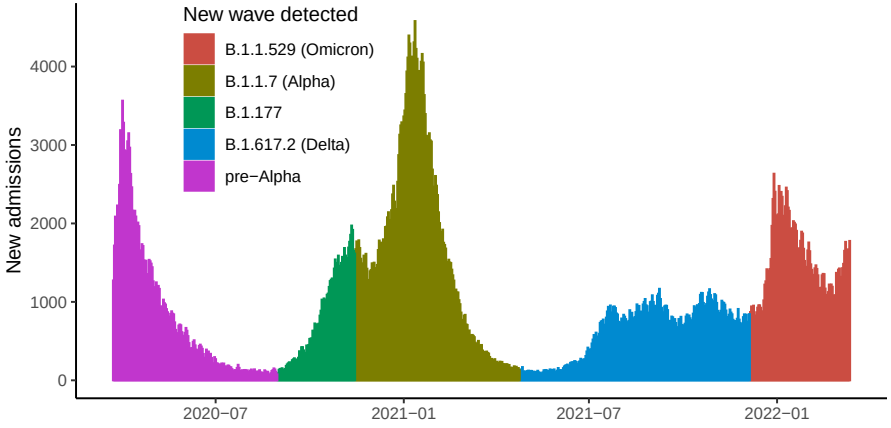
Methods for detecting and characterizing emerging SARS-CoV-2 variants

Erik Volz

12 June, 2023

covid19 hospital admissions in the United Kingdom

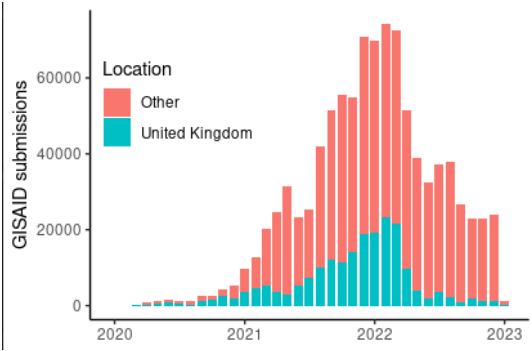


Outline

1. Genomic surveillance of SARS-CoV-2 variants
   - Review of variants that shaped the covid19 pandemic in the UK
2. Early warning signals provided by pathogen genomic surveillance
3. Future directions and open questions
   - Defining and detecting important lineages
   - Dealing with complex patterns of cocirculation
   - Design of Genomic Surivellance Systems (GSS)

COVID-19 Genomics UK Consortium (COG-UK)

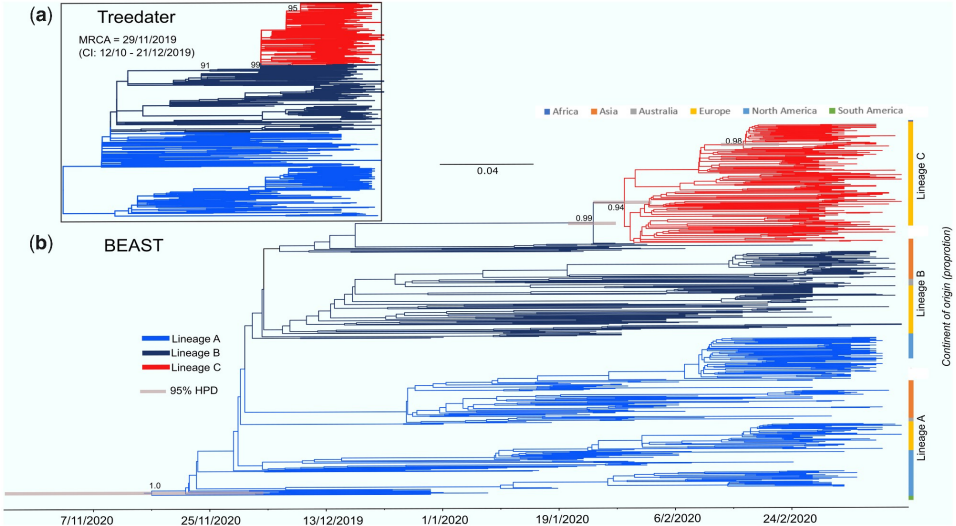
- Shared sequencing and bioinformatics resources
  - Clinical, academic and public health laboratories
  - Wellcome Sangar Institute (>90% of community-based samples)
- Standardised metadata
- Standardised phylogenetics pipeline



COG-UK

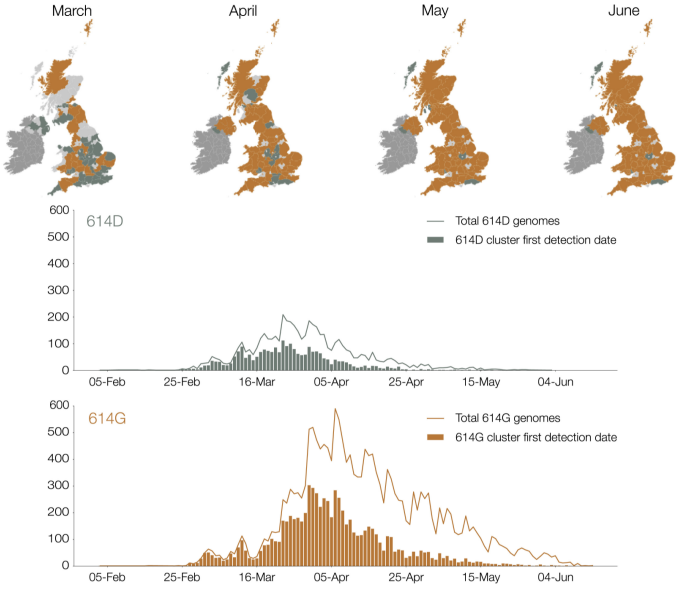
COVID-19
GENOMICS
UK CONSORTIUM

Emergence of D614G



Fountain-Jones et al, Virus Evolution, 2020

Expansion of the D614G variant Spring 2020



Is a variant more transmissible? If so, how much?

