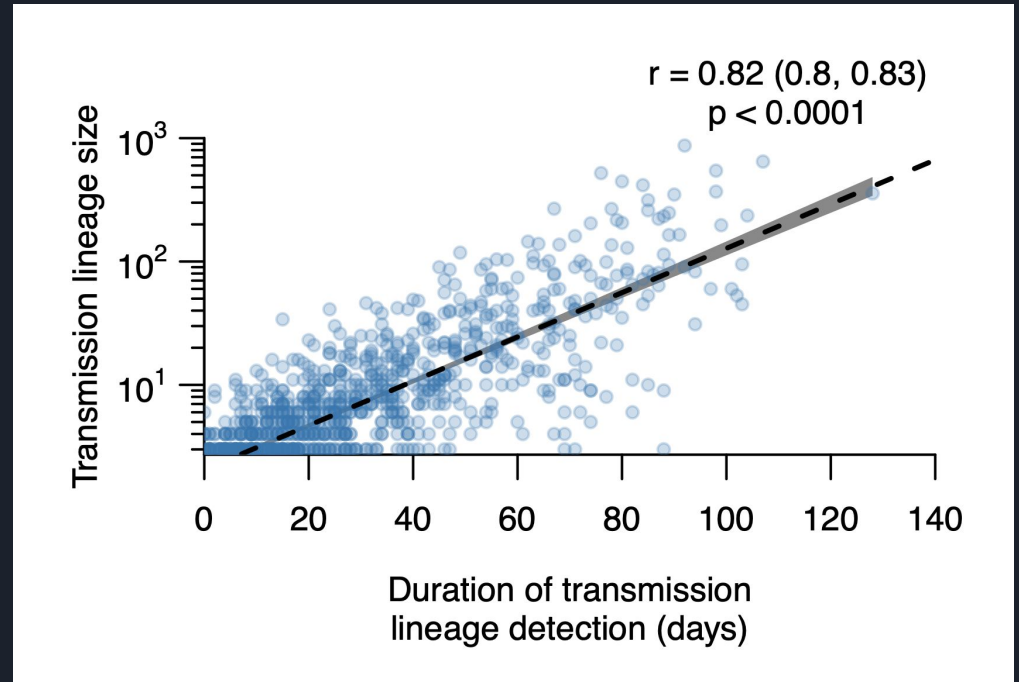
The distribution of lineage sizes follows the power-law distribution.

8 largest lineages account for 26% of all samples and were first detected before UK lockdown.



Distribution of lineage sizes

Larger lineages tend to be observed for a longer time.

