Lineage replacement and evolution captured by the United Kingdom Covid Infection Survey

Katrina A. Lythgoe*+[1,2], Tanya Golubchik*[1], Matthew Hall[1], Thomas House[3], George MacIntyre-Cockett[1,4], Helen Fryer[1], Laura Thomson[1], Anel Nurtay[1], David Buck[4], Angie Green[4], Amy Trebes[4], Paolo Piazza[4], Lorne J Lonie[4], Ruth Studley[5], Emma Rourke[5], Duncan Cook[5], Darren Smith[6], Matthew Bashton[6], Andrew Nelson[6], Matthew Crown[6], Clare McCann[7], Gregory R Young[6], Rui Andre Nunes dos Santos[7], Zack Richards[7], Adnan Tariq[7], Wellcome Sanger Institute COVID-19 Surveillance Team[**], COVID-19 Infection Survey Group, The COVID-19 Genomics UK (COG-UK) Consortium[++,8], Christophe Fraser[1,4,9], Ian Diamond[5], Jeff Barrett[9], Sarah Walker[10,11,12,13], David Bonsall[1,4,14]

*Equal contribution
+Corresponding author: katrina.lythgoe@bdi.ox.ac.uk

[1]Big Data Institute, Nuffield Department of Medicine, University of Oxford, Old Road Campus, Oxford OX3 7LF, UK.
[2]Department of Zoology, University of Oxford, Oxford OX1 3SZ, UK
[3]Department of Mathematics, University of Manchester, Manchester, M13 9PL.
[4]Wellcome Centre for Human Genetics, Nuffield Department of Medicine, NIHR Biomedical Research Centre, University of Oxford, Old Road Campus, Oxford OX3 7BN, UK.
Department of Zoology, University of Oxford, Oxford OX1 3SZ, UK
[5]Office for National Statistics, Newport, UK
[6]The Hub for Biotechnology in the Built Environment, Department of Applied Sciences, Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK
[7]Department of Applied Sciences, Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK
[8]https://www.cogconsortium.uk
[9]Wellcome Sanger Institute, Cambridge CB10 1SA, UK
[10]Nuffield Department of Medicine, University of Oxford, Oxford, UK
[11]The National Institute for Health Research Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at the University of Oxford, Oxford, UK.
[12]The National Institute for Health Research Oxford Biomedical Research Centre, University of Oxford,Oxford, UK
[13]MRC Clinical Trials Unit at UCL, UCL, London, UK
[14]Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK

[**]The full list of names is available at https://www.sanger.ac.uk/project/wellcome-sanger-institute-covid-19-surveillance-team/
[++]The full list of names and affiliations of COG-UK members is provided in the appendix

Abstract

The Office for National Statistics COVID-19 Infection Survey is a large household-based surveillance study based in the United Kingdom. Here, we report on the epidemiological and evolutionary dynamics of SARS-CoV-2 determined by analysing sequenced samples collected up until 13th November 2021. We observed four distinct sweeps or partial-sweeps, by lineages B.1.177, B.1.1.7/Alpha, B.1.617.2/Delta, and finally AY.4.2, a sublineage of B.1.617.2, with each sweeping lineage having a distinct growth advantage compared to their predecessors. Evolution was characterised by steady rates of evolution and increasing diversity within lineages, but with step increases in divergence associated with each sweeping major lineage, leading to a faster overall rate of evolution and fluctuating levels of diversity. These observations highlight the value of viral sequencing integrated into community surveillance studies to monitor the viral epidemiology and evolution of SARS-CoV-2, and potentially other pathogens, particularly as routine PCR testing is phased out or in settings where large-scale sequencing is not feasible.

Main text

A crucial component of the global response to COVID-19 is the identification, tracking and characterisation of new SARS-CoV-2 lineages. As well as enabling researchers to identify patterns of spread, variants can be identified that might pose a particular risk. For instance they may be able to transmit more easily, or evade immune responses. Prominent examples include the variants of concern (VOCs) Alpha, Beta, Gamma, Delta and Omicron [1], and individual mutations such as E484K, an immune escape mutation in the Spike protein [2]. At the time of writing, the COG-UK Genomics Consortium [3] has produced over 1.6 million SARS-CoV-2 sequences, primarily from positive RT PCR tests, with this substantial surveillance effort generating a snapshot of the leading edge of infection across the UK.

Estimating the prevalence of SARS-CoV-2 lineages and/or mutations can, however, be subject to biases as a consequence of the sampling regime [4–6]. Sampling has been heavily focussed on symptomatic infections, even though a high proportion of infections are asymptomatic or may not reach the criteria for testing [7]. For example, in the early phase of the UK epidemic most testing was conducted among hospitalised patients with severe disease, with a later focus on symptomatic individuals. Where testing of asymptomatic individuals has been conducted, it has often been in the context of specific settings, such as returning travellers, schools, or as part of surge testing in geographical areas where VOCs have been identified [8]. Large-scale community surveillance studies, such as the Office for National Statistics (ONS) Covid Infection Survey (CIS) [6], and the Real-time Assessment of Community Tranmission (REACT) [9,10] are thus valuable since sampling is not subject to these biases, they consist of a random, potentially more representative sample of the population, and, crucially, identify both symptomatic and asymptomatic infections. Moreover, community-based surveillance studies are not reliant on sequencing samples collected as part of national RT-PCR testing programmes. They will therefore become increasingly important as routine RT-PCR testing is scaled down, or as countries seek to enhance surveillance capabilities for SARS-CoV-2 and other pathogens .

The Office for National Statistics (ONS) COVID-19 Infection Survey (CIS) is a United Kingdom (UK) household-based surveillance study, with households approached at random

from address lists to ensure as representative a sample of the population as possible [6,11]. RT-PCR positive samples collected during the survey were sequenced as part of the COG-UK Genomics Consortium [3]. Here, we present an analysis of the 16817 consensus sequences from RT-PCR positive samples collected between 26th April 2020 to 13th November 2021 that had genomic coverage over 50%, with the aim of reconstructing the key epidemiological and evolutionary features of the UK epidemic. These data capture the sequential sweeps and partial sweeps of the B.1.177, B.1.1.7/Alpha, and B.1.617.2/Delta, and AY.4.2 lineages which the UK experienced, plus the sporadic appearance of other VOCs and Variants Under Investigation (VUIs), most notably B.1.351/Beta and P.1/Gamma. For each of the sweeping lineages we calculated their growth rate advantage, with each lineage having a progressively higher growth rate compared to previously circulating lineages.

A feature of B.1.1.7/Alpha is the RT-PCR S-gene target failure (SGTF) caused by the Spike DH69/V70 deletion. During the UK epidemic, SGTF greatly facilitated the rapid quantification of B.1.1.7/Alpha case numbers [12–15], and subsequently non-SGTF was used to quantify B.1.617.2/Delta case numbers [16,17]. By comparing the presence or absence of SGTF with genotype for samples in the ONS CIS, we determined the specifity of RT-PCR S-gene target failure (SGTF), for B.1.1.7/Alpha, and non-SGTF for B.1.617/Delta, with high specificity when lineage prevalences were high (~99%), but low specificity at low prevalences.

In addition to measuring the growth rate advantage of sweeping lineages, we determined how these sweeps impacted measures of the genetic diversity and divergence of the virus, both at the within-lineage and between-lineage levels. As well as VOCs, which are characterised by a large number (constellation) of mutations, single mutations can also be a cause for concern. For example, the appearance of the E484K mutation in Spike, which likely contributes to the partial immune escape [2] on the highly transmissible B.1.1.7/Alpha genetic background was rightly seen as a cause for concern [18]. We determined the number of samples in which we saw any of the Spike amino acid replacements reported to confer antigenic change to antibodies as listed by the COG-UK Mutation Explorer [2,19], taking particular note of those where the mutation was not lineage defining given the lineage of the sample, and therefore likely to represent recently acquired mutations. In addition, we calculated how often the ancestral nucleotide changes corresponding to these replacements appeared on the phylogenic tree of the ONS CIS samples. Most mutations were rare and sporadic, appearing in a single sample, but others were notably more common, with some having appeared independently on multiple phylogenetic lineages, suggesting convergent evolution.

Although sequences from the ONS CIS represented about 1.6% of the total number of SARS-CoV-2 sequences obtained in the UK during this period, we were able to reconstruct the key epidemiological and evolutionary aspects of the epidemic. Our observations highlight the value of incorporating sequencing into large-scale surveillance studies of infectious disease, and the important role that community-based genomic surveillance studies can have in the monitoring of infectious disease. Although the ONS CIS is based in the UK, in which sequencing effort has been unprecidented, this is of particular importance in settings where routine testing is likely to be scaled back, and for countries exploring the best strategies for tracking SARS-CoV-2 as well as other pathogens.

Results

Sequential replacement of lineages in the UK

Between April 2020 and late December 2020, a random selection of RT-PCR samples collected as part of the ONS CIS were sequenced. There was a substantial increase in the size of the ONS CIS between August and October 2020, which went from testing under 50,000 people per fortnight to around 180,000 people per fortnight [20], and consequently the number of sequenced samples also increased. In response to the emergence of B.1.1.7/Alpha [21], the sequencing effort was intensified again, with the aim of prospectively sequencing all RT-PCR positive samples in addition to unsequenced earlier samples where possible. Here we report on the sequenced samples with cycle threshold (Ct) <=30 and over 50% genome coverage collected between 26th April 2020 and 13 November 2021.

The Pango lineage [22] for all the samples was determined using Pangolin v3.1.16 [23]. Since 7 December 2020 we have provided publicly available weekly reports, giving the number of sequenced samples with >50% genome coverage by lineage [24]. Observing the raw data (Fig. 1), we see a small 2020 autumn peak in sequenced samples, dominated by B.1.177 and its sub-lineages. We then see a decline in cases due to the second national lockdown which lasted from 5th November to 2nd December 2020, before the number of sequenced samples started to rise again. This is attributed to a relaxation of restrictions during the Christmas period and corresponded to a rapid rise in the number of B.1.1.7/Alpha infections. After the commencement of a further lockdown in England, Scotland and Northern Ireland in early January, cases declined again, before another rapid increase in the number of sequenced samples that were dominated by B.1.617.2/Delta, with this increase corresponding to a phased reopening on the 19th May and 20th June 2021.

When the proportion of samples belonging to each lineage are plotted over time (Fig. 1), the sequential sweeps and partial sweeps of B.1.177, B.1.1.7/Alpha, and B.1.617.2/Delta, can be readily observed, followed by the slower partial sweep of the B.1.617.2 sublineage AY.4.2. The lineage dynamics observed using ONS CIS samples are broadly in line with those observed for pillar 2 (community testing) samples across the whole COG-UK consortium, but excluding the ONS CIS samples (Figs. S1, S2). However, some differences are noticeable. The first samples of each sweeping lineage were collected earlier by COG-UK, and sustained lineage growth rates were also noticeable two-to-four weeks earlier among COG-UK samples for lineages B.1.177, B.1.617.2/Delta and AY.4.2, but at a similar time to the ONS-CIS for B.1.1.7/Alpha. This reflects the much larger number of COG-UK compared to ONS-CIS sequences, but also may reflect differences in the demographics sampled. It is interesting that the lineages dynamics among the two datasets are most similar for B.1.1.7/Alpha, which most likely emerged within the UK, whereas initial growth of the other lineages was driven by importations.

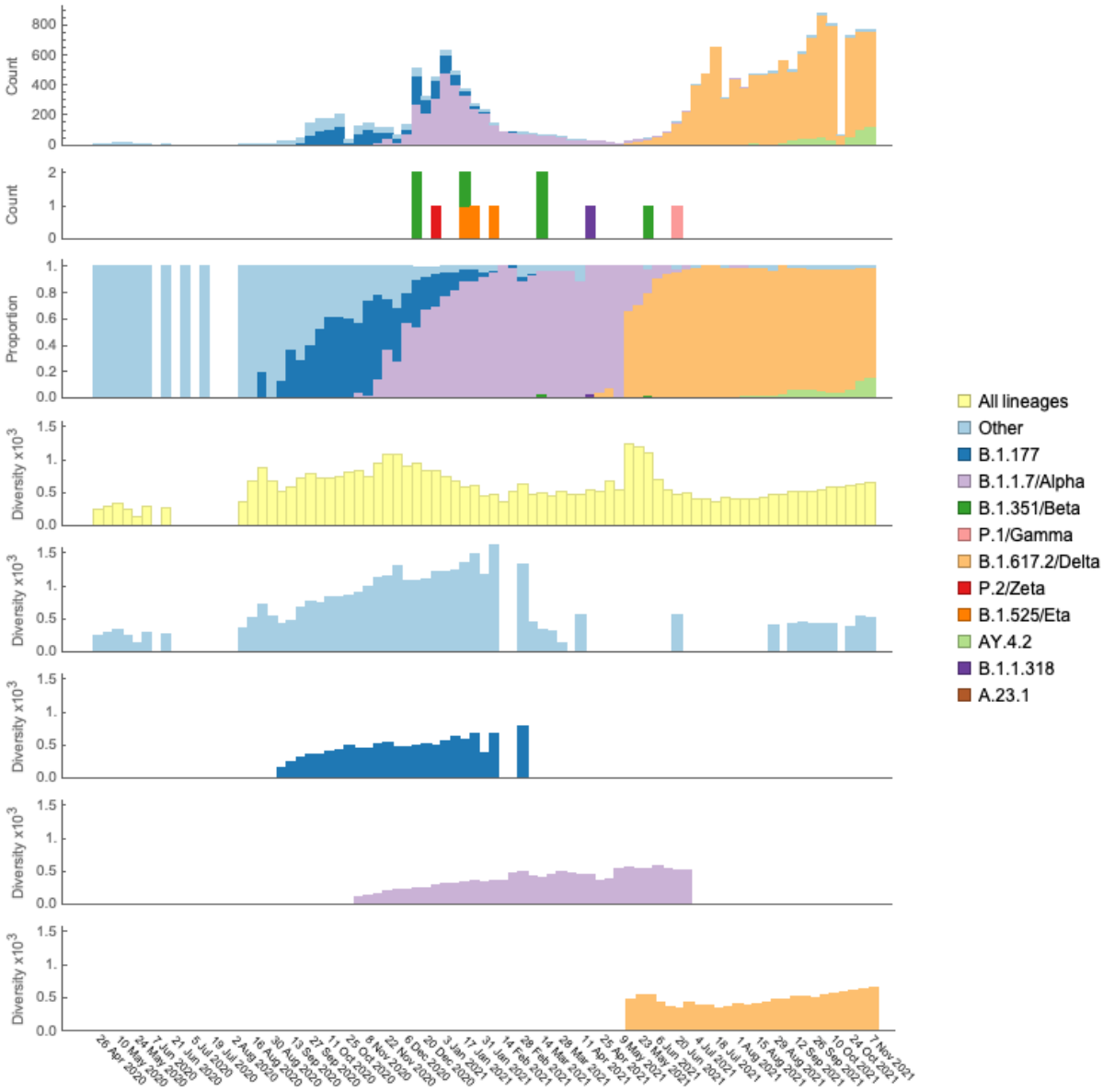

Figure 1. **Sequenced samples and genetic diversity by lineage.** A. Number of sequenced samples with >=50% genome coverage, coloured by lineage. Named lineages include all sub-lineages apart from B.1.617.2, for which sublineaege AY.4.2 is coloured separately. B. Number of VOC and VUI sequenced samples, as designated by Public Health England, but excluding Alpha and Delta. C. Proportion of samples belonging to each lineage. D. Genetic diversity among all samples, based on the consensus sequence of each sample. E. Genetic diversity among all samples with a fully resolved lineage, excluding B.1.177, Alpha, Delta and AY.4.2. F-H. Genetic diversity among all B.1.177 (E) B.1.1.7/Alpha (F) and B.1.617.2/Delta including AY.4.2 (G) samples. All samples are grouped by the week in which they were collected, with the date giving the first day of the collection week (every other week labelled for clarity).