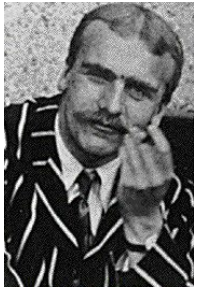
Classic population genetic model:

Suppose:

- Two alleles n and m with population sizes N_t and M_t.
- In each generation n makes R copies and m makes $R_m = (1+s)R$ copies.

$$N_{t+1} = N_t R_n$$

$$M_{t+1} = M_t(1+s)R_n$$



JBS Haldane

Proportion that is more transmissible over time?

$$p_t = \frac{M_t}{M_t + N_t}.$$

$$OR_t = \frac{p_t}{1 - p_t} = \frac{M_t}{N_t}.$$

How does the ratio change over time?

$$OR_{t+1} = \frac{(1+s)RM_t}{RN_t}$$

$$= (1+s)OR_t$$

What about in an SIR epidemic?

SIR model with two variants.

- Variants can differ by transmission rate β and/or generation time $1/\gamma$

$$\frac{d}{dt}I(t) = \beta I(t)\frac{Y(t)}{N} - \gamma I(t)$$
$$\frac{d}{dt}I'(t) = \beta' I'(t)\frac{Y(t)}{N} - \gamma' I(t)$$
$$\frac{d}{dt}Y(t) = -\frac{Y(t)}{N}(\beta I(t) + \beta' I'(t))$$

Evolution of log odds:

$$\psi(t) = log\left(\frac{I'(t)}{I(t)}\right)$$

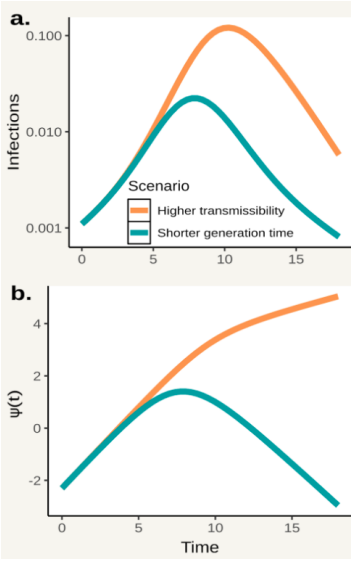
$$\frac{d}{dt}\psi(t) = \frac{Y}{N}(\beta' - \beta) + (\gamma - \gamma')$$

Scenario $\beta' = (1+s)\beta$ and $\gamma' = \gamma$:

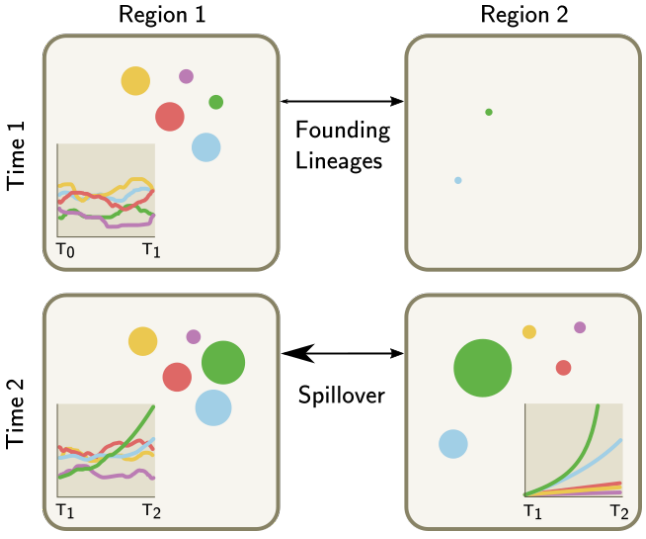
$$\left(\frac{d}{dt}\psi\right) \times T_g = \frac{Y(t)}{N\gamma}(\beta' - \beta) = sR_t$$

Scenario $\gamma' = (1+r)\gamma$ and $R' = R$:

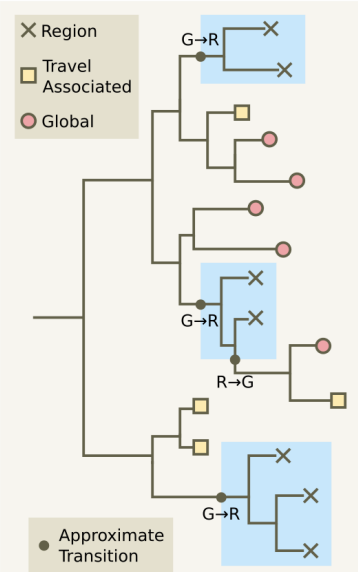
$$\frac{d}{dt}\psi \times T_g = \left(R_t - 1\right)\left(\frac{1}{T_g'} - \frac{1}{T_g}\right) = r\left(R_t - 1\right)$$