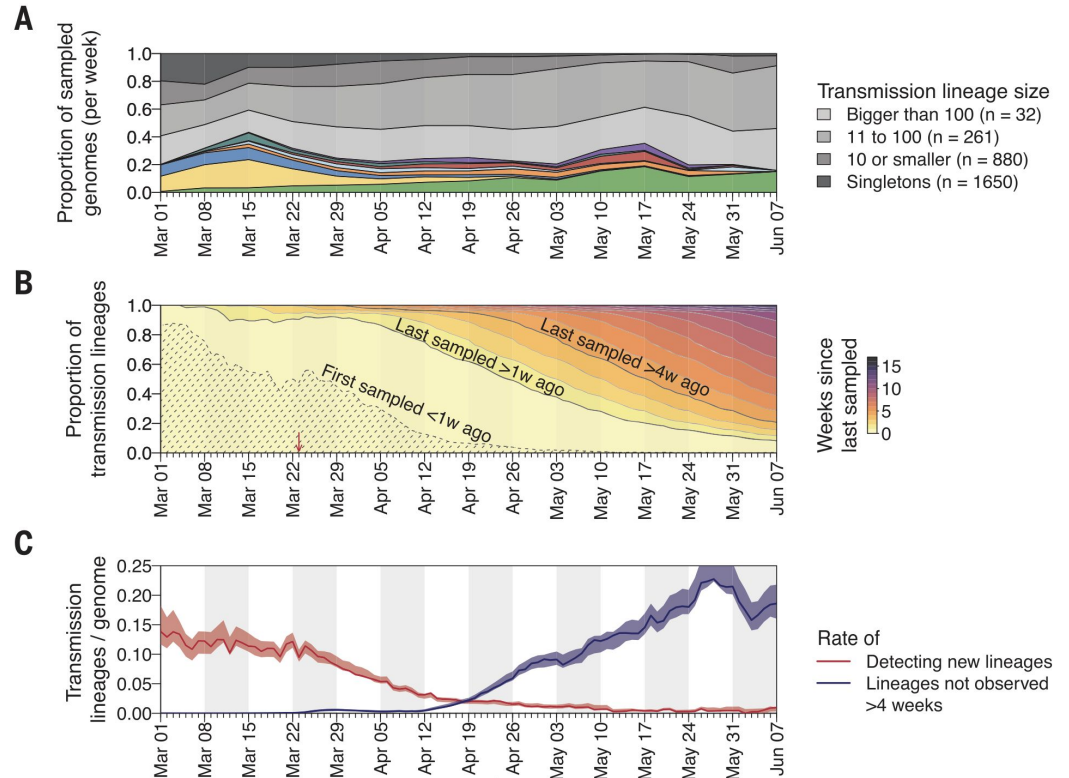


Dynamics of transmission lineages

Proportion of small lineages decreases

Most lineages has been rare/extinct.

- Less new lineages observed.
- More lineages have been not observed for 4+ weeks.

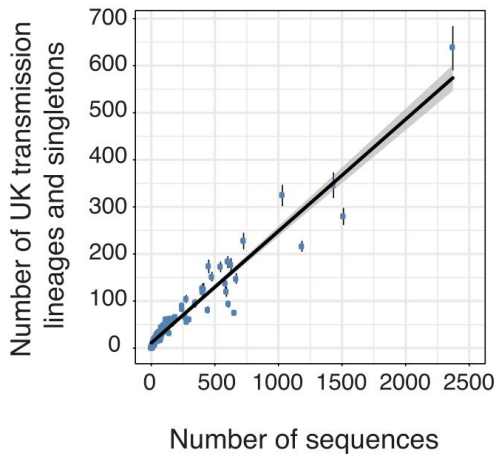


Spatial distribution of lineages

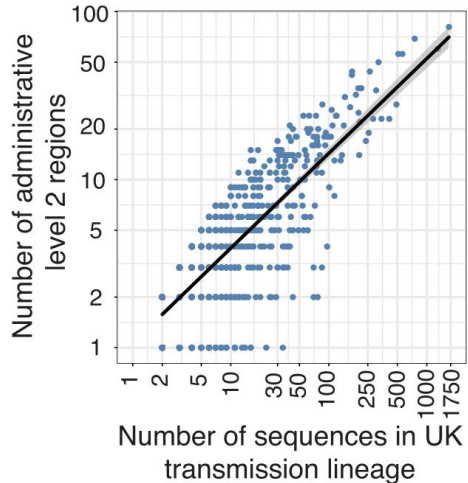
Bigger regional epidemics comprised a greater diversity of lineages.

Larger lineages tend to more geographically widespread.

