SARS-CoV-2 lineage dynamics in England from September to November 2021: high diversity of Delta sub-lineages and increased transmissibility of AY.4.2

Oliver Eales[1,2], Andrew J. Page[3], Leonardo de Oliveira Martins[3], Haowei Wang[1,2], Barbara Bodinier[1,4], David Haw[1,2], Jakob Jonnerby[1,2], Christina Atchison[1], The COVID-19 Genomics UK (COG-UK) Consortium, Deborah Ashby[1], Wendy Barclay[5], Graham Taylor[5], Graham Cooke[5,6,7], Helen Ward[1,6,7], Ara Darzi[6,7,8], Steven Riley[1,2], Marc Chadeau-Hyam[1,4], Christl A. Donnelly[1,2,9]* and Paul Elliott[1,2,6,7,10,11]*

Abstract

Background: Since the emergence of SARS-CoV-2, evolutionary pressure has driven large increases in the transmissibility of the virus. However, with increasing levels of immunity through vaccination and natural infection the evolutionary pressure will switch towards immune escape. Genomic surveillance in regions of high immunity is crucial in detecting emerging variants that can more successfully navigate the immune landscape.

Methods: We present phylogenetic relationships and lineage dynamics within England (a country with high levels of immunity), as inferred from a random community sample of individuals who provided a self-administered throat and nose swab for rt-PCR testing as part of the REal-time Assessment of Community Transmission-1 (REACT-1) study. During round 14 (9 September–27 September 2021) and 15 (19 October–5 November 2021) lineages were determined for 1322 positive individuals, with 27.1% of those which reported their symptom status reporting no symptoms in the previous month.

Results: We identified 44 unique lineages, all of which were Delta or Delta sub-lineages, and found a reduction in their mutation rate over the study period. The proportion of the Delta sub-lineage AY.4.2 was increasing, with a reproduction number 15% (95% CI 8–23%) greater than the most prevalent lineage, AY.4. Further, AY.4.2 was less associated with the most predictive COVID-19 symptoms (p = 0.029) and had a reduced mutation rate (p = 0.050). Both AY.4.2 and AY.4 were found to be geographically clustered in September but this was no longer the case by late October/early November, with only the lineage AY.6 exhibiting clustering towards the South of England.

Conclusions: As SARS-CoV-2 moves towards endemicity and new variants emerge, genomic data obtained from random community samples can augment routine surveillance data without the potential biases introduced due to higher sampling rates of symptomatic individuals.

Keywords: SARS-CoV-2, COVID-19, Delta variant, Genetic diversity, Transmission advantage, Mutation

*Correspondence: c.donnelly@imperial.ac.uk; p.elliott@imperial.ac.uk

[1] School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, UK
Full list of author information is available at the end of the article

Background

Since its first documented case in India in November 2020 [1] the Delta variant of SARS-CoV-2 has spread rapidly across the world and by 16 November 2021 was

Eales *et al. BMC Infectious Diseases*   (2022) 22:647

Page 2 of 17

responsible for 99.7% of all SARS-CoV-2 infections [2]. Its rapid rise to dominance has been attributed to greater levels of transmissibility [3, 4] than previously circulating variants with the reproduction number estimated to be over two-fold higher [5], as well as possible reduced vaccine effectiveness against infection [6]. Since its global dissemination, continued adaptive evolution has led to a diverse set of Delta sub-lineages, with distinct combinations of mutations (especially on the spike protein) [7, 8].

Since July 2021 the lineage AY.4.2 (Pango nomenclature [9]), a descendant of the original Delta variant (henceforth B.1.617.2) has increased in proportion in routine surveillance data for England from 8.5% the week beginning 4 October [10] to 14.7% the week beginning 31 October [11]. AY.4.2 was declared a variant under investigation (VUI) by the UK Health Security Agency on 20 October 2021 [12]. Globally AY.4.2 had been detected in 43 countries by 22 November 2021 [13] but had only been estimated at a cumulative proportion greater than 1% in Poland [14]. AY.4.2 has two defining mutations in the spike protein, Y145H and A222V, but is otherwise similar to AY.4, a lineage that is far more widespread. AY.4 was the most prevalent lineage in England on 29 October 2021 [11] and had been detected in 87 countries by 22 November 2021 [15], in some of which it had already been reported as the most prevalent lineage (by 23 November 2021) [16, 17].

England has recorded high levels of SARS-CoV-2 infection over the course of the pandemic [4, 18] and vaccinated, as part of its mass vaccination campaign (Pfizer/BioNTec, Oxford/AstraZeneca and Moderna), a large proportion of its population (80.3% of over 12 year olds double vaccinated by 27 November 2021), with further booster jabs (Pfizer/BioNTec or Moderna) being rolled out in adults (30.5% of over 12 year olds having received a booster dose by 27 November 2021) [18]. This has led to high levels of antibodies against coronavirus with 92.8% of adults in England estimated to test positive for antibodies (IgG antibodies against the SARS-CoV-2 trimeric spike protein) in the week beginning 1 November 2021 [19]. With high vaccination coverage in the population it is likely that there is substantial selective pressure on SARS-CoV-2 towards immune escape and vaccine breakthrough infections. Genomic surveillance in highly immunised regions is crucial to detect emerging variants that can more successfully navigate the immune landscape that has been created by both natural infection and vaccination.