Founder effects & spillover

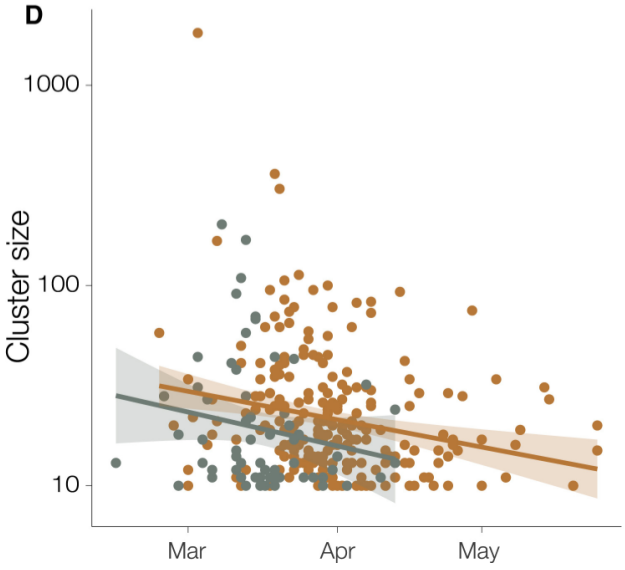


Adjusting for founder effects

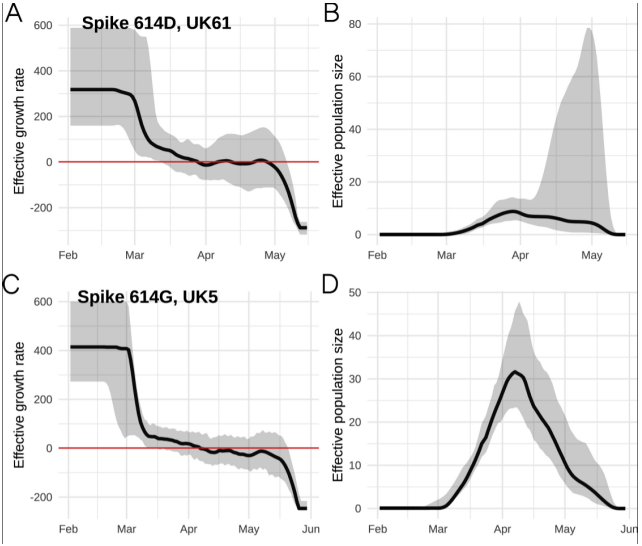


- Combined global sequences and sequences from travelers
- Ancestral state estimation enables identification of imported lineages
- & Time of importation

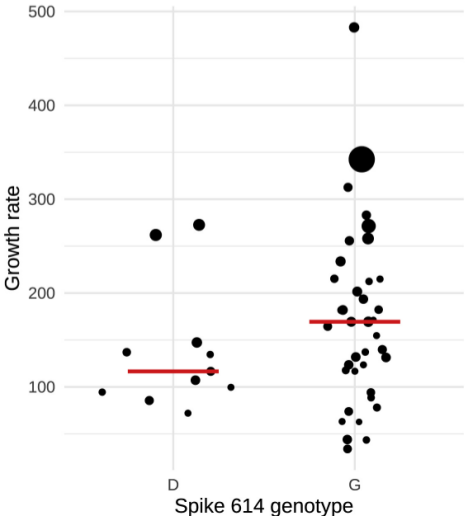
Phylogenetic cluster sizes D614G

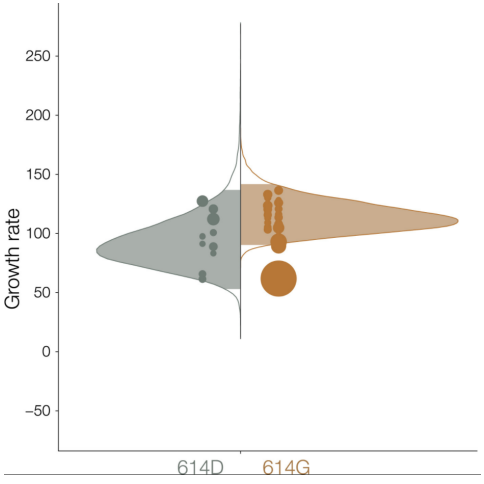


Effective population size



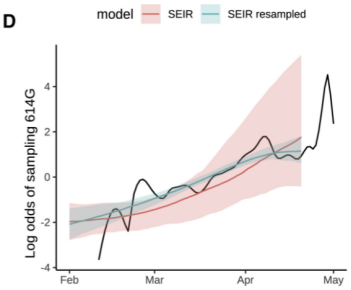
Effective growth rates





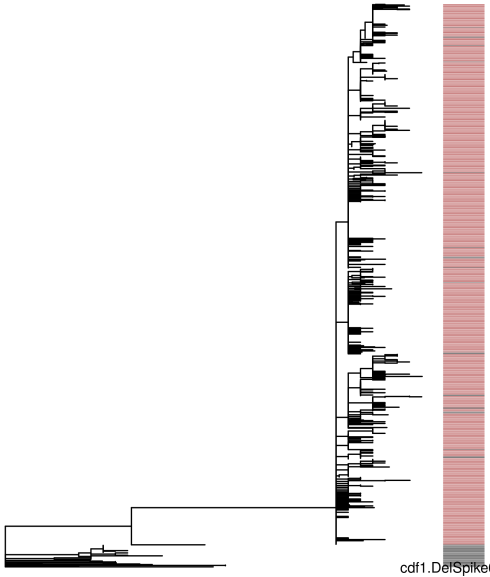
D614G conclusions

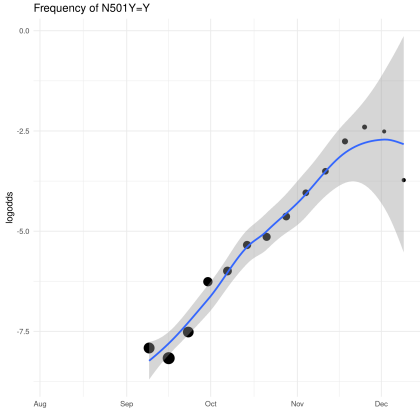
- Phylodynamic models were inconclusive but pointed towards a difference of about 20-30% transmissibility
- Simple comparisons of logistic growth rate were statistically significant and pointed towards an effect size comparable to phylodynamic models
- Using both logistic growth and phylodynamic (coalescent) modeling gave higher precision for size of the effect
- Subsequent laboratory and animal studies supported a large change in phenotype