A



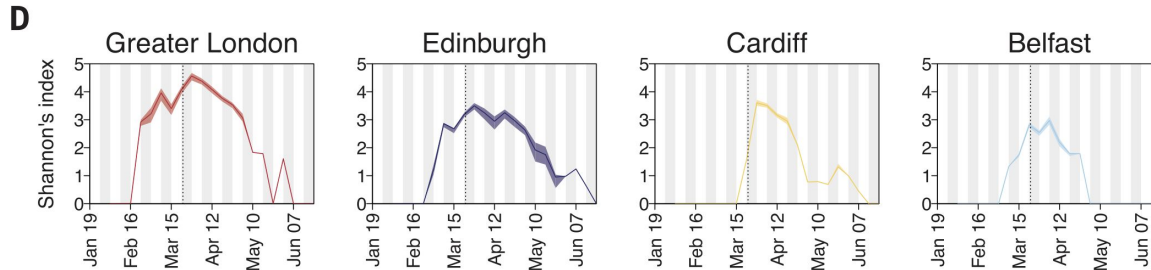
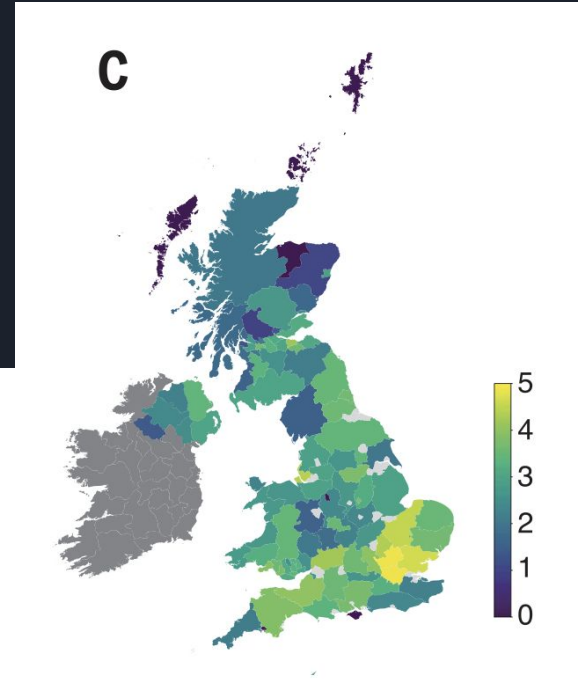
B



Spatial distribution of lineages

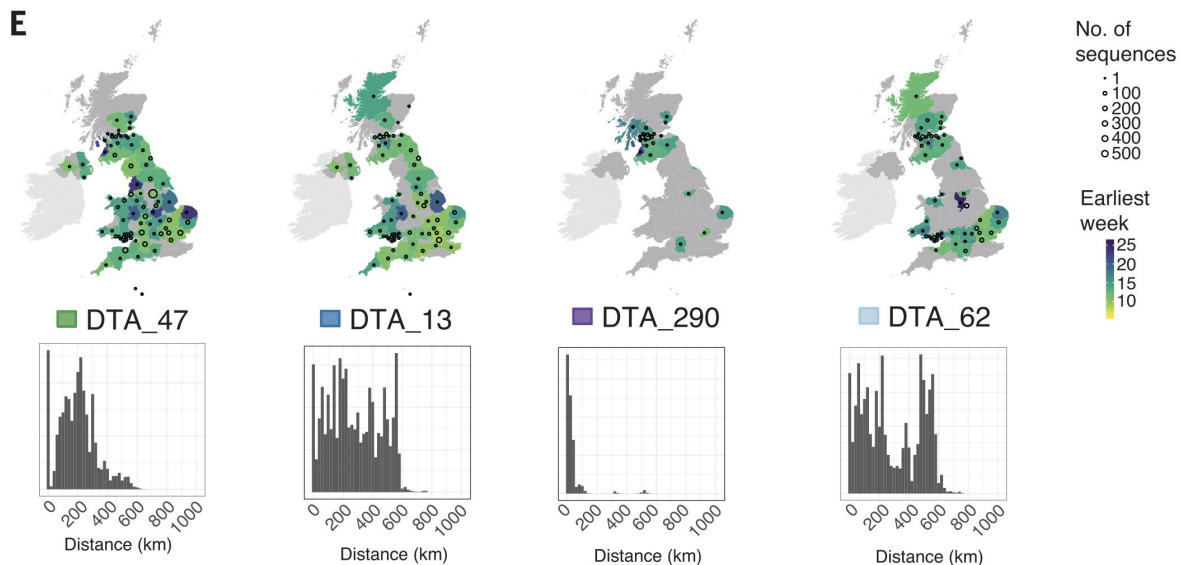
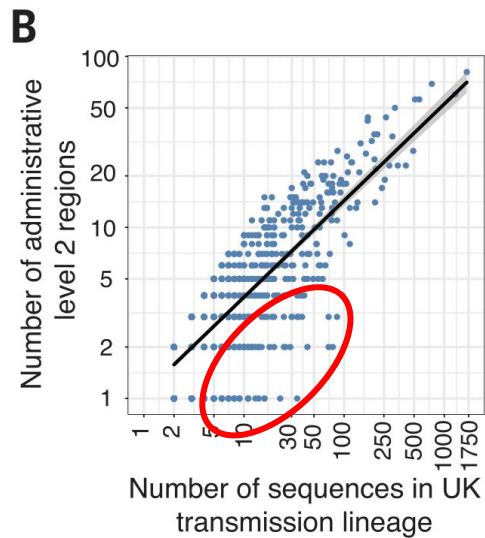
Shannon's index: When (1) the number of lineages and (2) the evenness of the lineage frequency increase, SI increases.