B.1.177, B.1.1.7/Alpha, B.1.617.2/Delta and AY.4.2 each had growth advantages

For each of the sweeping lineages in turn, we calculated the relative growth advantage compared to all other contemporary lineages using the ONS-CIS data (Fig. 2). In line with previous findings [25–27], we found that B.1.177 had a significant growth rate advantage compared to all other co-circulating SARS-CoV-2 lineages, which peaked at around 0.075

per day towards the end of September 2020 before slowly declining through October 2020. In other words, B.1.177 case numbers grew about 7.5% faster each day when compared to all other cases. The B.1.177 lineage most likely originated in Spain, with the first samples collected in June 2020 [26], and with international travel from Europe being a major driver of new infections of this lineage in the UK from August 2020 [25,26]. However, there is continuing debate as to whether the growth advantage of B.1.177 was also associated with increased transmissibility, with some reports arguing there is little evidence for increased intrinsic transmissibility [25,26], whilst others suggesting importations alone cannot explain the patterns of replacement [27]. In the ONS CIS data, we observed a continued growth advantage of B.1.177 in the UK thoughout October 2020, when the number of incident B.1.177 infections was relatively high (Fig. S1) whilst travel to the UK from other European countries had tailed off [25,26], which is consistent with increased B.1.177 transmissibility.

If we consider the growth advantage of B.1.1.7/Alpha, which largely replaced B.1.177 and other circulating lineages between October 2020 and February 2021, we continue to find a pattern consistent with a transmission advantage of B.1.177 over previously circulating lineages. The early growth advantage of B.1.1.7/Alpha, when compared to other contemporaneous lineages, reached a peak of approximately 0.14 per day in November 2020, when both B.1.177 and earlier lineages were in circulation (Figs 1, S1), before settling to around 0.05 per day in December 2020, by which time almost all non-Alpha lineages were B.1.177. This is the pattern we would expect if B.1.177 has a transmission advantage compared to other (non-B.1.1.7/Alpha) lineages; B.1.1.7/Alpha will grow relatively faster in a background of B.1.177 and other less transmissible lineages, but will grow relatively slower if the background consists of only B.1.177. This declining growth rate advantage of B.1.1.7/Alpha after initial high values has also been noted in previous reports [14,27–30].

There is little doubt, however, that an intrinsic transmission advantage of B.1.1.7/Alpha was the major force driving the rapid increase of B.1.1.7 infections [14,27–30]. Finally, by the time B.1.617.2/Delta emerged it was in a background of almost exclusively B.1.1.7/Alpha infections, making interpretation much more straightforward. We estimate that B.1.617.2/Delta had a growth advantage of around 0.12 per day compared to B.1.1.7/Alpha, in line with previous estimates [27,31], likely due to a combination of increased transmissibility and immune evasion [27]. More recently AY.4.2, a Delta sublineage with Y145H, a potential antibody escape mutation, is currently increasing in frequency in the UK [32,33]. We estimate AY.4.2 and its sublineages (which in turn is a sublineage of B.1.617.2), had a growth advantage of 0.02 per day compared to all other Delta lineages, with an estimated doubling time of 28 days.

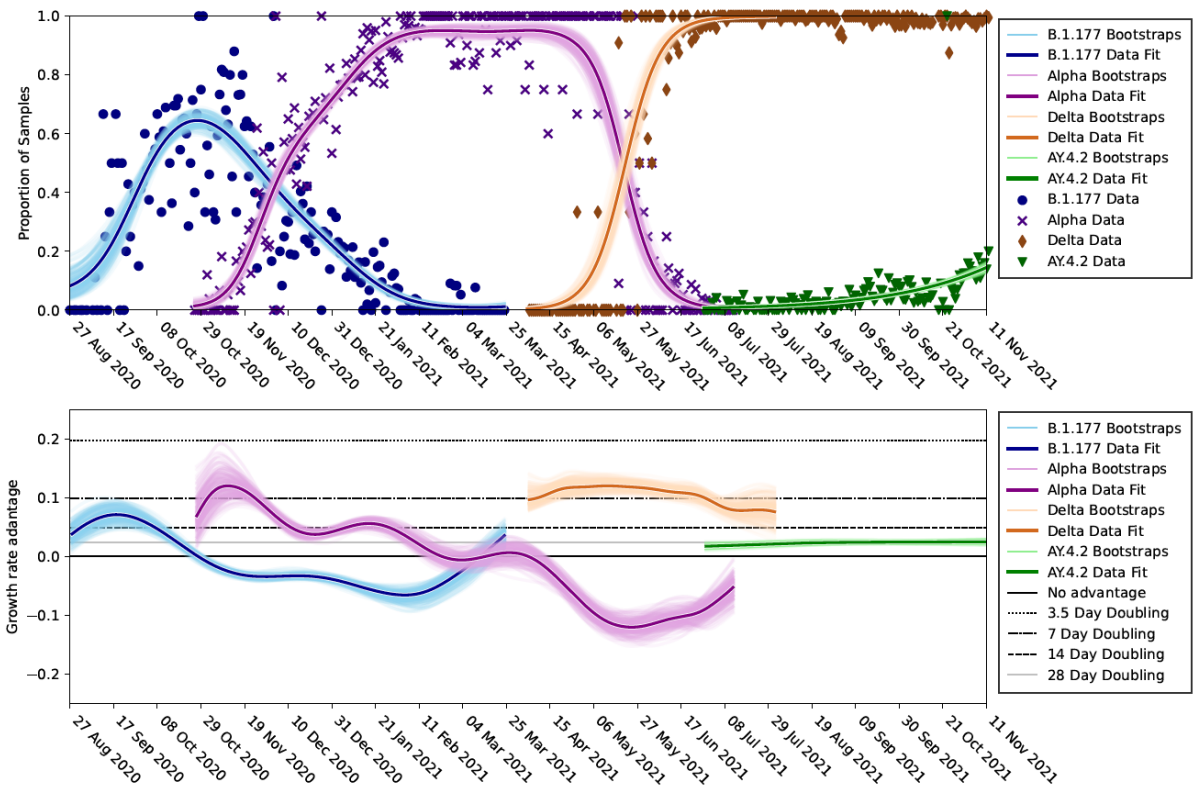

Figure 2. Relative growth advantages of B.1.177, B.1.1.7/Alpha, B.1.617.2/Delta and AY.4.2. Top. The proportion of samples belonging to each of the lineages compared to all other contemporary samples, with uncertainty represented by 200 data bootstraps. Bottom. Per day growth rate advantage of each of the four lineages compared to all other contemporary samples. The doubling times represent how long it would take for the frequency of the lineage to double if current trends continued. The Delta curve includes B.1.617.2 and all sublineages, including AY.4.2.

SGTF was a good proxy for lineage when lineage prevalence was high, but not when lineage prevalence was low

SGTF was an attractive surrogate marker for B.1.1.7/Alpha in the UK since it enabled the lineage to be tracked in the population without the delays incurred due to sequencing, and enabled samples with high cycle threshold (Ct) to be included in analyses even if they couldn't be sequenced. For all samples in the analysis, we classed those as having SGTF if, during RT-PCR testing, N and ORF1ab were successfully amplified, but S was not, and non-SGTF samples as those where N, ORF1ab and S were all amplified. Since we only consider samples with Ct<=30 (and hence those with relatively high viral loads), almost all samples were classified as SGTF or non-SGTF. Lineage was generally a good indicator of SGTF, with 99.0% (3424/33460) of B.1.1.7/Alpha and 100% (3/3) of B.1.525/Eta variant samples, both of which have the DH69/V70 deletion, having SGTF (Table S1). Other lineages, which do not have the DH69/V70 deletion, typically did not have SGTF; no B.1.351/Beta, P.1/Gamma had it, although a few B.1.617.2/Delta samples (14/10984) were SGTF. As previously reported [34], the exception was the B.1.258 lineage, of which around three-quarters of samples had the DH69/V70 deletion and SGTF (47/62).

While B.1.1.7/Alpha prevalence was high (between December 2020 and June 2021), SGTF was highly specific for Alpha, and similarly while B.1.617.2/Delta prevalence was high (May-

Nov 2021) non-SGTF was highly specific for B.1.617.2/Delta (Fig. 3, Table S2). However, when the prevalence of these variants was lower, SGTF was a poorer indicator of lineage, either due to the presence of other co-circulating lineages with or without the DH69/V70 deletion, or due to SGTF independent of the corresponding B.1.617.2/Delta lineage mutation as a result of RT-PCR technical failures or other errors. Prior to November 2020, most SGTF samples were associated with the B.1.258 lineage, and conversely, from February to April 2021, non-SGTF samples represented a broad range of lineages, including Alpha, making non-SGTF a poor indicator of any specific lineage during this period.

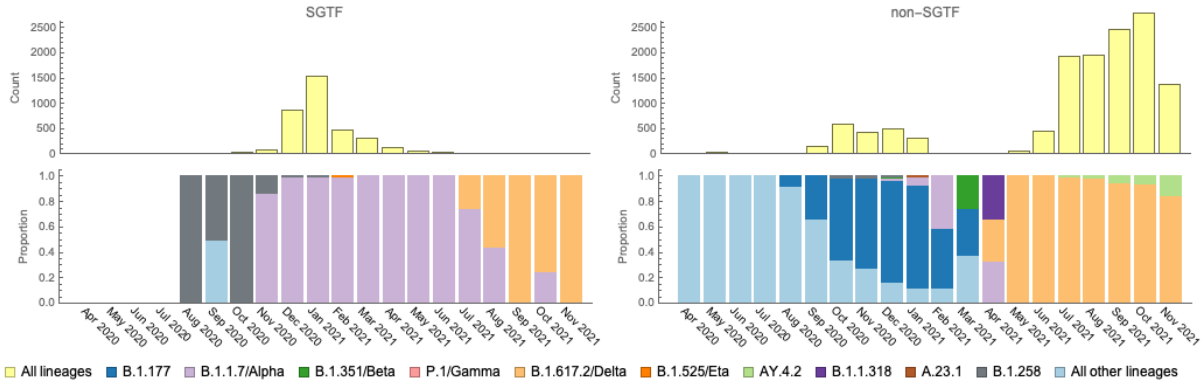

Figure 3. Comparison of S-gene target failure with lineage by calendar month. Top row. Number of SGTF (left) and non-SGTF (right) samples by calendar month.Bottom row. Proportion of SGTF (left) and non-SGTF (right) samples that are of a given lineage.

Diversity increases within lineages through time, but fluctuates when measured across all lineages

The sequential sweeps of B.1.177, B.1.1.7/Alpha and B.1.617.2/Delta (including AY.4.2) are readily observable on the time-scaled phylogeny of ONS CIS consensus sequences, with each of the lineages representing a distinct clade (Figs. 4, S3). Both the B.1.177 and B.1.1.7/Alpha clades have times of most recent common ancestor (tMRCAs) close to the time of first sampling, indicating the recent emergence of these lineages.

These patterns are also reflected in measurements of genetic diversity, with diversity among lineages (excluding B.1.177, B.1.1.7/Alpha and B.1.617.2/Delta) ("Other"; Fig. 1) showing a pattern of initial low diversity, followed by increasing diversity until February 2021, after which there are very few samples in the ONS CIS. Similarly, within both the B.1.177 and B.1.1.7/Alpha lineages, diversity was relatively low when they first appeared, and gradually increased through time (Fig. 1). This initial low diversity is a consequence of their relatively recent emergence in Spain and South East England, respectively, before first detection in the ONS CIS data. The slightly higher diversity in B.1.177 likely reflects its geographical spread and multiple introductions from Europe. B.1.617.2/Delta, on the other hand, had high initial diversity (Fig. 1), with multiple introductions into the UK of this lineage from an already diverse source population in India [16]. Diversity in this lineage then declined slightly corresponding to fewer introductions as a result of travel restrictions from India, before then steadily rising again.

In contrast, when we consider overall genetic diversity, we see transient increases, peaking when two or more distinct lineages are at relatively high frequencies (Fig. 1), but then

declining as single lineages dominate the population. It is notable that in August 2021 overall levels of diversity were lower than in August 2020 despite much higher prevalence (e.g. in England 1.32% in the week ending 31 July 2021 vs 0.05% the week ending 25 August 2020).

Divergence increases through time, but at different rates within and among lineages
As expected, divergence from the root of the phylogeny increased gradually through time, both within-lineages (Fig. S4) and across all lineages (Fig. 4), demonstrating the presence of a strong molecular clock. It has been noted previously that although divergence within the B.1.1.7/Alpha lineage increased at a similar rate to previously circulating lineages, it had accumulated a disproportionate number of lineage defining mutations [21]. We also observe this pattern, with similar estimated molecular clock rates (line gradients) of 0.00038 substitutions per site per year (s/s/y) for B.1.177, and of 0.00034 s/s/y for B.1.1.7/Alpha, but with the B.1.1.7 line appearing shifted upwards (Fig. S4). Since many of the B.1.1.7/Alpha lineage defining mutations were nonsynonymous and in the Spike region, it has been hypothesised Alpha arose during a long-term chronic infection [21].