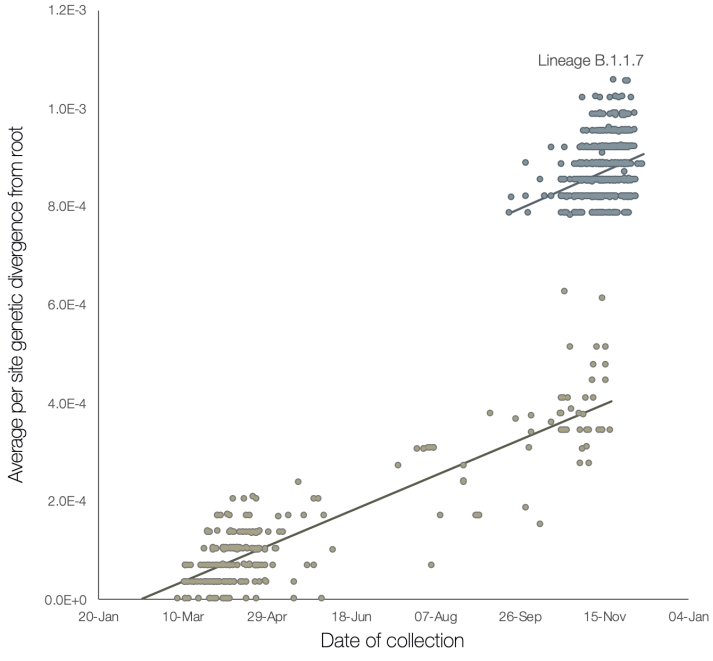


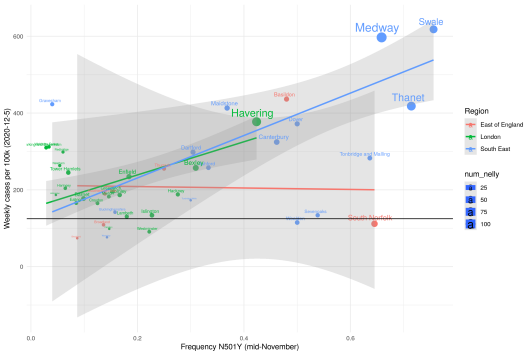
Emergence of B.1.1.7 (Alpha)

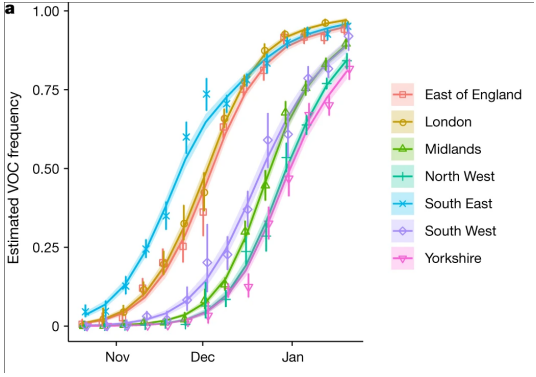
Lineage B.1.1.7






Lineage B.1.1.7
Average per site genetic divergence from root
1.2E-3
1.0E-3
8.0E-4
6.0E-4
4.0E-4
2.0E-4
0.0E+0
20-Jan
10-Mar
29-Apr
18-Jun
07-Aug
26-Sep
15-Nov
04-Jan
Date of collection


Medway
Swale
Thanet
Havering
Basildon
Canterbury
Tonbridge and Malling
South Norfolk
Region
East of England
London
South East
num_nelly
25
50
75
100
Weekly cases per 100k (2020-12-5)
Frequency N501Y (mid-November)
0
200
400
600
0.0
0.2
0.4
0.6

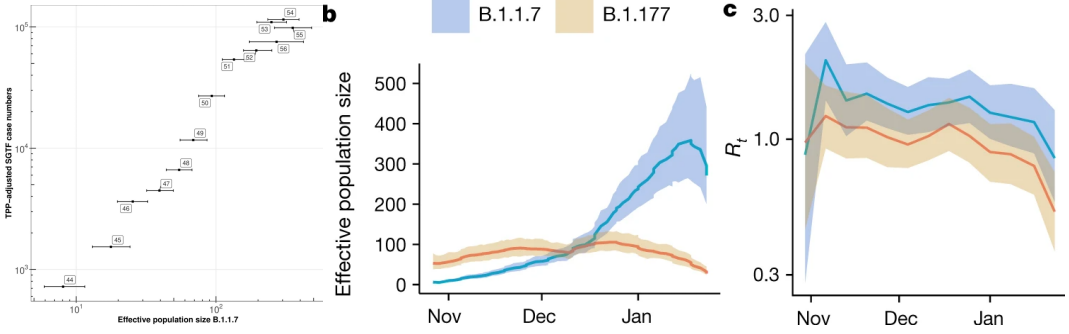

a
Estimated VOC frequency
1.00
0.75
0.50
0.25
0
Nov
Dec
Jan
East of England
London
Midlands
North West
South East
South West
Yorkshire

December 19, 2020

"NERVTAG's early analysis suggests the new variant could increase R by 0.4 or greater. Although there is considerable uncertainty, it may be up to 70% more transmissible than the old variant."

Phylodynamic case-control

For each B.1.1.7 patient, sample without replacement a non-B.1.1.7 patient in the same 1) region 2) week of sample collection.



$$\frac{R'_t}{R_t} = \frac{\int_{a=0}^{\infty} g(a) e^{-\int_{b=0}^{a} r(t-b)db} da}{\int_{a=0}^{\infty} g(a) e^{-\int_{b=0}^{a} r(t-b)+\rho(t-b)db} da}$$

B.1.617.2 (Delta)

Establishment of the UKHSA Variant Technical Group

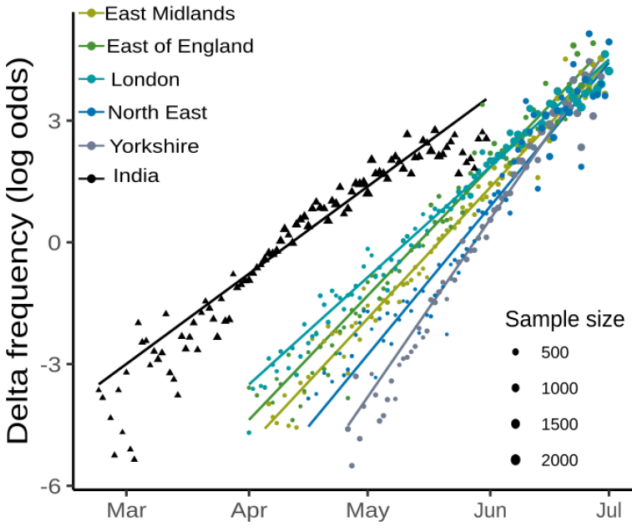
3 June 2021 Risk assessment for SARS-CoV-2 variant: Delta (VOC-21APR-02, B.1.617.2) **Public Health England**