- Higher around London, lower in Scotland.
- SI before and after the lockdown.



Spatial distribution of lineages

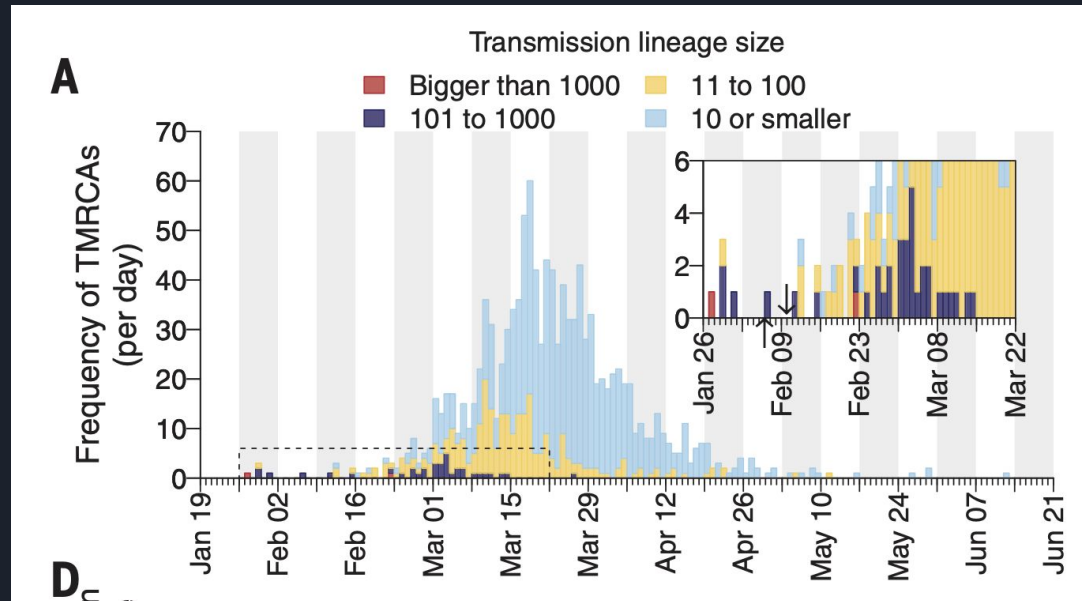
- Variation in the spatial range of lineages: Localized lineages.



Dynamics of lineage importation.

Inferred number of TMRCAs (Time of the most recent common ancestor)

- Most lineages are dated to March and early April.
- Lineages with earlier TMRCAs tend to be larger.