Meanwhile, the data indicate a higher substitution rate for B.1.617.2/Delta (0.00064 s/s/y), again with a shift up in the regression line, possibly as a consequence of divergence within India before importation into the UK. Finally, the observed substitution rate over all ONS sequences is 0.00094 s/s/y, which is faster than any of the within-lineage rates we measured, and is largely a result of the step increases in divergence associated with the new variants.

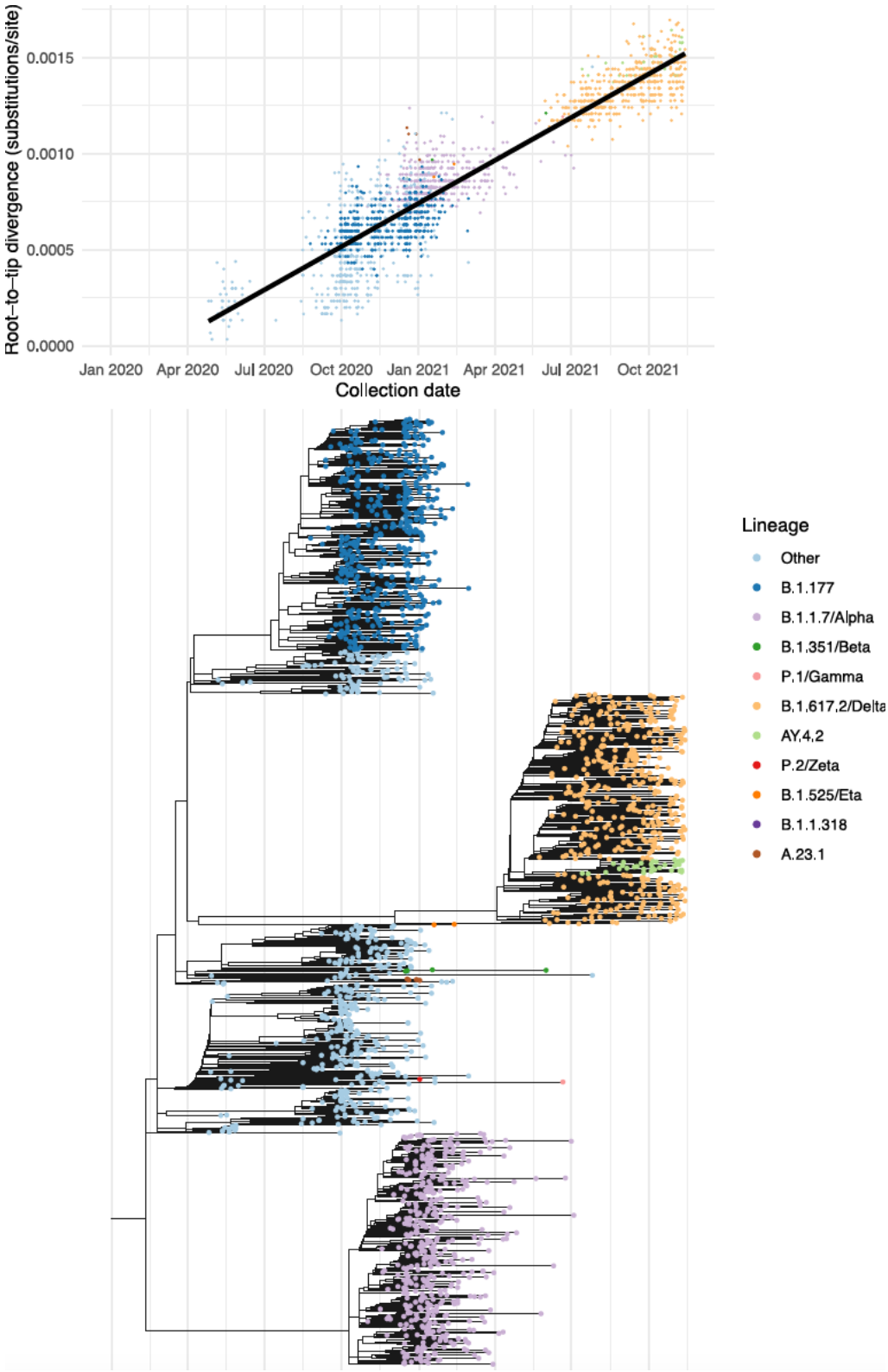

Figure 4. Dated phylogeny and root-to-tip distance of ONS CIS sequences. First, a maximum likelihood phylogeny of ONS sequences with over 95% genome coverage up to and including 13th November 2021 (Fig S3) was generated using RAxML-NG. A. Root to tip distance for samples from the maximum likelihood phylogeny. B. Time tree generated from the maximum likelihood phylogeny using TreeTime [35]. B.1.177, B.1.1.7/Alpha, B.1.617.2/Delta (including AY.4.2) and 'Other' were each randomly subsampled to a maximum of 500 sequences.

Spike mutations conferring antigenic change

There is justified concern that, as population levels of immunity increase via vaccination and/or prior infection, SARS-CoV-2 may acquire immune escape mutations on already highly transmissible genetic backgrounds. For example, concerns were raised over the presence of E484K in some B.1.1.7/Alpha isolates in the UK [36]. More recently the AY.4.2 sublineage of Delta has Y145H, a potential antibody escape mutation. We determined the number of consensus sequences with potential antibody escape mutations for each of the four sweeping lineages in our dataset: B.1.177, B.1.1.7/Alpha, B.1.617.2/Delta and AY.4.2. We considered Spike amino acid replacements reported to confer antigenic change to antibodies as listed by the COG-UK Mutation Explorer [2,19], and only considered mutations which were not lineage defining on the genetic background on which they were found. For example, L18F is found in the majority of B.1.177 samples, but only a small proportion of Alpha and Delta samples, and therefore L18F was excluded on a B.1.177 background but included on a B.1.1.7/Alpha or B.1.617.2/Delta background.

Of the 1415 B.1.177 samples in our dataset, 20 (1.4%) had a non-lineage defining antigenic mutation of concern; this was also true of 159 (4.6%) out of 3454 B.1.1.7/Alpha samples, 292 (2.8%) of 10421 B.617.2/Delta samples, and 7 (1.2%) of 578 AY.4.2 samples. In total, 65 unique mutations at 55 unique residue sites were observed among the four lineages (Fig.5). By performing ancestral state reconstruction on the phylogeny of all over 95% coverage ONS CIS samples collected before 17th July 2021 (Fig. S5) we determined the number of ancestral occurances of each of these mutations (multiple ancestral nodes are depicted on Fig. 5). Whereas most mutations have a single origin on the tree, some appear multiple times, including L18F (14 times), H146X (9 times), E484K (5 times), and S255F (6 times). Some of these mutations are lineage defining for multiple lineages, indicating convergent evolution and benefit on different genetic backgrounds. Most however, only appeared once, suggesting they may have had limited evolutionary advantage when occurring in isolation, at least during the sampling period considered. It remains possible that these mutations may have an advantage on different genetic backgrounds (epistasis) and/or in different environments, for example as acquired immunity increases due to infection or vaccination.

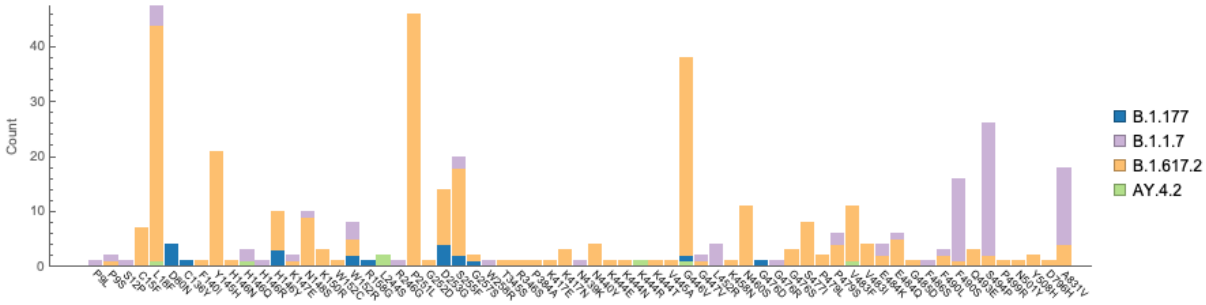

Figure 5. Mutations in Spike conferring antigenic change for the three most common lineages. The stacked bar chart indicates the number of samples in our dataset with the mutation. All antigenic mutations included in the COG-UK mutational explorer [19], were included. For clarity, mutations on a lineage that define that lineage are not considered. The y-axis has been truncated for clarity, but L18F was observed in 77 Alpha samples.

Discussion

Surveillance studies such as the ONS CIS are valuable tools for tracking the emergence and spread of infectious disease. Since participants are selected at random and are periodically tested for SARS-CoV-2 infection regardless of symptoms, the ONS CIS gives an accurate picture of SARS-CoV-2 prevalence in the UK that is not subject to biases due to, for example, increased sampling effort in different geographical areas or demographic groups, or among symptomatic individuals [4–6]. By sequencing RT-PCR positive samples collected as part of ONS-CIS data between April 2020 and November 2021, we observed four sweeps or partial sweeps by lineages B.1.177, B.1.1.7/Alpha, B.1.617.2/Delta, and AY.4.2. This resulted in a pattern of relatively steady within-lineage evolution, followed by periodioc replacement by faster growing lineages which were characterised by a step-increase in the number of substitutions. This in turn resulted in faster overall rates of evolution when measured across all lineages, and fluctuating levels of genetic diversity. Whether this pattern will be an ongoing feature of SARS-CoV-2 evolution remains to be seen.

Of the ~1.3 million UK sequences collected during the period studied here as part of the COG-UK consortium [3], about 1.6% were samples collected as part of the ONS CIS. The comparatively smaller sample sizes associated with the ONS-CIS makes it more difficult to identify small clusters of infection, and may delay the detection of lineages with a growth rate advantage. For example, only two B.1.258/Beta samples were identified from sequenced ONS CIS samples during February, March and April 2021, despite the occurance of a sizeable outbreak in South London during this time. Moreover, a clear increase in the proportions of B.1.177, B.1.617.2/Delta, and AY.4.2 cases was detected about two-four weeks later in the ONS CIS data compared to the whole COG-UK data. The sampling of all individuals regardless of symptoms in ONS may also generate a short lag in the detection of a growth advantage compared to COG-UK as a whole: ONS effectively measures pathogen prevalence (total number of infected individuals), whereas COG-UK data effectively measures incidence (total number of new infections).

As well as the smaller sample sizes contributing to the delayed detection of growing lineages, it is also possible that a higher proportion of COG-UK samples represented communities and/or demographics where imported cases took hold. We note that detection

of B.1.1.7/Alpha, which likely emerged in the UK, was not associated with delayed detection in the ONS-CIS data. Other community surveillance sampling strategies can partially compensate for the lower number of sequenced samples on a given day by concentrating sampling over a short period of time. For example, the REACT study [10] concentrates sampling over a few days each month, which enabled characterisation of the spread of B.617.2/Delta in the UK [37]. However, this is at the cost of the temporal granularity needed to rapidly detect fast growing variants such as B.1.1.529/Omicron.

A further drawback of relying on sequening for genomic surveillance is the delay between sample collection and subsequent sequencing (the ONS CIS has lags in the data of two weeks or more) and the need for high viral loads to produce adequate sequence data. This in turn could severely impact the success of any interventions. The earliest signals that both B.1.1.7/Alpha and B.1.1.529/Omicron [38] had a growth rate advantage were serendipitously inferred from the increasing incidence of SGTF during RT-PCR testing. The introduction of qPCR-based genotyping for specific VOCs and VUIs into diagnostic pipelines has the potential to speed-up detection of known variants. However, qPCR-based genotyping cannot be relied upon to characterise emerging variants fast enough to contain them due to the lead time required to manufacture and distribute specific assays. Moreover, substantial genome sequencing efforts will always be required to detect variants that have not previously been identified as of concern, to monitor the ongoing specificity of rapid genotyping in the face of ongoing evolution, and to better characterise the evolution and spread of the virus.

Due to foresight and investment, the COG-UK consortium has now sequenced over 1.6 million RT PCR positive samples spanning the UK SARS-CoV-2 epidemic. This has enabled the detection and tracking of genomic variants in the UK [21], quantification of their growth advantage [14], and inference regarding patterns of spread [28,39]. However, it is unlikely that this unprecedented sequencing effort can be sustained in the long term, and in most countries this level of sequencing has never been feasible, with patchy sequencing efforts among different countries and regions [40]. Although only a fraction of the COG-UK sequences were comprised of samples collected as part of the ONS CIS, we were able to use ONS CIS sequenced samples to monitor the emergence, spread and evolution of the major lineages and sublineages sweeping through the UK population.

Moving forwards, the implementation of genomic surveillance globally should be considered a key development goal, enabling the early detection of worrisome and/or rapidly growing lineages wherever they emerge. Community surveillance studies similar to ONS-CIS may therefore provide powerful cost-effective tools for pathogen genomic surveillance in the future, particularly if combined with the continued sequencing of a small proportion of samples from symptomatic individuals. Incorporating the detection and sequencing of other pathogens into the same community surveillance frameworks will only act to enhance the positive public health and scientific outcomes from these studies whilst maximising value for money.

Methods

ONS COVID-19 Infection Survey

The ONS CIS is a UK household-based surveillance study with selected households chosen to ensure a representative sample of the population. For a full description of the sampling design see [6], but in brief swabs were taken from individuals aged two years and older living in private households, from 26th April 2020 onwards. These households were selected randomly from address lists and previous ONS surveys to provide a representative sample of the population. Participants could provide consent for optional follow-up sampling weekly for the first five weeks, and monthly thereafter.