Indicator	RAG*	Confidence	Assessment and rationale
Transmissibility between humans		**HIGH**	**Transmissibility appears greater than wild type (first wave) SARS-CoV-2** Delta continues to demonstrate a substantially increased growth rate compared to Alpha, across multiple analyses. Delta cases are rising whilst Alpha cases are declining. Secondary attack rates, including household secondary attack rates, are higher for Delta, but these are not yet corrected for vaccination status. There is in vitro evidence suggestive of increased replication in biological systems that model human airway. It is highly likely that Delta is significantly more transmissible than Alpha.
Infection severity		**LOW**	**Increased severity (hospitalisation risk) when compared to Alpha** Early evidence from England and Scotland suggests there may be an increased risk of hospitalisation compared to contemporaneous Alpha cases. A large number of cases are still within the follow up period. In some areas, hospital admissions show early signs of increasing, but the national trend is not clear.
Immunity after natural infection		**LOW**	**Experimental evidence of functional evasion of natural immunity but insufficient epidemiological data** Pseudovirus and live virus neutralisation using convalescent sera from first wave and Alpha infections shows a reduction in neutralisation. National reinfection surveillance data are being analysed. There is no increase in numbers of reinfections in the SIREN national healthcare worker cohort.
Vaccines		**HIGH**	**Epidemiological and laboratory evidence of reduced vaccine effectiveness** There are now analyses from England and Scotland supporting a reduction in vaccine effectiveness for Delta compared to Alpha. This is more pronounced after one dose (absolute reduction in vaccine effectiveness against symptomatic infection of approximately 15-20% after 1 dose). Iterated analysis continues to show vaccine effectiveness against Delta is higher after 2 doses but that there is a reduction for Delta compared to Alpha. There is a high level of uncertainty around the magnitude of the change in vaccine effectiveness after 2 doses of Oxford-AstraZeneca vaccine. Although this is observational data subject to some biases, it holds true across several analytic approaches and the same effect is seen in both English and Scottish data. It is strongly supported by pseudovirus and live virus neutralisation data from multiple laboratories. There are no data on whether prevention of transmission is affected and insufficient data to assess vaccine effectiveness against severe disease. The acquisition of an additional mutation which may be antigenically significant in a small number of cases is noted.
Overall assessment			Delta is predominant and all analyses find that it has a very substantial growth advantage. The observed high growth rate is most likely to be due to a combination of place based context, transmissibility and immune escape. Both English and Scottish analyses continue to support the finding of reduced vaccine effectiveness which has increased to high confidence. New early data from England and Scotland suggest a possible increased risk of hospitalisation compared to Alpha. The priority investigations are vaccine effectiveness against hospitalisation and transmission, household secondary attack rate corrected for vaccination, characterisation of the generation time, viral load and period of infectivity, and epidemiological studies of reinfections.

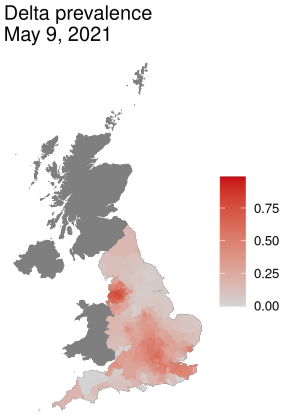
The therapeutics risk assessment is under review for all variants and is not included.
*refer to scale and confidence grading slide

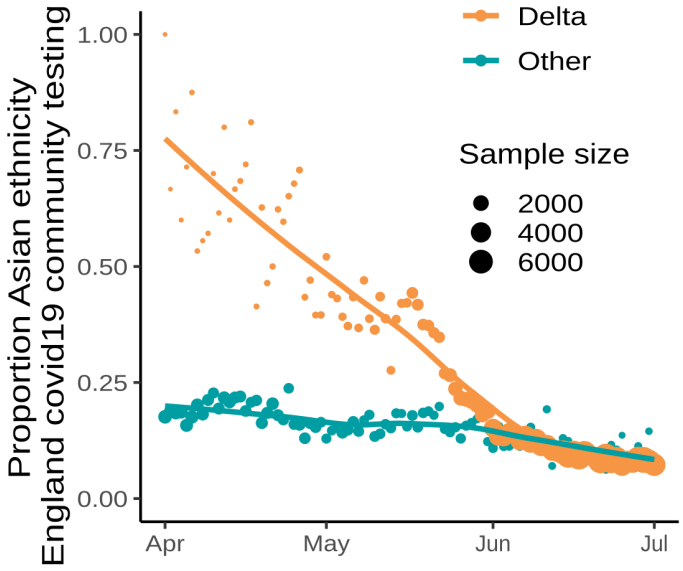
https://www.gov.uk/government/publications/investigation-of-sars-cov-2-variants-technical-briefings