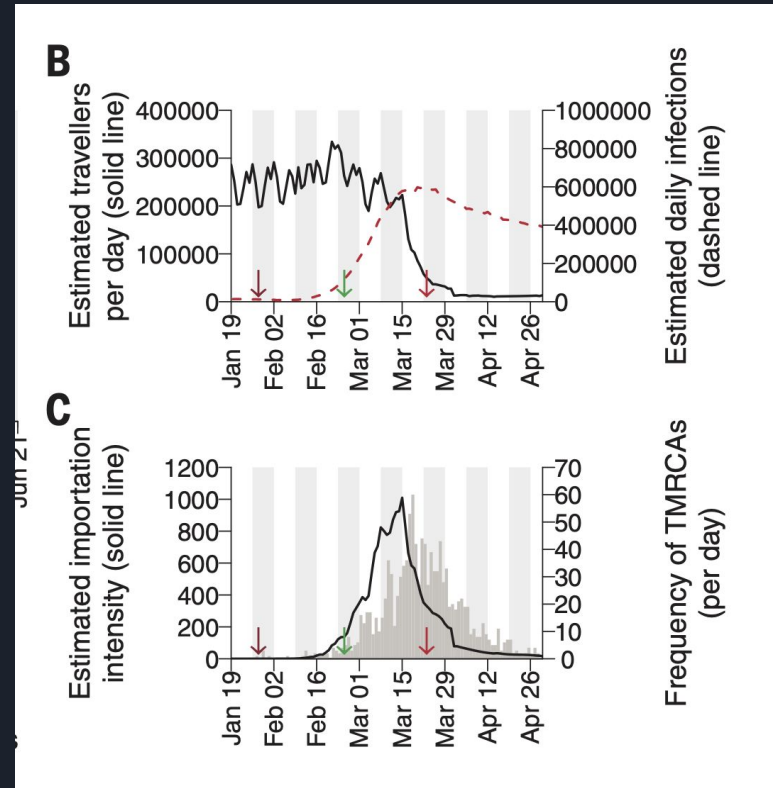


Dynamics of lineage importation.

TMRCA represents the date of the first inferred transmission of the lineage instead of importation date.

What's the delay between importation and transmission?

- B: Estimated importation intensity inferred from number of travelers and estimated daily infections all over the world.
- C. Same trends of EII and distribution of TMRCAs but with a delay.





Summary

With pathogen genome sequencing, we are able to draw the picture of the large, national-scale epidemic.

What picture:

- Before the lockdown, international importations contributes to the large number of lineages.
- After the lockdown, large lineages survive but small lineages extinct.

Drawbacks: Early lineages extinct without being recorded?

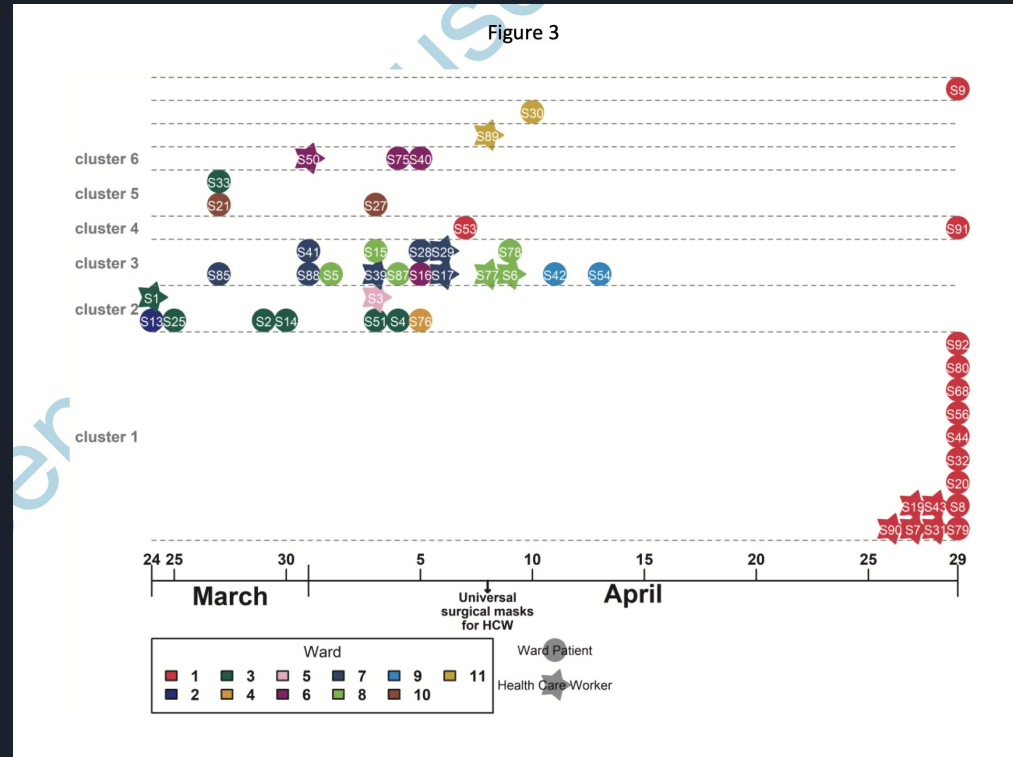
New ideas: What happens in the second-wave? : Local breakout? Or new transmission between different locations?

