Re: Expanding cluster of a SARS-COV-2 variant in Kent and London with multiple spike mutations

From: Meera Chand, Richard Myers, Gavin Dabrera, Christina Atchison, Nick Loman, Name Redacted Andrew Rambaut, Wendy Barclay

Date: 10 December 2020

Background

Exploratory phylogenetic analysis was undertaken on available genomes from Kent as part of the epidemiological investigation into the recent increase in cases. This revealed a wider cluster with changes in spike protein. Results of a preliminary investigation are presented here. This investigation uses the currently available routine data from PHE and COG-UK.

NERVTAG is asked to note the findings and review the proposed actions.

Findings of the preliminary investigation

Epidemiology and phylogeny (Data as of 7/12/2020)

1. Analysis of routinely available genomic data for Kent was undertaken as part of the epidemiological deep dive done to investigate increasing incidence. Although only 4% (255/6130) of Kent cases had available genomes through COG-UK sequencing, a large phylogenetic cluster of 117 genomically similar cases over the week 10-18 November was identified.

2. The Kent cluster, when examined in the national phylogeny, is part of a larger cluster (962 genomes at the time of analysis). This cluster is phylogenetically very distinct from the rest of the UK dataset. These cases are concentrated in Kent and NE London, with limited spread into the rest of London, Anglia and Essex. Although genomic data has a 2-3 week lag, as of the last available data, this cluster continued to grow at a rapid pace.

3. Of the 962 cases in the cluster, data was available for 915 individuals; most specimen dates were in November (828/915) followed by October (79/915), with a small number of cases in September (4/915). For 4 cases in the cluster, apparent previous PCR-positive episodes (April – August) are present in SGSS, including 3 cases where there are intervening negative PCR results between the two PCR-positive episodes.

4. Distribution of cases by patient sex is similar (51% female, 49% male) (Table 1). By age, just under 90% of individuals are aged <60 years; work is being undertaken to compare this age distribution to relevant comparators. Six of the 915 cases are deceased.

5. Using N501Y as a marker for this strain, the frequency with which this variant is sampled increased rapidly after mid-October (figure 1). Analysis of the frequency of the variant indicates that in weeks 46-47 the relative weekly growth of N501Y was 60% (95% CI 30-89%) greater than other variants circulating in this region.

6. Among 40 local authorities in East and South East England with more than five N501Y samples there is a significant trend of increasing reported cases with increasing frequency of N501Y (Figure 2, weighted linear regression $p=10^{-6}$). A 10% difference in N501Y frequency in

mid-November corresponds to approximately 50 more weekly cases per 100 thousand in early December. Local authorities with few N501Y samples have similar reported cases as the rest of the UK (linear regression intercept = 137 cases per 100k versus UK median 130.4 per 100k).

Assessment of spike changes

7. This cluster is defined by a combination of 13 non synonymous mutations, including a series of spike protein mutations (deletion 69-70, deletion 144-145, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H). Other notable mutations include a stop codon in ORF8. This is an unusually large number of mutations in a single cluster. Similar to Danish Cluster 5, it may suggest that the virus has replicated under different selective pressures, for example in an alternative host or possibly in an immunocompromised patient, although this is speculative. A recent case report of an immunocompromised individual persistently infected with SARS-COV-2 acquired approximately 10 mutations in the spike protein over 154 days, notably including N501Y.

8. The most unusual and concerning single mutation in this cluster is N501Y. However, the summative effect of this large number of mutations is also unknown and of concern.

Spike variant N501Y

9. **Transmissibility:** It is highly likely that N501Y affects the receptor binding affinity of the spike protein and it is possible that this mutation alone or in combination with the deletion at 69/70 in the N terminal domain (NTD) is enhancing the transmissibility of the virus. This is based on the position of the 501 residue in the spike receptor binding domain and data showing that N501Y increases spike interactions with human ACE2. N501Y is one of a number of artificially generated RBD variants shown to do this (others include Y453F and N439K). It should be noted that this mutation is the only spike variant found to date in mouse-adapted SARS-CoV2 and is also seen in ferret infections.

10. **Antigenicity**: Position 501 is in the RBD, the target of neutralising antibodies, and therefore it is possible that variants at this position affect the efficacy of neutralisation of virus. Of several monoclonal antibodies tested across different studies, mutations at position 501 decreased the ability of one to neutralise(LYCoV016). N501Y was not included. There is no neutralisation data from polyclonal sera from natural infection.

Other spike variants

11. Much less is known about the other spike variants present in this cluster, with the exception of D614G which is well characterised and already highly prevalent in the UK. Their significance cannot be judged at present. Deletion at position 69/70 was present in the Danish Cluster 5 and has been seen in other clusters. Its significance is unknown. Deletions in the 145 area have been noted in infections in immunocompromised patients. Residues at positions 570, 681, 716, 982 and 1118 are of unclear significance although they fall in potentially structurally important areas of the spike protein.