This work contains statistical data from ONS which is Crown Copyright. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

Sequencing

For samples collected from 26th April 2020 to approximately mid-December 2020, a random selection were selected for sequencing. RNA extracts were amplified using the ARTIC amplicon protocol [41] and most were sequenced on Illumina Novaseq with consensus fasta sequences produced using the ARTIC nextflow processing pipeline [41]. A small number of samples (36) in our study were sequenced using Oxford Nanopore GridION or MINION. Thereafter, the ambition was to sequence all positive samples including retrospective sequencing of stored RT-PCR positive samples where available (after a couple of months it was decided to only sequence samples with Ct<30, since samples with higher Ct values had a high failure rate due to low levels of virus). The move to sequence all RT-PCR positive samples coincided with a move to veSeq, an RNASeq protocol based on a quantitative targeted enrichment strategy [42,43] and sequenced on Illumina Novaseq. Consensus sequences were produced using *shiver* [44]. If the same sample was sequenced twice, the consensus sequence with the lowest genome coverage was excluded from the analysis. Fig. S6 shows the proportion of all ONS CIS RT PCR positive samples with Ct <=30, and the proportion of these samples that have sequence >50% coverage. Finally, from mid-July 2021 onwards, sequences were again sequenced using ARTIC.

We have previously shown that Log10 viral load is positively correlated with the Log10 number of mapped reads obtained using veSeq, and that Ct is negatively correlated with the Log10 number of mapped reads [42,45]. For all samples sequenced using the veSeq protocol we compared the Log10 mapped reads with Ct, obtaining a strong negative correlation [Fig. S7]. Outliers may have a number of different causes including PCR amplification or sequence failures. Samples from individual sequencing plates where no clear correlation was observed were excluded from all analyses. All remaining sequences with coverage >=50% were included. In total, 10,042 ARTIC/Illumina, 36 ARTIC/Nanopore, and 6775 veSeq/Illumina sequences were included.

Lineage calling

Lineages using the Pango nomenclature [22] were determined using the Pangolin software [23]. Reported lineages include any sub-lineages, except where stated otherwise. For example, B.1.177 includes all sub-lineages of B.1.177, and B.1.617.2 includes all AY.x

lineages. When comparing SGTF with lineage, we excluded samples where the lineage resolved to A, B, B.1 or B.1.1 since B.1.1.7/Alpha samples can be given these Pango lineages if an insufficient number of loci have coverage at lineage defining sites.

Lineage growth rates and doubling times

For each of the three most common lineages observed in our dataset, B.1.177, B.1.1.7/Alpha, and B.1.617.2/Delta, we calculated their relative growth rate advantage compared to all other lineages in our data using a generalised smoothing method.

Suppose we have two species with exponential growth rates r_1 and r_2 and therefore expected counts

$$Z_1(t) = Z_1(0)e^{r_1 t}, Z_2(t) = Z_2(0)e^{r_2 t},$$

respectively. Then the probability of a uniform random sample from both being of type 1 is

$$p(t) := \frac{Z_1(t)}{Z_1(t) + Z_2(t)}.$$

We now consider the **log odds**, $f(t)$ of being type 1 over time; using the standard definition of these we obtain

$$f(t) := log(\frac{p(t)}{1 - p(t)}) = log(\frac{Z_1(0)}{Z_2(0)}) + (r_1 - r_2)t.$$

Therefore, if we take the **derivative** of this quantity, we obtain the relative growth rate advantage for species 1 and associated doubling time,

$$r_1 - r_2 = \dot{f}(t), \tau_D = log(2)/(r_1 - r_2).$$

This doubling time when calculated gives the time for Z_1/Z_2 to double if the current trends continue.

In practice, we do not measure $p(t)$ directly, but rather work with a sample, with $\boldsymbol{X}$ being a vector of times that samples are taken and $\boldsymbol{y}$ being an associated vector with values equal to 0 if species 2 is observed and 1 if species 1 is observed. Because we wish to differentiate the time trend, traditional splines that have penalised or zero derivatives may not be appropriate, and so we place a Gaussian process prior on f with the Radial Basis Function (RBF) kernel, which has C_∞ samples (i.e., all derivatives exist).

To implement this we can use the approach from Chapters 3 and 5 of C. E. Rasmussen and C. K. I. Williams. Gaussian Processes for Machine Learning [46], as implemented in Scikit-learn's *GaussianProcessClassifier* class. This returns an estimate for $p(t)$ as well as optimised hyperparameters for the RBF kernel. To assess uncertainty, we are most interested in the role of finite data size and the distribution over possible trajectories of the

relative growth advantage, and so bootstrap the data vectors $\boldsymbol{X}$ and $\boldsymbol{y}$, then use the kernel hyperparameters optimised on real data to produce an ensemble of bootstrapped curves for $\pi(t)$. For the original data and bootstrapped curves, we can then also produce estimates of growth advantage from the equations above.

Phylogenetics

The alignment of consensus sequences with at least 95% coverage was used for phylogenetic reconstruction using RAxML-NG version 0.9.0 [47]. The resulting tree was rooted and fit to calendar time using TreeTime version 0.8.2 [35]. Ancestral sequence reconstruction was performed on the TreeTime divergence tree using IQ-TREE version 1.6.12 [48]. Visualisation used ggtree [49].

Nucleotide genetic diversity

We calculated the genetic diversity among consensus sequences for all sequenced samples collected in a particular week and with >50% coverage. Nucleotide genetic diversity was calculated using the π statistic, since this has been shown to be the least sensitive to differences in the number of sequences used in the analysis [50]. Mean pairwise genetic diversity across the genome is given by:

$$\pi = \frac{\sum_{i=1}^{L} D_i}{L}$$

Where L represents the length of the genome, and D_i the pairwise genetic diversity at locus i. This is calculated as:

$$D_i = \frac{\sum_{i \neq j} n_i n_j}{0.5\, N(N-1)}$$

Where n_i represents the number of alleles i observed at that locus, and N the number of samples with a consensus base call.

Within-lineage genetic diversity was calculated as above, but limiting only to sequences identified as belonging to B.1.177, B.1.1.7/Alpha or B.1.617.2/Delta, as well as for all of the other samples with a defined lineage but not B.1.177, B.1.1.7Alpha or B.1.617.2/Delta.

Funding Statement

The CIS is funded by the Department of Health and Social Care with in-kind support from the Welsh Government, the Department of Health on behalf of the Northern Ireland Government and the Scottish Government. COG-UK is supported by funding from the Medical Research Council (MRC) part of UK Research & Innovation (UKRI), the National Institute of Health Research (NIHR) [grant code: MC_PC_19027], and Genome Research Limited, operating as the Wellcome Sanger Institute. The authors acknowledge the support of the NHS Test and Trace Genomics Programme through sequencing of SARS-CoV-2 genomes analysed in this study. ASW is supported by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Healthcare Associated Infections and Antimicrobial Resistance at the University of Oxford in partnership with the UK Health Security Agency (UK HSA) (NIHR200915) and the NIHR Oxford Biomedical Research Centre, and is an NIHR Senior Investigator. KAL is supported by the Royal Society and the Wellcome Trust

(107652/Z/15/Z). The views expressed are those of the authors and not necessarily those of the National Health Service, NIHR, Department of Health, or UKHSA.

Ethics: The study received ethical approval from the South Central Berkshire B Research Ethics Committee (20/SC/0195)

References

1. WHO. Tracking SARS-CoV-2 variants. 2021 [cited 21 Sep 2021]. Available: https://www.who.int/activities/tracking-SARS-CoV-2-variants

2. Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, et al. SARS-CoV-2 variants, spike mutations and immune escape. Nat Rev Microbiol. 2021;19: 409–424.

3. The COVID-19 Genomics UK (COG-UK) consortium. An integrated national scale SARS-CoV-2 genomic surveillance network. The Lancet Microbe. 2020. doi:10.1016/S2666-5247(20)30054-9

4. Franceschi VB, Santos AS, Glaeser AB, Paiz JC, Caldana GD, Machado Lessa CL, et al. Population-based prevalence surveys during the Covid-19 pandemic: A systematic review. Rev Med Virol. 2021;31: e2200.

5. Kraemer MUG, Cummings DAT, Funk S, Reiner RC, Faria NR, Pybus OG, et al. Reconstruction and prediction of viral disease epidemics. Epidemiology & Infection. 2019;147. doi:10.1017/S0950268818002881

6. Pouwels KB, House T, Pritchard E, Robotham JV, Birrell PJ, Gelman A, et al. Community prevalence of SARS-CoV-2 in England from April to November, 2020: results from the ONS Coronavirus Infection Survey. Lancet Public Health. 2021;6: e30–e38.

7. Sah P, Fitzpatrick MC, Zimmer CF, Abdollahi E, Juden-Kelly L, Moghadas SM, et al. Asymptomatic SARS-CoV-2 infection: A systematic review and meta-analysis. Proc Natl Acad Sci U S A. 2021;118. doi:10.1073/pnas.2109229118

8. UKHSA. Surge testing for new coronavirus (COVID-19) variants. 2021 [cited 21 Sep 2021]. Available: https://www.gov.uk/guidance/surge-testing-for-new-coronavirus-covid-19-variants

9. Eales O, Page AJ, Tang SN, Walters CE, Wang H, Haw D, et al. SARS-CoV-2 lineage dynamics in England from January to March 2021 inferred from representative community samples. medRxiv. 2021; 2021.05.08.21256867.

10. Riley S, Atchison C, Ashby D, Donnelly CA, Barclay W, Cooke GS, et al. REal-time Assessment of Community Transmission (REACT) of SARS-CoV-2 virus: Study protocol. Wellcome Open Research. 2021;5: 200.

11. Office for National Statistics. Coronavirus (COVID-19) Infection Survey QMI. 2021 [cited 22 Nov 2021]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/methodologies/coronaviruscovid19infectionsurveyqmi

12. Kidd M, Richter A, Best A, Cumley N, Mirza J, Percival B, et al. S-Variant SARS-CoV-2 Lineage B1.1.7 Is Associated With Significantly Higher Viral Load in Samples Tested by

TaqPath Polymerase Chain Reaction. J Infect Dis. 2021;223: 1666–1670.

13. Walker AS, Vihta K-D, Gethings O, Pritchard E, Jones J, House T, et al. Increased infections, but not viral burden, with a new SARS-CoV-2 variant. medRxiv. 2021; 2021.01.13.21249721.

14. Volz E, Mishra S, Chand M, Barrett JC, Johnson R, Geidelberg L, et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. Nature. 2021;593: 266–269.

15. Davies NG, Jarvis CI, Edmunds WJ, Jewell NP, Diaz-Ordaz K, Keogh RH. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. Nature. 2021;593: 270–274.

16. Public Health England. Variants of Concern Technical Briefing 10. 2021 [cited 17 Nov 2021]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/984274/Variants_of_Concern_VOC_Technical_Briefing_10_England.pdf

17. Office for National Statistics. Coronavirus (COVID-19) Infection Survey, UK - Office for National Statistics. Office for National Statistics; 2021 [cited 22 Nov 2021]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/16july2021

18. UKHSA. Variants: distribution of case data, 17 September 2021. 2021 [cited 17 Nov 2021]. Available: https://www.gov.uk/government/publications/covid-19-variants-genomically-confirmed-case-numbers/variants-distribution-of-case-data-17-september-2021

19. COG-UK. COG-UK/Mutation Explorer. 2021 [cited 22 Oct 2021]. Available: http://sars2.cvr.gla.ac.uk/cog-uk/

20. Pritchard E, Jones J, Vihta K, Stoesser N, Matthews PC, Eyre DW, et al. Monitoring populations at increased risk for SARS-CoV-2 infection in the community. medRxiv. 2021; 2021.09.02.21263017.

21. Rambaut A, Loman N, Pybus O, Barclay W, Barrett J, Carabelli A, et al. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. 18 Dec 2020 [cited 21 Sep 2021]. Available: https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563

22. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nature Microbiology. 2020;5: 1403–1407.

23. O'Toole Á, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, et al. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evol. 2021;7. doi:10.1093/ve/veab064

24. Office for National Statistics. Coronavirus (COVID-19) Infection Survey: technical data - Office for National Statistics. 2021 [cited 22 Sep 2021]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/covid19infectionsurveytechnicaldata/2021

25. Lemey P, Ruktanonchai N, Hong SL, Colizza V, Poletto C, Van den Broeck F, et al. Untangling introductions and persistence in COVID-19 resurgence in Europe. Nature.

2021;595: 713–717.

26. Hodcroft EB, Zuber M, Nadeau S, Vaughan TG, Crawford KHD, Althaus CL, et al. Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. Nature. 2021;595: 707–712.

27. Vöhringer HS, Sanderson T, Sinnott M, De Maio N, Nguyen T, Goater R, et al. Genomic reconstruction of the SARS-CoV-2 epidemic in England. Nature. 2021; 1–11.

28. Kraemer MUG, Hill V, Ruis C, Dellicour S, Bajaj S, McCrone JT, et al. Spatiotemporal invasion dynamics of SARS-CoV-2 lineage B.1.1.7 emergence. Science. 2021;373: 889–895.

29. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. Science. 2021;372. doi:10.1126/science.abg3055

30. Jones TC, Biele G, Mühlemann B, Veith T, Schneider J, Beheim-Schwarzbach J, et al. Estimating infectiousness throughout SARS-CoV-2 infection course. Science. 2021 [cited 3 Nov 2021]. doi:10.1126/science.abi5273

31. Sonabend R, Whittkles LK, Imai N, Perez-Guzman PN, Knock ES, Rawson T, et al. Non-pharmaceutical interventions, vaccination, and the SARS-CoV-2 delta variant in England: a mathematical modelling study. Lancet. 2021 [cited 17 Nov 2021]. doi:10.1016/S0140-6736(21)02276-5

32. UKHSA. SARS-CoV-2 variants of concern and variants under investigation in England Technical briefing 27. 2021 [cited 17 Nov 2021]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1029715/technical-briefing-27.pdf

33. Office for National Statistics. Coronavirus (COVID-19) Infection Survey, UK - Office for National Statistics. Office for National Statistics; 2021 [cited 22 Nov 2021]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/19november2021

34. Thomson EC, Rosen LE, Shepherd JG, Spreafico R, da Silva Filipe A, Wojcechowskyj JA, et al. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. Cell. 2021;184: 1171–1187.e20.

35. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis. Virus Evol. 2018;4. doi:10.1093/ve/vex042

36. Public Health England. Variants of Concern Technical Briefing 6. 2021 [cited 17 Nov 2021]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/961299/Variants_of_Concern_VOC_Technical_Briefing_6_England-1.pdf

37. Elliott P, Haw D, Wang H, Eales O, Walters CE, Ainslie KEC, et al. Exponential growth, high prevalence of SARS-CoV-2, and vaccine effectiveness associated with the Delta variant. Science. 2021; eabl9551.

38. Viana R, Moyo S, Amoako DG, Tegally H, Scheepers C, Althaus CL, et al. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. medRxiv. 2021; 2021.12.19.21268028.

39. du Plessis L, McCrone JT, Zarebski AE, Hill V, Ruis C, Gutierrez B, et al. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. Science. 2021;371: 708–712.

40. Wilkinson E, Giovanetti M, Tegally H, San JE, Lessells R, Cuadros D, et al. A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa. Science. 2021;374: 423–431.

41. COG-UK. COG-UK publication. 2020 [cited 1 Nov 2021]. Available: https://www.protocols.io/workspaces/coguk/publication

42. Lythgoe KA, Hall M, Ferretti L, de Cesare M, MacIntyre-Cockett G, Trebes A, et al. SARS-CoV-2 within-host diversity and transmission. Science. 2021;372. doi:10.1126/science.abg0821

43. Bonsall D, Golubchik T, de Cesare M, Limbada M, Kosloff B, MacIntyre-Cockett G, et al. A Comprehensive Genomics Solution for HIV Surveillance and Clinical Monitoring in Low-Income Settings. J Clin Microbiol. 2020;58. doi:10.1128/JCM.00382-20

44. Wymant C, Blanquart F, Golubchik T, Gall A, Bakker M, Bezemer D, et al. Easy and accurate reconstruction of whole HIV genomes from short-read sequence data with shiver. Virus Evol. 2018;4: vey007.

45. Golubchik T, Lythgoe KA, Hall M, Ferretti L, Fryer HR, MacIntyre-Cockett G, et al. Early analysis of a potential link between viral load and the N501Y mutation in the SARS-COV-2 spike protein. medRxiv. 2021; 2021.01.12.20249080.

46. Rasmussen CE, Williams CKI. Gaussian Processes for Machine Learning. 2005. doi:10.7551/mitpress/3206.001.0001

47. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019;35: 4453–4455.

48. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 2014;32: 268–274.

49. Yu G, Smith DK, Tsan-Yuk ZHL. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution. 2017;8: 28–36.

50. Zhao L, Illingworth CJR. Measurements of intrahost viral diversity require an unbiased diversity metric. Virus Evol. 2019;5: vey041.

51. COG-UK. Public Data & Analysis. 12 Jan 2021 [cited 8 Nov 2021]. Available: https://www.cogconsortium.uk/tools-analysis/public-data-analysis-2/

Supplementary Tables and Figures

Table S1. S-gene target failure by lineage

Lineage	SGTF	Non-SGTF	Percentage SGTF*
Other	8	576	1.4 (0.6, 2.7)
B.1.177	3	1377	0.2 (0.1, 0.6)
B.1.1.7/Alpha	3424	36	99.0 (98.6, 99.3)
B.1.351/Beta	0	6	0.0 (0, 45.9)
P.1/Gamma	0	1	0.0 (0, 97.5)
B.1.617.2/Delta	14	10970	0.1 (0.1, 0.2)
B.1.525/Eta	3	0	100 (29.2, 100)
B.1.1.318	0	1	0.0 (0, 97.5)
A.23.1	0	4	0.0 (0, 60.2)
B.1.258	47	15	75.8 (63.5, 85.8)

*Values in parentheses represent 95% confidence intervals using the Clopper-Pearson method

Table S2. Sensitivity of SGTF for B.1.1.7/Alpha and non-SGTF for B.1.617.2/Delta by month of sampling (new)

Month	SGTF count	SGTF + B.1.1.7	Percentage SGTF that are B.1.1.7 *	Non-SGTF count	Non-SGTF + B.1.617.2	Percentage Non-SGTF that are B.1.617.2*
Apr 2020	0	0		2	0	0 (0, 84.2)
May 2020	0	0		22	0	0 (0, 15.4)
Jun 2020	0	0		2	0	0 (0, 84.2)
Jul 2020	0	0		1	0	0 (0, 97.5)
Aug 2020	4	0	0 (0, 60.2)	12	0	0 (0, 26.5)
Sep 2020	4	0	0 (0, 60.2)	148	0	0 (0, 2.5)
Oct 2020	21	0	0 (0, 16.1)	580	0	0 (0, 0.6)
Nov 2020	84	73	86.9 (77.9, 93.2)	420	0	0.(0, 0.9)
Dec 2020	864	854	98.8 (97.9, 99.4)	491	0	0 (0, 0.7)
Jan 2021	1534	1525	99.4 (98.9, 99.7)	307	0	0 (0, 1.2)
Feb 2021	473	472	99.8 (98.8, 100.0)	17	0	0 (0, 19.5)
Mar 2021	303	302	99.7 (98.2, 100.0)	8	0	0, (0, 36.9)
Apr 2021	124	124	100 (97.1, 100)	3	1	33.3 (3.2, 66.0)
May 2021	42	42	100 (91.6, 100)	44	44	100 (92.0, 100)
Jun 2021	24	24	100 (85.8, 100)	445	443	99.6 (98.4, 99.9)
Jul 2021	4	3	75 (22.2, 97.1)	1920	1919	99.9 (99.7, 100)
Aug 2021	9	4	44.4 (15.2, 77.1)	1949	1949	100 (99.8, 100)
Sep 2021	1	0	0 (0, 97.5)	2455	2454	100 (99.8, 100)
Oct 2021	4	1	25.0 (2.5, 58.9)	2791	2789	99.9 (99.7, 100)
Nov 2021	4	0	0 (0, 60.2)	1371	1371	100 (99.7, 100)

*Values in parentheses represent 95% confidence intervals using the Clopper-Pearson method

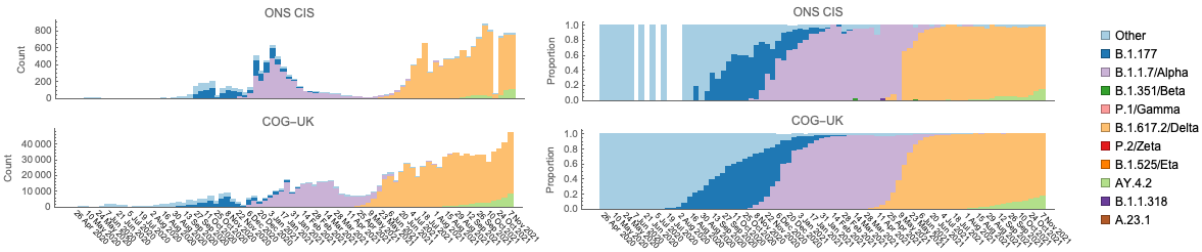

Figure S1. Comparison of lineage dynamics within the ONS CIS and among all COG-UK sequences. Top row: Number of sequenced samples and relative prevalence among ONS CIS sequences. Bottom row: Number of sequenced samples and relative incidence among COG-UK pillar 2 sequences [51], but excluding ONS sequences. All samples are grouped by the week in which they were collected, with the date giving the first day of the collection week (every other week labelled for clarity).

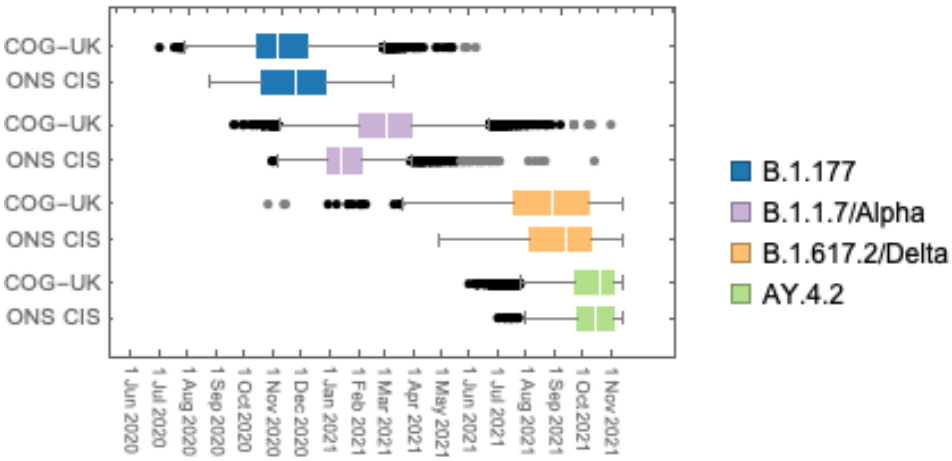

Figure S2. Comparison of sampling dates within the ONS CIS and among all COG-UK sequences. Box-whisker plot showing the distribution of sampling dates for the four sweeping lineages for COG-UK and ONS-CIS samples. COG-UK sampling dates represent all publicly available data for COG-UK pillar 2 sequences [51], but excluding ONS sequences. Some of the early reported sampling dates for COG-UK samples for each lineage may represent recording errors.

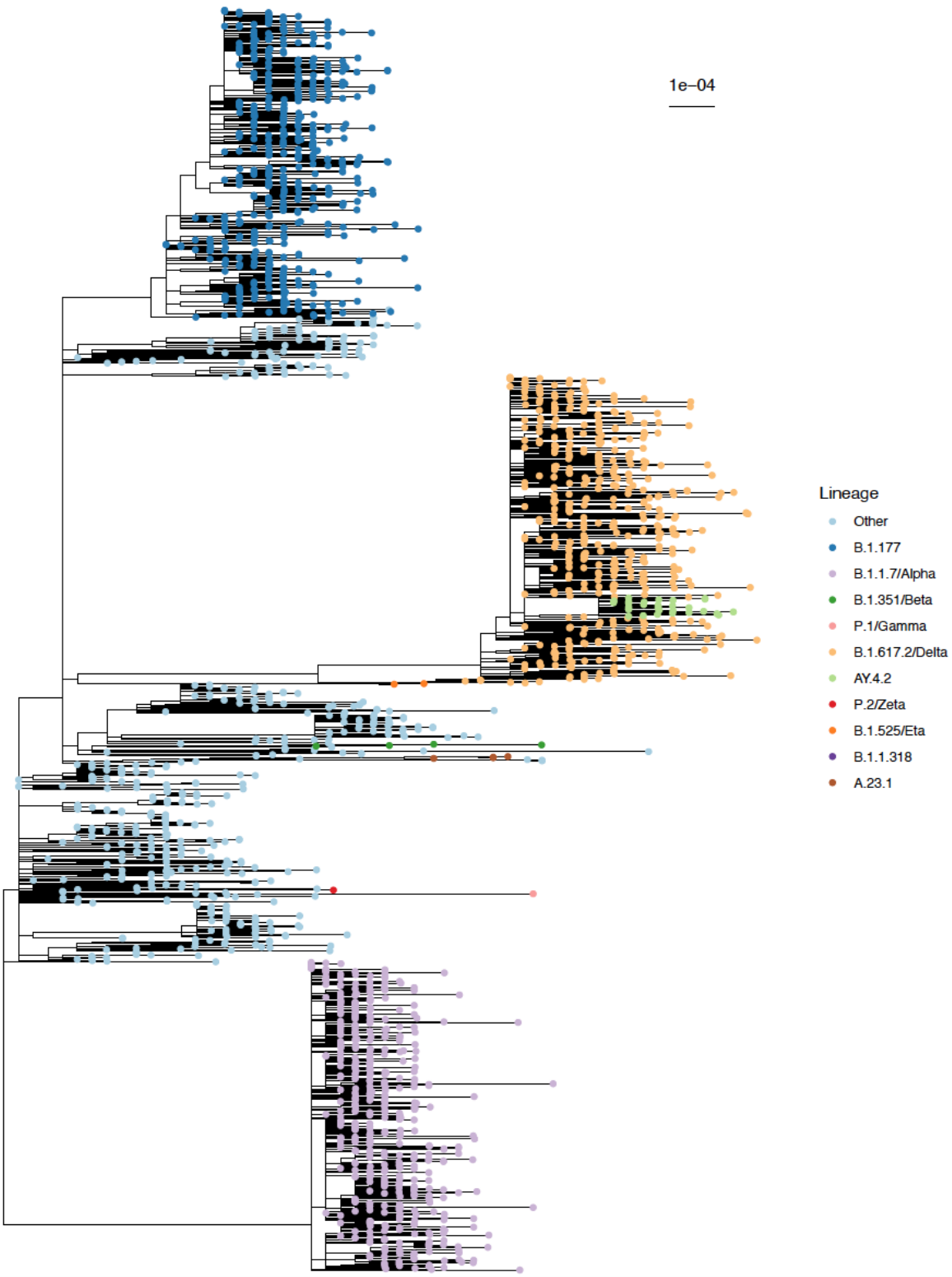

Figure S3. Maximum likelihood phylogeny. A maximum likelihood phylogeny of ONS sequences with over 95% genome coverage up to and including 13th November 2021 generated using RAxML-NG. B.1.177, B.1.1.7/Alpha, B.1.617.2/Delta (including AY.4.2) and 'Other' were each randomly subsampled to a maximum of 500 sequences.