Whole-genome sequencing to track SARS-CoV-2 transmission in nosocomial outbreaks

Main idea: Use genome sequencing data to trace the nosocomial transmission from HCW to HA patients.

Input Data: (1) Genome sequencing data. (2) Contact info between HCW and patients.

Conclusion: Inferred HA transmission paths: Close contact with HCWs & HCWs intra-transmission because of not wearing masks.

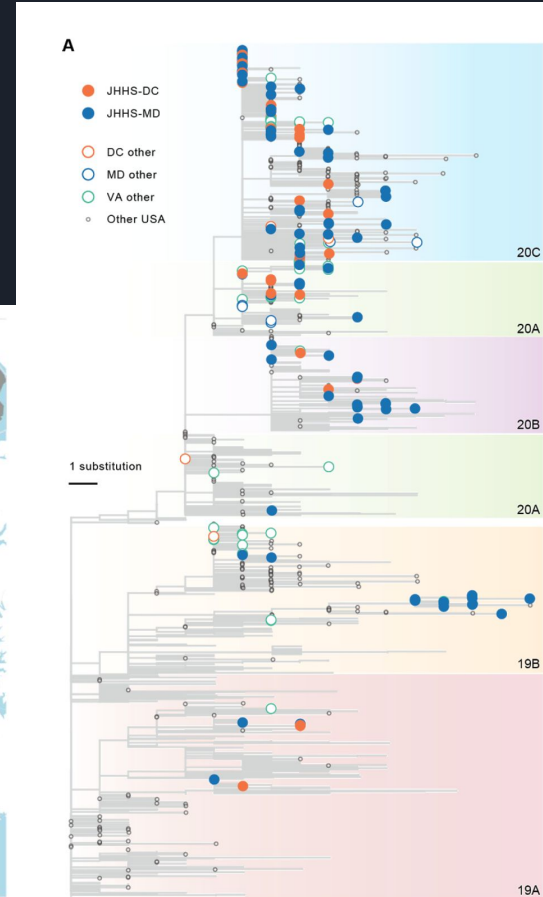
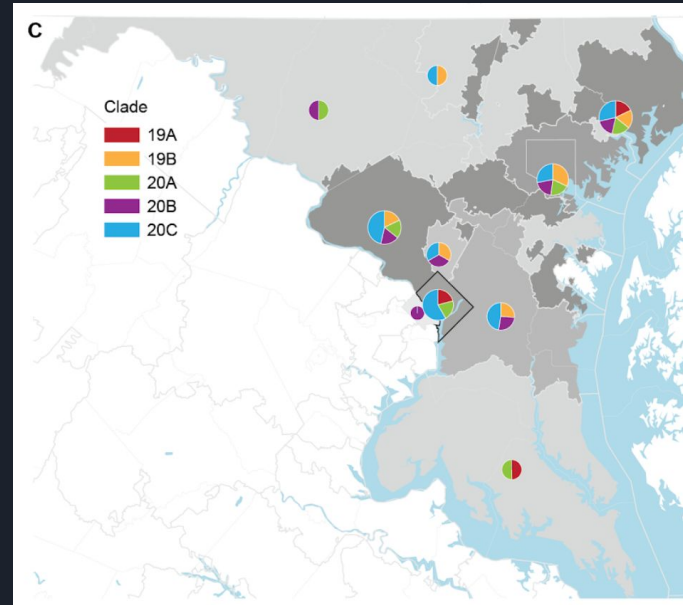


Genomic Diversity of SARS-CoV-2 During Early Introduction into the United States National Capital Region

Main idea: The diversity of genomics in early stages of the COVID.

Input data: Genomic sequences collected by JHHS (DC & MD)

Output: The sequences belong to all 5 major phylogenetic clades of the COVID:
High diversity at early stage!



Genomic Diversity of SARS-CoV-2 During Early Introduction into the United States National Capital Region

Different genomic diversity at different location: Help to tell which are more infected by importations, which are more infected by local transmissions.