Rapid growth & displacement of B.1.1.7

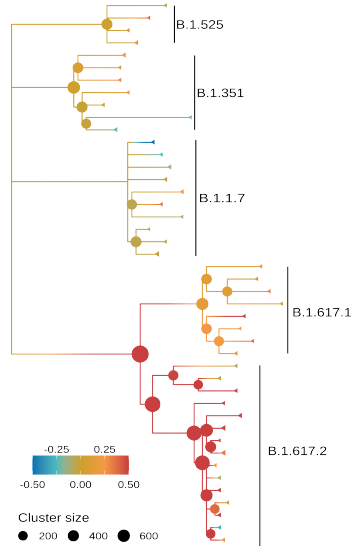


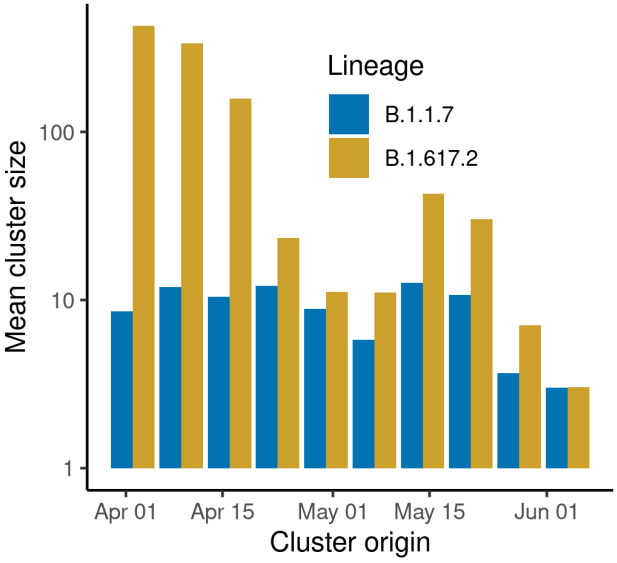
Difficult to evaluate



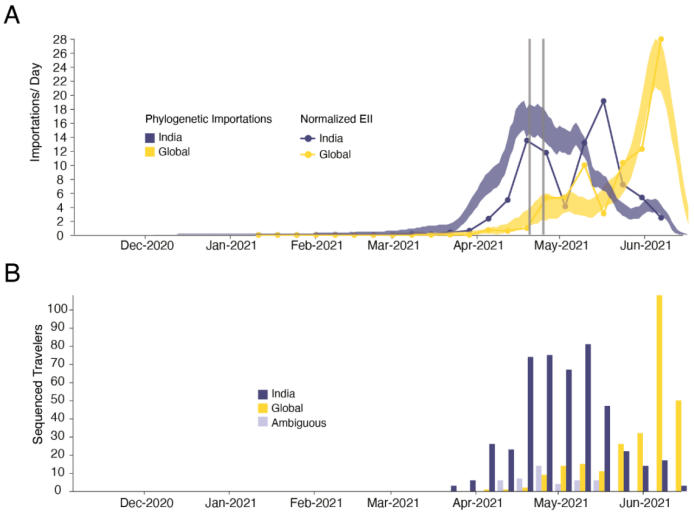


Phylogenetic evidence





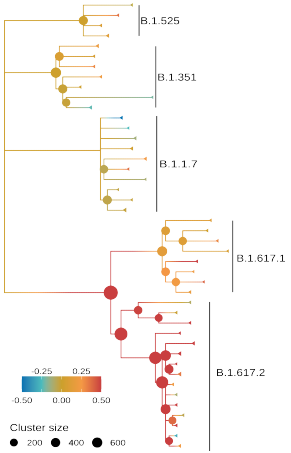
Travel restrictions



JT McCrone et al., Science 2022

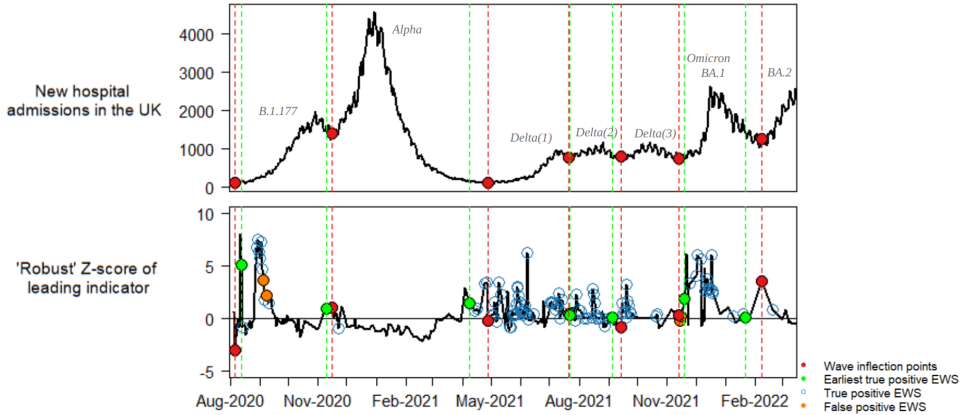
Early warning signals (EWS)

Designing EWS from genomic data sources



- Recall the figure at the beginning– most epidemic waves are associated with particular variants
- raises possibility that fast detection of novel variants can be an ews for rising cases
- TFP Scanner: EWS based on outlier growth among phylogenetic clades
- Statistics based on
  - logistic growth models (GAM), propensity-score matched based on geography
  - adjusts for geography
  - molecular clock, changes in evolutionary rates
- Consistent growth across related lineages & geography provides confidence that EWS is detecting a true change in transmissibility

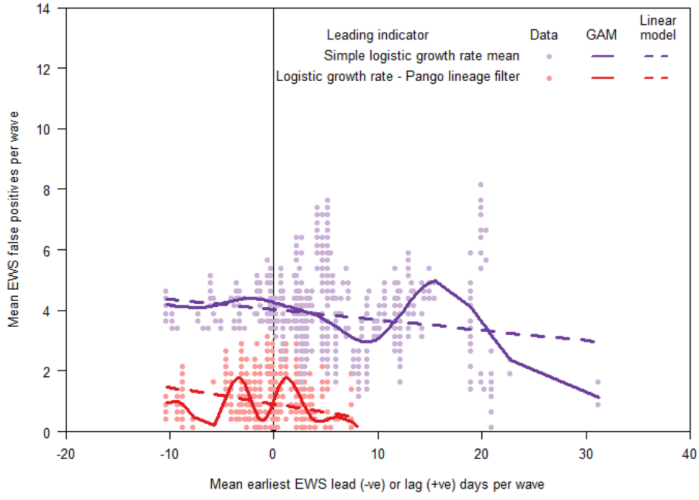
Phylogenomic scanning provided consistent EWS for epidemic waves in the UK