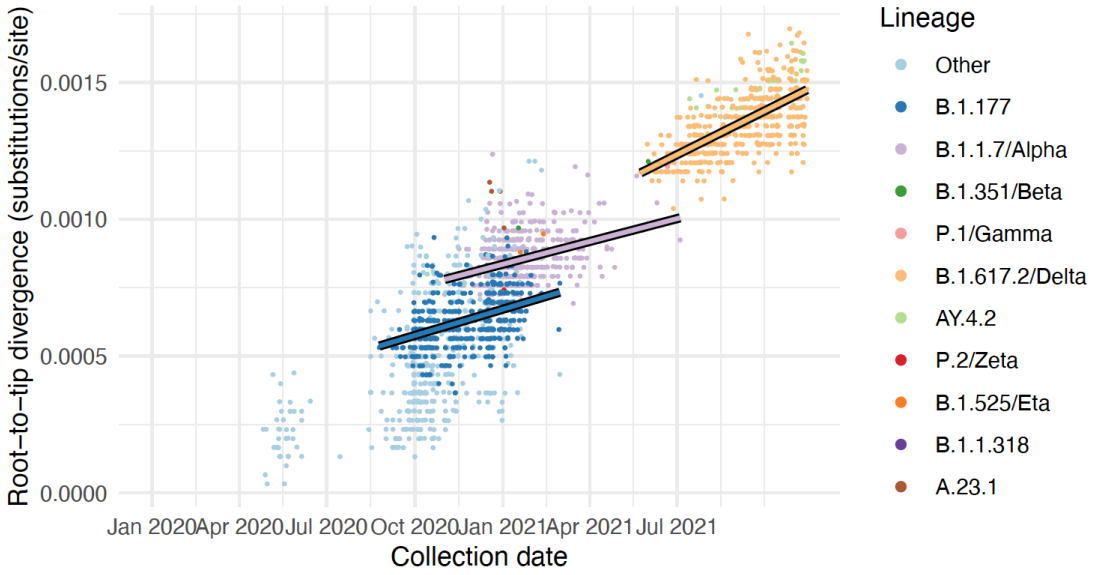

Figure S4. Divergence by lineage for all samples with over 95% genome coverage. A maximum likelihood phylogeny of ONS sequences with over 95% genome coverage up to and including 13th November 2021 (Fig S3) was generated using RAxML-NG, from which root-to-tip distances were calculated. Regressions were performed separately for B.1.177, B.1.1.7/Alpha, and B.1.617.2/Delta (including AY.4.2). B.1.177, B.1.1.7/Alpha, B.1.617.2/Delta (including AY.4.2) and 'Other' were each randomly subsampled to a maximum of 500 sequences.

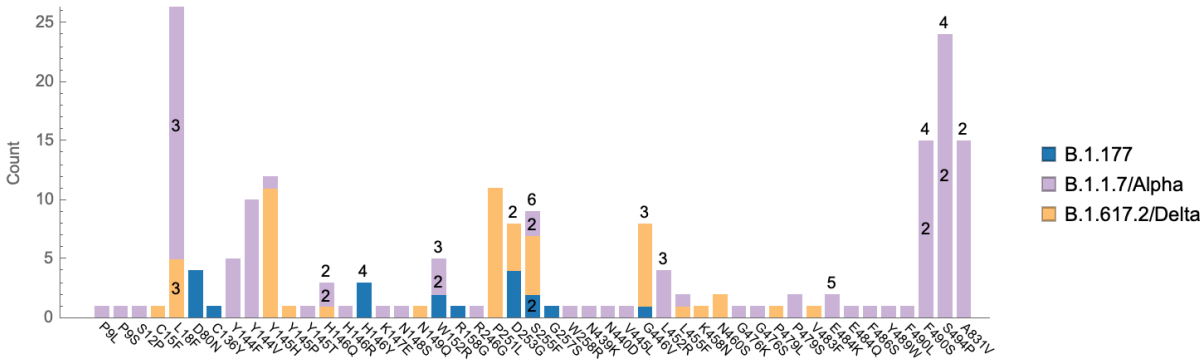

Figure S5. Mutations in Spike conferring antigenic change for the three most common lineages. The stacked bar chart indicates the number of samples in our dataset with the mutation, up until 17th July 2021. All antigenic mutations included in the COG-UK mutational explorer [19], were included. For clarity, mutations on a lineage that define that lineage are not considered. Numbers above the bars indicated the number of ancestral nucleotide changes on the phylogeny of all ONS CIS sequences with >95% coverage, if greater than one, whereas numbers superimposed onto the bars indicate the number of ancestral nodes associated with that lineage. The y-axis has been truncated at 26 for clarity, but up until 17th July 2021 L18F was observed in 76 Alpha samples, and was associated with 14 ancestral nucleotide changes. Apparent discrepancies occur because samples with <95% coverage were not included in the phylogeny (hence e.g. L455F only has one ancestral node on the phylogeny yet appears on both B.1.1.7/Alpha and B.1.617.2/Delta backgrounds on the bar chart), and because mutations may occur on lineages other than B.1.177, B.1.1.7/Alpha or B.1.617.2/Delta (e.g. E484K). The samples with Y145H on a B.1.617.2/Delta background are sublineage AY.4.2.

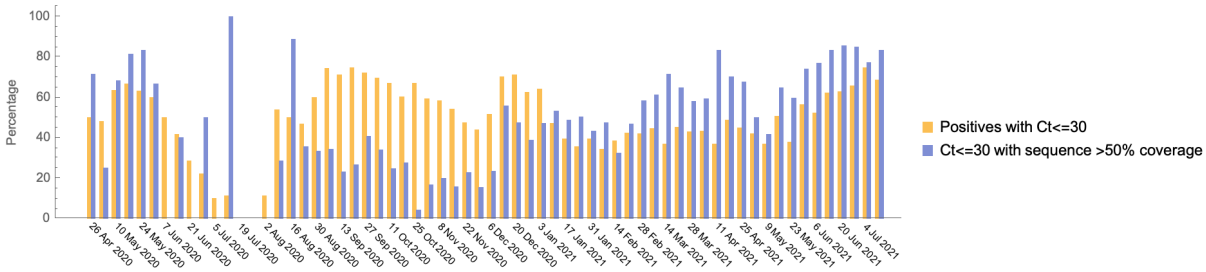

Figure S6: Sample proportions by collection week. The figure shows the percentage of ONS CIS RT PCR positive samples with Ct <=30 (yellow), and the percentage of Ct <=30 samples that have sequence with >50% genome coverage.

Figure S7: Plot of Log10 mapped reads versus Ct for all samples seqeunced using veSeq included in this study.

Appendix

The COVID-19 Genomics UK (COG-UK) consortium

June 2021 V.1

Funding acquisition, Leadership and supervision, Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, Software and analysis tools, and Visualisation:
Dr Samuel C Robson PhD [13, 84]

Funding acquisition, Leadership and supervision, Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, and Software and analysis tools:
Dr Thomas R Connor PhD [11, 74] and Prof Nicholas J Loman PhD [43]

Leadership and supervision, Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, Software and analysis tools, and Visualisation:
Dr Tanya Golubchik PhD [5]

Funding acquisition, Leadership and supervision, Metadata curation, Samples and logistics, Sequencing and analysis, and Visualisation:
Dr Rocio T Martinez Nunez PhD [46]

Funding acquisition, Leadership and supervision, Project administration, Samples and logistics, Sequencing and analysis, and Software and analysis tools:
Dr David Bonsall PhD [5]

Funding acquisition, Leadership and supervision, Project administration, Sequencing and analysis, Software and analysis tools, and Visualisation:
Prof Andrew Rambaut DPhil [104]

Funding acquisition, Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, and Software and analysis tools:
Dr Luke B Snell MSc, MBBS [12]

Leadership and supervision, Metadata curation, Project administration, Samples and logistics, Software and analysis tools, and Visualisation:
Rich Livett MSc [116]

Funding acquisition, Leadership and supervision, Metadata curation, Project administration, and Samples and logistics:
Dr Catherine Ludden PhD [20, 70]

Funding acquisition, Leadership and supervision, Metadata curation, Samples and logistics, and Sequencing and analysis:
Dr Sally Corden PhD [74] and Dr Eleni Nastouli FRCPath [96, 95, 30]

Funding acquisition, Leadership and supervision, Metadata curation, Sequencing and analysis, and Software and analysis tools:

Dr Gaia Nebbia PhD, FRCPath [12]

Funding acquisition, Leadership and supervision, Project administration, Samples and logistics, and Sequencing and analysis:
Ian Johnston BSc [116]

Leadership and supervision, Metadata curation, Project administration, Samples and logistics, and Sequencing and analysis:
Prof Katrina Lythgoe PhD [5], Dr M. Estee Torok FRCP [19, 20] and Prof Ian G Goodfellow PhD [24]
Leadership and supervision, Metadata curation, Project administration, Samples and logistics, and Visualisation:
Dr Jacqui A Prieto PhD [97, 82] and Dr Kordo Saeed MD, FRCPath [97, 83]

Leadership and supervision, Metadata curation, Project administration, Sequencing and analysis, and Software and analysis tools:
Dr David K Jackson PhD [116]

Leadership and supervision, Metadata curation, Samples and logistics, Sequencing and analysis, and Visualisation:
Dr Catherine Houlihan PhD [96, 94]

Leadership and supervision, Metadata curation, Sequencing and analysis, Software and analysis tools, and Visualisation:
Dr Dan Frampton PhD [94, 95]

Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, and Software and analysis tools:
Dr William L Hamilton PhD [19] and Dr Adam A Witney PhD [41]

Funding acquisition, Samples and logistics, Sequencing and analysis, and Visualisation:
Dr Giselda Bucca PhD [101]

Funding acquisition, Leadership and supervision, Metadata curation, and Project administration:
Dr Cassie F Pope PhD[40, 41]

Funding acquisition, Leadership and supervision, Metadata curation, and Samples and logistics:
Dr Catherine Moore PhD [74]

Funding acquisition, Leadership and supervision, Metadata curation, and Sequencing and analysis:
Prof Emma C Thomson PhD, FRCP [53]

Funding acquisition, Leadership and supervision, Project administration, and Samples and logistics:
Dr Ewan M Harrison PhD [116, 102]

Funding acquisition, Leadership and supervision, Sequencing and analysis, and Visualisation:
Prof Colin P Smith PhD [101]

Leadership and supervision, Metadata curation, Project administration, and Sequencing and analysis:
Fiona Rogan BSc [77]

Leadership and supervision, Metadata curation, Project administration, and Samples and logistics:
Shaun M Beckwith MSc [6], Abigail Murray Degree [6], Dawn Singleton HNC [6], Dr Kirstine Eastick PhD, FRCPath [37], Dr Liz A Sheridan PhD [98], Paul Randell MSc, PgD [99], Dr Leigh M Jackson PhD [105], Dr Cristina V Ariani PhD [116] and Dr Sónia Gonçalves PhD [116]

Leadership and supervision, Metadata curation, Samples and logistics, and Sequencing and analysis:
Dr Derek J Fairley PhD [3, 77], Prof Matthew W Loose PhD [18] and Joanne Watkins MSc [74]

Leadership and supervision, Metadata curation, Samples and logistics, and Visualisation:
Dr Samuel Moses MD [25, 106]

Leadership and supervision, Metadata curation, Sequencing and analysis, and Software and analysis tools:
Dr Sam Nicholls PhD [43], Dr Matthew Bull PhD [74] and Dr Roberto Amato PhD [116]

Leadership and supervision, Project administration, Samples and logistics, and Sequencing and analysis:
Prof Darren L Smith PhD [36, 65, 66]

Leadership and supervision, Sequencing and analysis, Software and analysis tools, and Visualisation:
Prof David M Aanensen PhD [14, 116] and Dr Jeffrey C Barrett PhD [116]

Metadata curation, Project administration, Samples and logistics, and Sequencing and analysis:
Dr Dinesh Aggarwal MRCP[20, 116, 70], Dr James G Shepherd MBCHB, MRCP [53], Dr Martin D Curran PhD [71] and Dr Surendra Parmar PhD [71]

Metadata curation, Project administration, Sequencing and analysis, and Software and analysis tools:
Dr Matthew D Parker PhD [109]

Metadata curation, Samples and logistics, Sequencing and analysis, and Software and analysis tools:
Dr Catryn Williams PhD [74]

Metadata curation, Samples and logistics, Sequencing and analysis, and Visualisation:

Dr Sharon Glaysher PhD [68]

Metadata curation, Sequencing and analysis, Software and analysis tools, and Visualisation:
Dr Anthony P Underwood PhD [14, 116], Dr Matthew Bashton PhD [36, 65], Dr Nicole Pacchiarini PhD [74], Dr Katie F Loveson PhD [84] and Matthew Byott MSc [95, 96]

Project administration, Sequencing and analysis, Software and analysis tools, and Visualisation:
Dr Alessandro M Carabelli PhD [20]

Funding acquisition, Leadership and supervision, and Metadata curation:
Dr Kate E Templeton PhD [56, 104]

Funding acquisition, Leadership and supervision, and Project administration:
Dr Thushan I de Silva PhD [109], Dr Dennis Wang PhD [109], Dr Cordelia F Langford PhD [116] and John Sillitoe BEng [116]

Funding acquisition, Leadership and supervision, and Samples and logistics:
Prof Rory N Gunson PhD, FRCPath [55]

Funding acquisition, Leadership and supervision, and Sequencing and analysis:
Dr Simon Cottrell PhD [74], Dr Justin O'Grady PhD [75, 103] and Prof Dominic Kwiatkowski PhD [116, 108]

Leadership and supervision, Metadata curation, and Project administration:
Dr Patrick J Lillie PhD, FRCP [37]

Leadership and supervision, Metadata curation, and Samples and logistics:
Dr Nicholas Cortes MBCHB [33], Dr Nathan Moore MBCHB [33], Dr Claire Thomas DPhil [33], Phillipa J Burns MSc, DipRCPath [37], Dr Tabitha W Mahungu FRCPath [80] and Steven Liggett BSc [86]

Leadership and supervision, Metadata curation, and Sequencing and analysis:
Angela H Beckett MSc [13, 81] and Prof Matthew TG Holden PhD [73]

Leadership and supervision, Project administration, and Samples and logistics:
Dr Lisa J Levett PhD [34], Dr Husam Osman PhD [70, 35] and Dr Mohammed O Hassan-Ibrahim PhD, FRCPath [99]

Leadership and supervision, Project administration, and Sequencing and analysis:
Dr David A Simpson PhD [77]

Leadership and supervision, Samples and logistics, and Sequencing and analysis:
Dr Meera Chand PhD [72], Prof Ravi K Gupta PhD [102], Prof Alistair C Darby PhD [107] and Prof Steve Paterson PhD [107]

Leadership and supervision, Sequencing and analysis, and Software and analysis tools:
Prof Oliver G Pybus DPhil [23], Dr Erik M Volz PhD [39], Prof Daniela de Angelis PhD [52], Prof

David L Robertson PhD [53], Dr Andrew J Page PhD [75] and Dr Inigo Martincorena PhD [116]

Leadership and supervision, Sequencing and analysis, and Visualisation:
Dr Louise Aigrain PhD [116] and Dr Andrew R Bassett PhD [116]

Metadata curation, Project administration, and Samples and logistics:
Dr Nick Wong DPhil, MRCP, FRCPath [50], Dr Yusri Taha MD, PhD [89], Michelle J Erkiert BA [99] and Dr Michael H Spencer Chapman MBBS [116, 102]