12. There is a small amount of data about variants affecting ORF8, a viral accessory protein which may be involved in immune evasion by downregulation of MHC class I. In Singapore

an ORF8 deletion was associated with attenuated disease, but in this was not supported by findings in primary human airway cell experiments.

Summary

13. Investigations of Kent have revealed a large and rapidly growing cluster, primarily in Kent and London, with an unusually large number of spike mutations including N501Y, for which there is some concerning associated data. There is no neutralisation data from polyclonal sera.

14. Recommendations for next steps

a. WGS surveillance should be enhanced in Kent and London. The extent and spread of this cluster should be monitored and further epidemiological investigation undertaken.
b. Sampling should be undertaken rapidly in Kent to obtain material for viral culture
c. Fitness of this mutant should be assessed in primary human airway cultures
d. Neutralization studies should be undertaken using live virus and pseudovirus.
e. Information should be sought on the genomes present in international data in GISAID with variants at position 501.
f. Data should be compiled for a full variant risk assessment.

Figures and tables

Age Group	Female	Male	Female %	Male %	Total
<20	73	69	51%	49%	142
20-29	85	97	47%	53%	182
30-39	91	81	53%	47%	172
40-49	87	72	55%	45%	159
50-59	83	73	53%	47%	156
60-69	28	35	44%	56%	63
70-79	12	13	48%	52%	25
>=80	7	9	44%	56%	16
Total	466	449	51%	49%	915

Table 1 Age and sex distribution of cases in the cluster

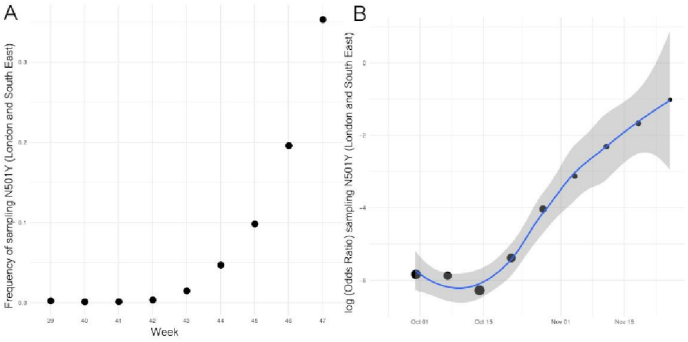

Figure 1 Frequency of the Spike N501Y variant over time in London and the South East of England. Values are based on lighthouse sampling only. A. Absolute frequency of N501Y by Epi Week. B. A local polynomial regression to the log odds of sampling the N501Y variant.

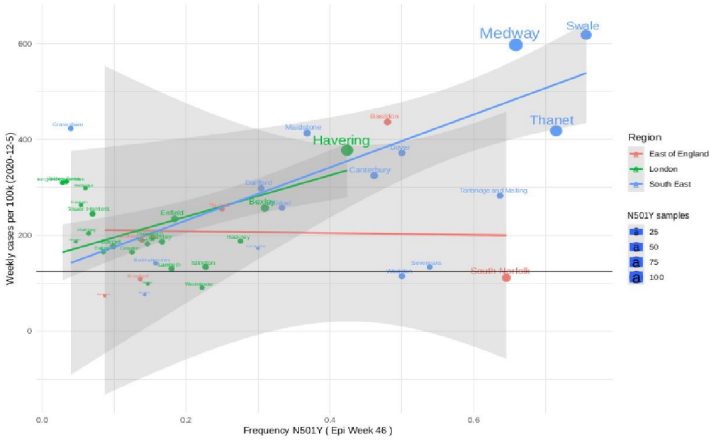

Figure 2 | Reported weekly cases per 100 thousand versus sample frequency of the Spike N501Y variant in week 46. Data are shown for 40 local authorities in England with more than 5 N501Y samples. Lines and confidence intervals are based on linear regression weighted by sample size in each local authority. The horizontal black line shows the median weekly cases among local authorities with no samples of N501Y.

Data sources and methods

Data used in this investigation is routine data from the COG-UK dataset, PHE Second Generation Surveillance System and the PHE Rapid Investigation Team Kent investigation.

For the frequency analysis:
Weekly case data for local authorities were obtained from https://imperialcollegelondon.github.io/covid19local. COG samples with available outer postcode were mapped to local authorities. Genotypes (N501Y) were inferred from the December 9 COG global alignment using an R package developed for variant analysis of COG data (https://github.com/emvolz-phylodynamics/variantAnalysis). Sequences with missing or ambiguous characters at this position were excluded. Samples taken from the hospital setting were excluded due to concerns about randomness of sampling. To correct for different rates of sampling over time and space, we weighted each sample inversely to the proportion of reported cases with sequence data in each local authority (https://github.com/robj411/sequencing_coverage). These weights were used in all analyses. Relative growth of N501Y was modeled using a weighted local polynomial regression fitted to log OR of sampling this variant by week. The relative growth is inferred from the time derivative of estimated log OR which can also be interpreted as a selection coefficient if the difference in growth rates is due to phenotypic differences in the virus.