- Kieran Drake, MRC GIDA

Phylogenomic scanning produces few false positive signals

. . . Even when scanning with maximum sensitivity



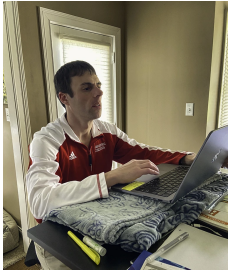
Future directions

1. Genomic surveillance - identifying important lineages
   - How do we define important lineages for investigation?
2. Quickly & robustly estimate fitness effects for lineages with partial co-circulation
   - How do we predict how a lineage will perform in a region where it is not yet circulating?
3. Sample design for genomic surveillance
   - Optimal size, density, and data sources

Detecting and defining important lineages

- How do we define important lineages for investigation?
- PANGO lineages cov-lineages.org
  - Designates lineages based on
    1. Robust identification of a new mutation on the background of an existing lineage
    2. Circulation of lineage within a new geographical space (typically country)
  - Reliant on large network of human volunteers
- Potential problems with PANGO
  - High noise-to-signal. Most PANGO lineages are unimportant.
  - Dependent on enthusiasm of humans
  - Can also be slow to detect important lineages

Ryan Hisner



https://www.nature.com/articles/d41586-023-01047-4

Systematic Biology

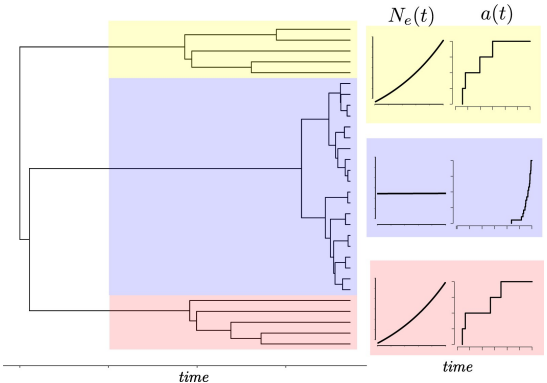
Identification of Hidden Population Structure in Time-Scaled Phylogenies

Erik M Volz, Wiuf Carsten, Yonatan H Grad, Simon D W Frost, Ann M Dennis, Xavier Didelot

Bayesian Inference of Clonal Expansions in a Dated Phylogeny

David Helekal, Alice Ledda, Erik Volz, David Wyllie, Xavier Didelot

- Detecting important phylogenetic structure can be automated
- Lineages which are growing quickly leave detectable signatures

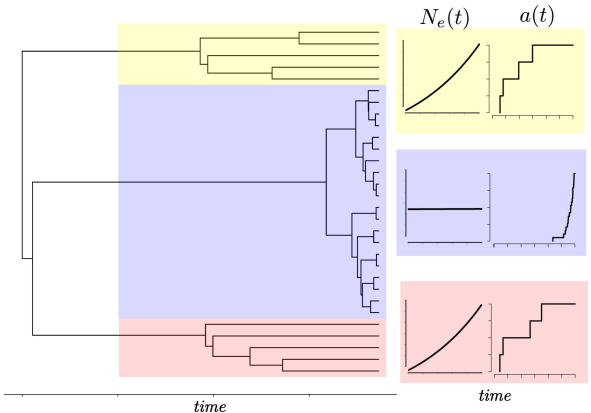


- Non-parametric method based on Kingman coalescent null hypothesis
- Test-statistic (rank sum) is derived based on ordering of coalescent events in adjacent clades.

$$\rho(X|Y) = \sum_{i=1}^{K} i\, \mathbf{1}_{D_X}(w_i)$$

- Transition probabilities

$$(z, w) \mapsto (z-1, w) \quad \text{with probability} \quad \frac{z+1}{z+w}$$



- Fast. Processes SARS-CoV-2 phylogeny with 10^5 tips in < 1 hour.
- Will also detect structure associated with geography-associated differences in transmission and/or sampling patterns

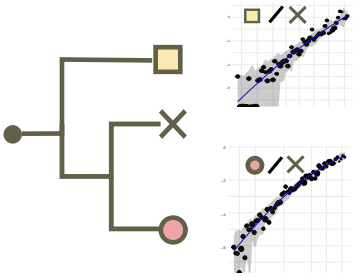
Current circulating lineages May 2023

Pipeline

- Usher → Chronumental
  - https://taxonium.org/
- → `treestructure`
  - "Phylotypes"
- → Logistic and coalescent growth rate for co-circulating phylotypes
  - How do we estimate relative fitness for phylotypes which are not yet co-circulating?

Estimating fitness for non-cocirculating variants

Relative fitness of variants which are not co-circulating?

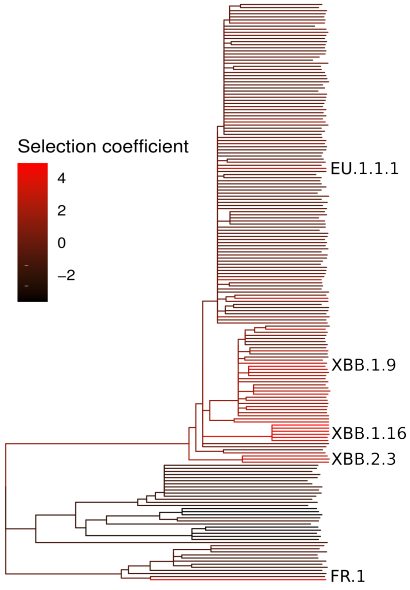


- Important question that has implications for forecasting
  - e.g. relative fitness of BQ & XBB lineages
- Fitness as a heritable trait
- Hierarchical Bayesian regression

$$r_{uv} \sim s_u - s_v$$

$$s \sim MVN(0, \sigma^2/D)$$

- Approach will fail where variant fitness depends on host-population properties
  - Specific immune landscape varies depending on vaccination and previous exposures