Metadata curation, Project administration, and Sequencing and analysis:
Dr Rebecca Dewar PhD [56] and Martin P McHugh MSc [56, 111]

Metadata curation, Project administration, and Software and analysis tools:
Siddharth Mookerjee MPH [38, 57]

Metadata curation, Project administration, and Visualisation:
Stephen Aplin [97], Matthew Harvey [97], Thea Sass [97], Dr Helen Umpleby FRCP [97] and Helen Wheeler [97]

Metadata curation, Samples and logistics, and Sequencing and analysis:
Dr James P McKenna PhD [3], Dr Ben Warne MRCP [9], Joshua F Taylor MSc [22], Yasmin Chaudhry BSc [24], Rhys Izuagbe [24], Dr Aminu S Jahun PhD [24], Dr Gregory R Young PhD [36, 65], Dr Claire McMurray PhD [43], Dr Clare M McCann PhD [65, 66], Dr Andrew Nelson PhD [65, 66] and Scott Elliott [68]

Metadata curation, Samples and logistics, and Visualisation:
Hannah Lowe MSc [25]

Metadata curation, Sequencing and analysis, and Software and analysis tools:
Dr Anna Price PhD [11], Matthew R Crown BSc [65], Dr Sara Rey PhD [74], Dr Sunando Roy PhD [96] and Dr Ben Temperton PhD [105]

Metadata curation, Sequencing and analysis, and Visualisation:
Dr Sharif Shaaban PhD [73] and Dr Andrew R Hesketh PhD [101]

Project administration, Samples and logistics, and Sequencing and analysis:
Dr Kenneth G Laing PhD[41], Dr Irene M Monahan PhD [41] and Dr Judith Heaney PhD [95, 96, 34]

Project administration, Samples and logistics, and Visualisation:
Dr Emanuela Pelosi FRCPath [97], Siona Silviera MSc [97] and Dr Eleri Wilson-Davies MD, FRCPath [97]

Samples and logistics, Software and analysis tools, and Visualisation:
Dr Helen Fryer PhD [5]

Sequencing and analysis, Software and analysis tools, and Visualization:
Dr Helen Adams PhD [4], Dr Louis du Plessis PhD [23], Dr Rob Johnson PhD [39], Dr William T

Harvey PhD [53, 42], Dr Joseph Hughes PhD [53], Dr Richard J Orton PhD [53], Dr Lewis G Spurgin PhD [59], Dr Yann Bourgeois PhD [81], Dr Chris Ruis PhD [102], Áine O'Toole MSc [104], Marina Gourtovaia MSc [116] and Dr Theo Sanderson PhD [116]

Funding acquisition, and Leadership and supervision:
Dr Christophe Fraser PhD [5], Dr Jonathan Edgeworth PhD, FRCPath [12], Prof Judith Breuer MD [96, 29], Dr Stephen L Michell PhD [105] and Prof John A Todd PhD [115]

Funding acquisition, and Project administration:
Michaela John BSc [10] and Dr David Buck PhD [115]

Leadership and supervision, and Metadata curation:
Dr Kavitha Gajee MBBS, FRCPath [37] and Dr Gemma L Kay PhD [75]

Leadership and supervision, and Project administration:
Prof Sharon J Peacock PhD [20, 70] and David Heyburn [74]

Leadership and supervision, and Samples and logistics:
Katie Kitchman BSc [37], Prof Alan McNally PhD [43, 93], David T Pritchard MSc, CSci [50], Dr Samir Dervisevic FRCPath [58], Dr Peter Muir PhD [70], Dr Esther Robinson PhD [70, 35], Dr Barry B Vipond PhD [70], Newara A Ramadan MSc, CSci, FIBMS [78], Dr Christopher Jeanes MBBS [90], Danni Weldon BSc [116], Jana Catalan MSc [118] and Neil Jones MSc [118]

Leadership and supervision, and Sequencing and analysis:
Dr Ana da Silva Filipe PhD [53], Dr Chris Williams MBBS [74], Marc Fuchs BSc [77], Dr Julia Miskelly PhD [77], Dr Aaron R Jeffries PhD [105], Karen Oliver BSc [116] and Dr Naomi R Park PhD [116]

Metadata curation, and Samples and logistics:
Amy Ash BSc [1], Cherian Koshy MSc, CSci, FIBMS [1], Magdalena Barrow [7], Dr Sarah L Buchan PhD [7], Dr Anna Mantzouratou PhD [7], Dr Gemma Clark PhD [15], Dr Christopher W Holmes PhD [16], Sharon Campbell MSc [17], Thomas Davis MSc [21], Ngee Keong Tan MSc [22], Dr Julianne R Brown PhD [29], Dr Kathryn A Harris PhD [29, 2], Stephen P Kidd MSc [33], Dr Paul R Grant PhD [34], Dr Li Xu-McCrae PhD [35], Dr Alison Cox PhD [38, 63], Pinglawathee Madona [38, 63], Dr Marcus Pond PhD [38, 63], Dr Paul A Randell MBBCh [38, 63], Karen T Withell FIBMS [48], Cheryl Williams MSc [51], Dr Clive Graham MD [60], Rebecca Denton-Smith BSc [62], Emma Swindells BSc [62], Robyn Turnbull BSc [62], Dr Tim J Sloan PhD [67], Dr Andrew Bosworth PhD [70, 35], Stephanie Hutchings [70], Hannah M Pymont MSc [70], Dr Anna Casey PhD [76], Dr Liz Ratcliffe PhD [76], Dr Christopher R Jones PhD [79, 105], Dr Bridget A Knight PhD [79, 105], Dr Tanzina Haque PhD, FRCPath [80], Dr Jennifer Hart MRCP [80], Dr Dianne Irish-Tavares FRCPath [80], Eric Witele MSc [80], Craig Mower BA [86], Louisa K Watson DipHE [86], Jennifer Collins BSc [89], Gary Eltringham BSc [89], Dorian Crudgington [98], Ben Macklin [98], Prof Miren Iturriza-Gomara PhD [107], Dr Anita O Lucaci PhD [107] and Dr Patrick C McClure PhD [113]

Metadata curation, and Sequencing and analysis:
Matthew Carlile BSc [18], Dr Nadine Holmes PhD [18], Dr Christopher Moore PhD [18], Dr Nathaniel Storey PhD [29], Dr Stefan Rooke PhD [73], Dr Gonzalo Yebra PhD [73], Dr Noel Craine DPhil [74], Malorie Perry MSc [74], Dr Nabil-Fareed Alikhan PhD [75], Dr Stephen Bridgett PhD [77],

Kate F Cook MScR [84], Christopher Fearn MSc [84], Dr Salman Goudarzi PhD [84], Prof Ronan A Lyons MD [88], Dr Thomas Williams MD [104], Dr Sam T Haldenby PhD [107], Jillian Durham BSc [116] and Dr Steven Leonard PhD [116]

Metadata curation, and Software and analysis tools:
Robert M Davies MA (Cantab) [116]

Project administration, and Samples and logistics:
Dr Rahul Batra MD [12], Beth Blane BSc [20], Dr Moira J Spyer PhD [30, 95, 96], Perminder Smith MSc [32, 112], Mehmet Yavus [85, 109], Dr Rachel J Williams PhD [96], Dr Adhyana IK Mahanama MD [97], Dr Buddhini Samaraweera MD [97], Sophia T Girgis MSc [102], Samantha E Hansford CSci [109], Dr Angie Green PhD [115], Dr Charlotte Beaver PhD [116], Katherine L Bellis [116, 102], Matthew J Dorman [116], Sally Kay [116], Liam Prestwood [116] and Dr Shavanthi Rajatileka PhD [116]

Project administration, and Sequencing and analysis:
Dr Joshua Quick PhD [43]

Project administration, and Software and analysis tools:
Radoslaw Poplawski BSc [43]

Samples and logistics, and Sequencing and analysis:
Dr Nicola Reynolds PhD [8], Andrew Mack MPhil [11], Dr Arthur Morriss PhD [11], Thomas Whalley BSc [11], Bindi Patel BSc [12], Dr Iliana Georgana PhD [24], Dr Myra Hosmillo PhD [24], Malte L Pinckert MPhil [24], Dr Joanne Stockton PhD [43], Dr John H Henderson PhD [65], Amy Hollis HND [65], Dr William Stanley PhD [65], Dr Wen C Yew PhD [65], Dr Richard Myers PhD [72], Dr Alicia Thornton PhD [72], Alexander Adams BSc [74], Tara Annett BSc [74], Dr Hibo Asad PhD [74], Alec Birchley MSc [74], Jason Coombes BSc [74], Johnathan M Evans MSc [74], Laia Fina [74], Bree Gatica-Wilcox MPhil [74], Lauren Gilbert [74], Lee Graham BSc [74], Jessica Hey BSc [74], Ember Hilvers MPH [74], Sophie Jones MSc [74], Hannah Jones [74], Sara Kumziene-Summerhayes MSc [74], Dr Caoimhe McKerr PhD [74], Jessica Powell BSc [74], Georgia Pugh [74], Sarah Taylor [74], Alexander J Trotter MRes [75], Charlotte A Williams BSc [96], Leanne M Kermack MSc [102], Benjamin H Foulkes MSc [109], Marta Gallis MSc [109], Hailey R Hornsby MSc [109], Stavroula F Louka MSc [109], Dr Manoj Pohare PhD [109], Paige Wolverson MSc [109], Peijun Zhang MSc [109], George MacIntyre-Cockett BSc [115], Amy Trebes MSc [115], Dr Robin J Moll PhD [116], Lynne Ferguson MSc [117], Dr Emily J Goldstein PhD [117], Dr Alasdair Maclean PhD [117] and Dr Rachael Tomb PhD [117]

Samples and logistics, and Software and analysis tools:
Dr Igor Starinskij MSc, MRCP [53]

Sequencing and analysis, and Software and analysis tools:
Laura Thomson BSc [5], Joel Southgate MSc [11, 74], Dr Moritz UG Kraemer DPhil [23], Dr Jayna Raghwani PhD [23], Dr Alex E Zarebski PhD [23], Olivia Boyd MSc [39], Lily Geidelberg MSc [39], Dr Chris J Illingworth PhD [52], Dr Chris Jackson PhD [52], Dr David Pascall PhD [52], Dr Sreenu Vattipally PhD [53], Timothy M Freeman MPhil [109], Dr Sharon N Hsu PhD [109], Dr Benjamin B Lindsey MRCP [109], Dr Keith James PhD [116], Kevin Lewis [116], Gerry Tonkin-Hill [116] and Dr Jaime M Tovar-Corona PhD [116]

Sequencing and analysis, and Visualisation:
MacGregor Cox MSci [20]

Software and analysis tools, and Visualisation:
Dr Khalil Abudahab PhD [14, 116], Mirko Menegazzo [14], Ben EW Taylor MEng [14, 116], Dr Corin A Yeats PhD [14], Afrida Mukaddas BTech [53], Derek W Wright MSc [53], Dr Leonardo de Oliveira Martins PhD [75], Dr Rachel Colquhoun DPhil [104], Verity Hill [104], Dr Ben Jackson PhD [104], Dr JT McCrone PhD [104], Dr Nathan Medd PhD [104], Dr Emily Scher PhD [104] and Jon-Paul Keatley [116]

Leadership and supervision:
Dr Tanya Curran PhD [3], Dr Sian Morgan FRCPath [10], Prof Patrick Maxwell PhD [20], Prof Ken Smith PhD [20], Dr Sahar Eldirdiri MBBS, MSc, FRCPath [21], Anita Kenyon MSc [21], Prof Alison H Holmes MD [38, 57], Dr James R Price PhD [38, 57], Dr Tim Wyatt PhD [69], Dr Alison E Mather PhD [75], Dr Timofey Skvortsov PhD [77] and Prof John A Hartley PhD [96]

Metadata curation:
Prof Martyn Guest PhD [11], Dr Christine Kitchen PhD [11], Dr Ian Merrick PhD [11], Robert Munn BSc [11], Dr Beatrice Bertolusso Degree [33], Dr Jessica Lynch MBCHB [33], Dr Gabrielle Vernet MBBS [33], Stuart Kirk MSc [34], Dr Elizabeth Wastnedge MD [56], Dr Rachael Stanley PhD [58], Giles Idle [64], Dr Declan T Bradley PhD [69, 77], Dr Jennifer Poyner MD [79] and Matilde Mori BSc [110]

Project administration:
Owen Jones BSc [11], Victoria Wright BSc [18], Ellena Brooks MA [20], Carol M Churcher BSc [20], Mireille Fragakis HND [20], Dr Katerina Galai PhD [20, 70], Dr Andrew Jermy PhD [20], Sarah Judges BA [20], Georgina M McManus BSc [20], Kim S Smith [20], Dr Elaine Westwick PhD [20], Dr Stephen W Attwood PhD [23], Dr Frances Bolt PhD [38, 57], Dr Alisha Davies PhD [74], Elen De Lacy MPH [74], Fatima Downing [74], Sue Edwards [74], Lizzie Meadows MA [75], Sarah Jeremiah MSc [97], Dr Nikki Smith PhD [109] and Luke Foulser [116]

Samples and logistics:
Dr Themoula Charalampous PhD [12, 46], Amita Patel BSc [12], Dr Louise Berry PhD [15], Dr Tim Boswell PhD [15], Dr Vicki M Fleming PhD [15], Dr Hannah C Howson-Wells PhD [15], Dr Amelia Joseph PhD [15], Manjinder Khakh [15], Dr Michelle M Lister PhD [15], Paul W Bird MSc, MRes [16], Karlie Fallon [16], Thomas Helmer [16], Dr Claire L McMurray PhD [16], Mina Odedra BSc [16], Jessica Shaw BSc [16], Dr Julian W Tang PhD [16], Nicholas J Willford MSc [16], Victoria Blakey BSc [17], Dr Veena Raviprakash MD [17], Nicola Sheriff BSc [17], Lesley-Anne Williams BSc [17], Theresa Feltwell MSc [20], Dr Luke Bedford PhD [26], Dr James S Cargill PhD [27], Warwick Hughes MSc [27], Dr Jonathan Moore MD [28], Susanne Stonehouse BSc [28], Laura Atkinson MSc [29], Jack CD Lee MSc [29], Dr Divya Shah PhD [29], Adela Alcolea-Medina Clinical scientist [32, 112], Natasha Ohemeng-Kumi MSc [32, 112], John Ramble MSc [32, 112], Jasveen Sehmi MSc [32, 112], Dr Rebecca Williams BMBS [33], Wendy Chatterton MSc [34], Monika Pusok MSc [34], William Everson MSc [37], Anibolina Castigador IBMS HCPC [44], Emily Macnaughton FRCPath [44], Dr Kate El Bouzidi MRCP [45], Dr Temi Lampejo FRCPath [45], Dr Malur Sudhanva FRCPath [45], Cassie Breen BSc [47], Dr Graciela Sluga MD, MSc [48], Dr Shazaad SY Ahmad MSc [49, 70], Dr Ryan P George PhD [49], Dr Nicholas W Machin MSc [49, 70], Debbie Binns BSc [50], Victoria James BSc [50], Dr Rachel Blacow MBCHB [55], Dr Lindsay Coupland PhD [58], Dr Louise Smith