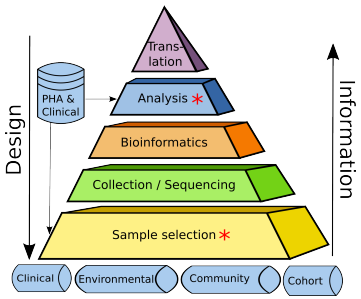


selcoef	pango_lineage_usher	country
4.91	FL.2	China
4.89	FU.1	China
4.40	XBB.1.16	England
4.26	XBB.1.16	England
4.25	FR.1	China
4.09	XBB.1.16	England
3.95	EG.5.1	China
3.86	XBB.1.16	China
3.77	XBB.1.16	England
3.57	FL.2	China
3.52	EU.1.1.1	England

- XBB sub-variants dominating
- XBB.1.16 growing quickly in the UK
  - Four distinct phylotypes which PANGO currently does not distinguish between
- One representative of BA.2.75 sub-lineage still showing fitness advantage
- Geographic distribution reflects which countries are submitting sequences in a timely way, not representative

Design of genomic surveillance systems (GSS)

Sequence first and ask questions later

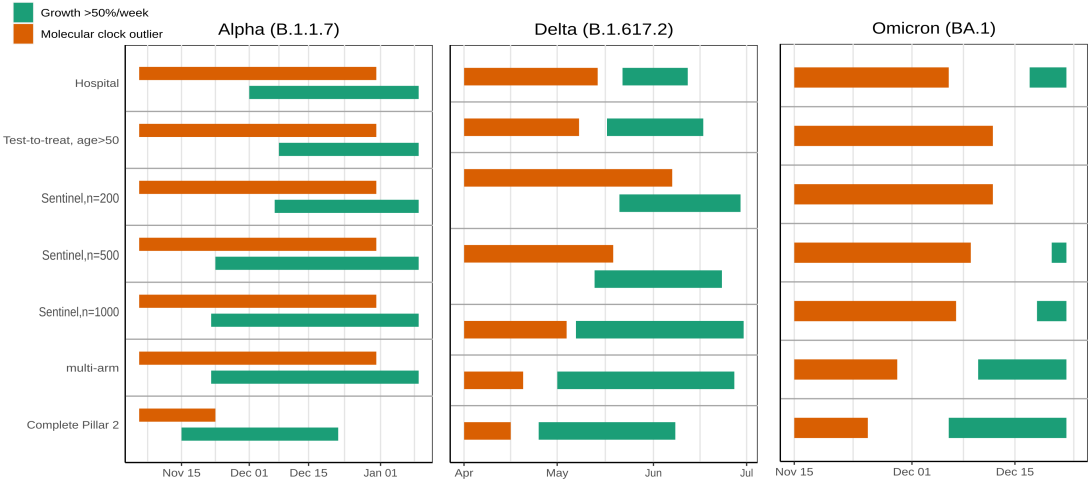


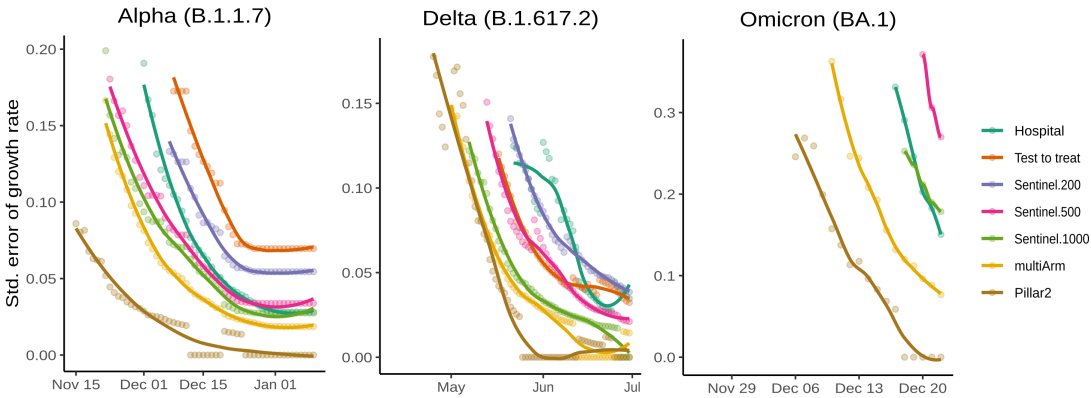
- Pathogen GSS sampling is often haphazard, convenience-based
- The UK achieved relatively standardised & uniform sampling for SARS-CoV-2 in community settings
  - Clinical sampling still often convenience-based
- Loss of community sampling (April 2022) is detrimental to future surveillance
- Future GSS should be informed by analytical goals (top-down)
  - Did the UK sequence too much?

Did the UK sequence too much? How important is community sampling?

- Retrospective simulation of how quickly Alpha, Delta & Omicron would be detected if:
  - Sampling from clinical settings only
  - Sampling only older individuals (>50)
  - Sentinal surveillance (n=200,500,1000/week) from general practictioners
  - Complete community sampling (no clinical)
- Outlier clades detected by logistic growth rate & molecular clock statistic

Timeline of TFP Scanner EWS for down-sampled UK data




Alpha (B.1.1.7)
Delta (B.1.617.2)
Omicron (BA.1)
Std. error of growth rate
0.20
0.15
0.10
0.05
0.00
Nov 15
Dec 01
Dec 15
Jan 01
May
Jun
Jul
0.3
0.2
0.1
0.0
Nov 29
Dec 06
Dec 13
Dec 20
Hospital
Test to treat
Sentinel.200
Sentinel.500
Sentinel.1000
multiArm
Pillar2

GSS design conclusions

- Sampling younger demographics is key to early detection and assessment
- Adaptive sampling strategy may be cost-effective
  - Example: Surge RGCP sequencing when genomically novel/interesting variant detected
- These recommendations are specific to EWS for fast-growing variants. Separate surveillance systems need to be evaluated regarding severity and vaccine escape.
- EWS for changes in severity is currently impractical in most locales, since clincal data can not be matched with community samples & severe outcomes (deaths) are rare

Imperial College London

MRC Medical Research Council

Thank you!

COG UK
COVID-19 GENOMICS UK CONSORTIUM

wellcome

& UKHSA genomics cell