PhD [59], Dr Edward Barton MD [60], Debra Padgett BSc [60], Garren Scott BSc [60], Dr Aidan Cross MBCHB [61], Dr Mariyam Mirfenderesky FRCPath [61], Jane Greenaway MSc [62], Kevin Cole [64], Phillip Clarke [67], Nichola Duckworth [67], Sarah Walsh [67], Kelly Bicknell [68], Robert Impey MSc [68], Dr Sarah Wyllie PhD [68], Richard Hopes [70], Dr Chloe Bishop PhD [72], Dr Vicki Chalker PhD [72], Dr Ian Harrison PhD [72], Laura Gifford MSc [74], Dr Zoltan Molnar PhD [77], Dr Cressida Auckland FRCPath [79], Dr Cariad Evans PhD [85, 109], Dr Kate Johnson PhD [85, 109], Dr David G Partridge FRCP, FRCPath [85, 109], Dr Mohammad Raza PhD [85, 109], Paul Baker MD [86], Prof Stephen Bonner PhD [86], Sarah Essex [86], Leanne J Murray [86], Andrew I Lawton MSc [87], Dr Shirelle Burton-Fanning MD [89], Dr Brendan Al Payne MD [89], Dr Sheila Waugh MD [89], Andrea N Gomes MSc [91], Maimuna Kimuli MSc [91], Darren R Murray MSc [91], Paula Ashfield MSc [92], Dr Donald Dobie MBCHB [92], Dr Fiona Ashford PhD [93], Dr Angus Best PhD [93], Dr Liam Crawford PhD [93], Dr Nicola Cumley PhD [93], Dr Megan Mayhew PhD [93], Dr Oliver Megram PhD [93], Dr Jeremy Mirza PhD [93], Dr Emma Moles-Garcia PhD [93], Dr Benita Percival PhD [93], Megan Driscoll BSc [96], Leah Ensell BSc [96], Dr Helen L Lowe PhD [96], Laurentiu Maftei BSc [96], Matteo Mondani MSc [96], Nicola J Chaloner BSc [99], Benjamin J Cogger BSc [99], Lisa J Easton MSc [99], Hannah Huckson BSc [99], Jonathan Lewis MSc, PgD, FIBMS [99], Sarah Lowdon BSc [99], Cassandra S Malone MSc [99], Florence Munemo BSc [99], Manasa Mutingwende MSc [99], Roberto Nicodemi BSc [99], Olga Podplomyk FD [99], Thomas Somassa BSc [99], Dr Andrew Beggs PhD [100], Dr Alex Richter PhD [100], Claire Cormie [102], Joana Dias MSc [102], Sally Forrest BSc [102], Dr Ellen E Higginson PhD [102], Mailis Maes MPhil [102], Jamie Young BSc [102], Dr Rose K Davidson PhD [103], Kathryn A Jackson MSc [107], Dr Lance Turtle PhD, MRCP [107], Dr Alexander J Keeley MRCP [109], Prof Jonathan Ball PhD [113], Timothy Byaruhanga MSc [113], Dr Joseph G Chappell PhD [113], Jayasree Dey MSc [113], Jack D Hill MSc [113], Emily J Park MSc [113], Arezou Fanaie MSc [114], Rachel A Hilson MSc [114], Geraldine Yaze MSc [114] and Stephanie Lo [116]

Sequencing and analysis:
Safiah Afifi BSc [10], Robert Beer BSc [10], Joshua Maksimovic FD [10], Kathryn McCluggage Masters [10], Karla Spellman FD [10], Catherine Bresner BSc [11], William Fuller BSc [11], Dr Angela Marchbank BSc [11], Trudy Workman HNC [11], Dr Ekaterina Shelest PhD [13, 81], Dr Johnny Debebe PhD [18], Dr Fei Sang PhD [18], Dr Marina Escalera Zamudio PhD [23], Dr Sarah Francois PhD [23], Bernardo Gutierrez MSc [23], Dr Tetyana I Vasylyeva DPhil [23], Dr Flavia Flaviani PhD [31], Dr Manon Ragonnet-Cronin PhD [39], Dr Katherine L Smollett PhD [42], Alice Broos BSc [53], Daniel Mair BSc [53], Jenna Nichols BSc [53], Dr Kyriaki Nomikou PhD [53], Dr Lily Tong PhD [53], Ioulia Tsatsani MSc [53], Prof Sarah O'Brien PhD [54], Prof Steven Rushton PhD [54], Dr Roy Sanderson PhD [54], Dr Jon Perkins MBCHB [55], Seb Cotton MSc [56], Abbie Gallagher BSc [56], Dr Elias Allara MD, PhD [70, 102], Clare Pearson MSc [70, 102], Dr David Bibby PhD [72], Dr Gavin Dabrera PhD [72], Dr Nicholas Ellaby PhD [72], Dr Eileen Gallagher PhD [72], Dr Jonathan Hubb PhD [72], Dr Angie Lackenby PhD [72], Dr David Lee PhD [72], Nikos Manesis [72], Dr Tamyo Mbisa PhD [72], Dr Steven Platt PhD [72], Katherine A Twohig [72], Dr Mari Morgan PhD [74], Alp Aydin MSci [75], David J Baker BEng [75], Dr Ebenezer Foster-Nyarko PhD [75], Dr Sophie J Prosolek PhD [75], Steven Rudder [75], Chris Baxter BSc [77], Sílvia F Carvalho MSc [77], Dr Deborah Lavin PhD [77], Dr Arun Mariappan PhD [77], Dr Clara Radulescu PhD [77], Dr Aditi Singh PhD [77], Miao Tang MD [77], Helen Morcrette BSc [79], Nadua Bayzid BSc [96], Marius Cotic MSc [96], Dr Carlos E Balcazar PhD [104], Dr Michael D Gallagher PhD [104], Dr Daniel Maloney PhD [104], Thomas D Stanton BSc [104], Dr Kathleen A Williamson PhD [104], Dr Robin Manley PhD [105], Michelle L Michelsen BSc [105], Dr Christine M Sambles PhD [105], Dr David J Studholme PhD [105], Joanna Warwick-Dugdale BSc [105], Richard Eccles MSc [107], Matthew Gemmell MSc [107], Dr Richard

Gregory PhD [107], Dr Margaret Hughes PhD [107], Charlotte Nelson MSc [107], Dr Lucille Rainbow PhD [107], Dr Edith E Vamos PhD [107], Hermione J Webster BSc [107], Dr Mark Whitehead PhD [107], Claudia Wierzbicki BSc [107], Dr Adrienn Angyal PhD [109], Dr Luke R Green PhD [109], Dr Max Whiteley PhD [109], Emma Betteridge BSc [116], Dr Iraad F Bronner PhD [116], Ben W Farr BSc [116], Scott Goodwin MSc [116], Dr Stefanie V Lensing PhD [116], Shane A McCarthy [116, 102], Dr Michael A Quail PhD [116], Diana Rajan MSc [116], Dr Nicholas M Redshaw PhD [116], Carol Scott [116], Lesley Shirley MSc [116] and Scott AJ Thurston BSc [116]

Software and analysis tools:
Dr Will Rowe PhD[43], Amy Gaskin MSc [74], Dr Thanh Le-Viet PhD [75], James Bonfield BSc [116], Jennifier Liddle [116] and Andrew Whitwham BSc [116]

1 Barking, Havering and Redbridge University Hospitals NHS Trust, 2 Barts Health NHS Trust, 3 Belfast Health & Social Care Trust, 4 Betsi Cadwaladr University Health Board, 5 Big Data Institute, Nuffield Department of Medicine, University of Oxford, 6 Blackpool Teaching Hospitals NHS Foundation Trust, 7 Bournemouth University, 8 Cambridge Stem Cell Institute, University of Cambridge, 9 Cambridge University Hospitals NHS Foundation Trust, 10 Cardiff and Vale University Health Board, 11 Cardiff University, 12 Centre for Clinical Infection and Diagnostics Research, Department of Infectious Diseases, Guy's and St Thomas' NHS Foundation Trust, 13 Centre for Enzyme Innovation, University of Portsmouth, 14 Centre for Genomic Pathogen Surveillance, University of Oxford, 15 Clinical Microbiology Department, Queens Medical Centre, Nottingham University Hospitals NHS Trust, 16 Clinical Microbiology, University Hospitals of Leicester NHS Trust, 17 County Durham and Darlington NHS Foundation Trust, 18 Deep Seq, School of Life Sciences, Queens Medical Centre, University of Nottingham, 19 Department of Infectious Diseases and Microbiology, Cambridge University Hospitals NHS Foundation Trust, 20 Department of Medicine, University of Cambridge, 21 Department of Microbiology, Kettering General Hospital, 22 Department of Microbiology, South West London Pathology, 23 Department of Zoology, University of Oxford, 24 Division of Virology, Department of Pathology, University of Cambridge, 25 East Kent Hospitals University NHS Foundation Trust, 26 East Suffolk and North Essex NHS Foundation Trust, 27 East Sussex Healthcare NHS Trust, 28 Gateshead Health NHS Foundation Trust, 29 Great Ormond Street Hospital for Children NHS Foundation Trust, 30 Great Ormond Street Institute of Child Health (GOS ICH), University College London (UCL), 31 Guy's and St. Thomas' Biomedical Research Centre, 32 Guy's and St. Thomas' NHS Foundation Trust, 33 Hampshire Hospitals NHS Foundation Trust, 34 Health Services Laboratories, 35 Heartlands Hospital, Birmingham, 36 Hub for Biotechnology in the Built Environment, Northumbria University, 37 Hull University Teaching Hospitals NHS Trust, 38 Imperial College Healthcare NHS Trust, 39 Imperial College London, 40 Infection Care Group, St George's University Hospitals NHS Foundation Trust, 41 Institute for Infection and Immunity, St George's University of London, 42 Institute of Biodiversity, Animal Health & Comparative Medicine, 43 Institute of Microbiology and Infection, University of Birmingham, 44 Isle of Wight NHS Trust, 45 King's College Hospital NHS Foundation Trust, 46 King's College London, 47 Liverpool Clinical Laboratories, 48 Maidstone and Tunbridge Wells NHS Trust, 49 Manchester University NHS Foundation Trust, 50 Microbiology Department, Buckinghamshire Healthcare NHS Trust, 51 Microbiology, Royal Oldham Hospital, 52 MRC Biostatistics Unit, University of Cambridge, 53 MRC-University of Glasgow Centre for Virus Research, 54 Newcastle University, 55 NHS Greater Glasgow and Clyde, 56 NHS Lothian, 57 NIHR Health Protection Research Unit in HCAI and AMR, Imperial College London, 58 Norfolk and Norwich University Hospitals NHS Foundation Trust, 59 Norfolk County Council, 60 North Cumbria Integrated Care NHS Foundation Trust, 61 North Middlesex University Hospital NHS Trust, 62 North Tees and Hartlepool NHS Foundation Trust, 63 North West London Pathology, 64 Northumbria Healthcare NHS Foundation Trust, 65 Northumbria University, 66 NU-OMICS, Northumbria University, 67 Path Links, Northern Lincolnshire and Goole NHS Foundation Trust, 68 Portsmouth Hospitals University NHS Trust, 69 Public Health Agency, Northern Ireland, 70

Public Health England, 71 Public Health England, Cambridge, 72 Public Health England, Colindale, 73 Public Health Scotland, 74 Public Health Wales, 75 Quadram Institute Bioscience, 76 Queen Elizabeth Hospital, Birmingham, 77 Queen's University Belfast, 78 Royal Brompton and Harefield Hospitals, 79 Royal Devon and Exeter NHS Foundation Trust, 80 Royal Free London NHS Foundation Trust, 81 School of Biological Sciences, University of Portsmouth, 82 School of Health Sciences, University of Southampton, 83 School of Medicine, University of Southampton, 84 School of Pharmacy & Biomedical Sciences, University of Portsmouth, 85 Sheffield Teaching Hospitals NHS Foundation Trust, 86 South Tees Hospitals NHS Foundation Trust, 87 Southwest Pathology Services, 88 Swansea University, 89 The Newcastle upon Tyne Hospitals NHS Foundation Trust, 90 The Queen Elizabeth Hospital King's Lynn NHS Foundation Trust, 91 The Royal Marsden NHS Foundation Trust, 92 The Royal Wolverhampton NHS Trust, 93 Turnkey Laboratory, University of Birmingham, 94 University College London Division of Infection and Immunity, 95 University College London Hospital Advanced Pathogen Diagnostics Unit, 96 University College London Hospitals NHS Foundation Trust, 97 University Hospital Southampton NHS Foundation Trust, 98 University Hospitals Dorset NHS Foundation Trust, 99 University Hospitals Sussex NHS Foundation Trust, 100 University of Birmingham, 101 University of Brighton, 102 University of Cambridge, 103 University of East Anglia, 104 University of Edinburgh, 105 University of Exeter, 106 University of Kent, 107 University of Liverpool, 108 University of Oxford, 109 University of Sheffield, 110 University of Southampton, 111 University of St Andrews, 112 Viapath, Guy's and St Thomas' NHS Foundation Trust, and King's College Hospital NHS Foundation Trust, 113 Virology, School of Life Sciences, Queens Medical Centre, University of Nottingham, 114 Watford General Hospital, 115 Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, 116 Wellcome Sanger Institute, 117 West of Scotland Specialist Virology Centre, NHS Greater Glasgow and Clyde, 118 Whittington Health NHS Trust