The REal-time Assessment of Community Transmission-1 (REACT-1) study is a series of cross-sectional surveys of the population of England that seeks to estimate the prevalence of SARS-CoV-2 on a monthly basis [4, 20], with genomic sequencing performed on all positive samples with a low enough cycle threshold (Ct) value (a proxy for viral load) and high enough volume. Due to its sampling procedure it does not suffer from the biases of routine surveillance that can be heavily biased towards symptomatic individuals [21]; symptom status can be highly dependent on levels of immunity [22]. Here we present the genomic analysis of the (N = 2163) positive samples for round 14 and round 15 which were collected from 9 to 27 September 2021 and 19 October to 5 November 2021 respectively.

Material and methods
Viral genome sequencing
The methods of the REACT-1 study have been described elsewhere [23]. REACT-1 is a repeat cross-sectional study whereby in each round a random subset of the English population (selected from the National Health Service general practitioners' patient list) is invited to obtain a self-administered swab test (parent/guardian administered for 5–12 year olds). These tests are then sent to a laboratory to undergo rt-PCR testing for the presence of SARS-CoV-2. A round of the study covers a ~ 2- to 3-week period and has occurred approximately monthly since May 2020 with between 100,000 and 185,000 individuals taking part in each round. Since round 8 in January 2021 all positive samples with a low enough N-gene Ct value (the threshold was 34 in rounds 14 and 15 presented here) and sufficient volume have been sent for genome sequencing. Amplification of the extracted RNA was performed using the ARTIC protocol [24] (version 4 primers), with sequence libraries prepared using CoronaHiT [25]; sequencing was performed on the Illumina NextSeq 500 platform. Raw sequences were analysed using the bioinformatic pipeline [26] and then uploaded to CLIMB [27]. Lineages were assigned using PangoLEARN [28] (database version 2021-11-04), a machine learning-based assignment algorithm, using Pango nomenclature [9]. For some sequences of low overall quality, a lineage designation was not possible and so they were not included in the analyses. Samples with less than 50% of bases covered were further excluded from the analysis. Of the 1322 lineages determined during rounds 14 and 15, 1160 individuals provided information on their symptoms in the previous month with 314 (27.1%) reporting no symptoms.

Phylogeographic model
For all sequences from REACT-1 rounds 11 (15 April–3 May 2021), 12 (20 May–7 June 2021), 13 (24 June–12 July 2021), 14 (9 September–27 September 2021) and 15 (19 October–5 November 2021), in which the lineage designated was Delta or a Delta sub-lineage, a maximum likelihood phylogenetic tree was constructed using a

Eales *et al. BMC Infectious Diseases*   (2022) 22:647

Page 3 of 17

HKY model implemented in IQ-TREE [29]. An uncorrelated relaxed clock model implemented in TreeTime [30], assuming a normal distribution of rates with mean 0.0008 substitutions per site per year and a single coalescent rate for the time scale, was then fit to the maximum likelihood phylogenetic tree producing a time-resolved phylogenetic tree. The mutation rates at the tree's tips were extracted from the model and a Gaussian regression model was fit to the samples obtained during round 14 and 15 for the 8 most prevalent lineages (AY.39, AY.4, AY.4.2, AY.43, AY.44, AY.5, AY.6, B.1.617.2) including lineage and round as covariates. A mugration model (implemented in TreeTime [30]) was run on the time-resolved phylogenetic tree, treating the region in which each sample was isolated as a discrete state. This allowed estimates of the migration rates between regions to be calculated (assumed to be symmetric).

Statistical analyses

The 95% confidence intervals for lineage proportions were calculated using the Wilson method [31] assuming a Binomial distribution. This method is preferred when the number of positives is low but is still valid when this is not the case [32]. Higher accuracy in confidence interval estimates for when the number of positives is low was chosen so that lower bounds on case numbers for rarer lineages were as accurate as possible.

Estimates of the true number of swab-positive infections in England during round 14 and round 15 for lineages in which only one sample was detected in a round were calculated by multiplying the estimated proportion of the lineage for each round, the weighted prevalence estimated for each round [33], and the population size of England [34]. The 95% confidence intervals were estimated by simulating the entire distribution for proportion and weighted prevalence and multiplying the two together. The distribution of weighted prevalence was estimated by randomly sampling from a normal distribution with mean value the central estimate, and standard deviation the width of the 95% confidence interval divided by 3.92 (2 times 1.96). The distribution of the lineage proportion was estimated by calculating the Wilson confidence intervals at different levels (0.00001 to 0.99999 in intervals of 0.00001).

The significance of differences in proportions of particular lineages by age group and region was calculated using Fisher's exact test with a binary outcome variable (lineage of interest or not). Differences with a p-value less than 0.05 were considered statistically significant. Analysis was only completed for a lineage in a round if there were more than 90 samples (AY.4 round 14, AY.4 round 15, AY.4.2 round 15, B.1.617.2 round 15), so that there

were, on average, more than 10 samples per parameter (9 regions in England).

Shannon diversity was calculated using all data for round 14 and round 15, and for each region for round 14 and round 15 [35]. The significance of any differences in Shannon diversity between round 14 and 15 (for all data) and between regions in each round was assessed using the Hutcheson T-test [36] and its associated p-value.

The relative growth rate of a lineage compared to all other lineages was estimated using a Bayesian logistic regression model fit to the binary outcome variable (lineage of interest or not) over time. The two model parameters (intercept and gradient) were given uninformative constant prior distributions. The probability that the growth rate was greater than zero was calculated from the model's posterior. Lineages were deemed to be different to zero if the posterior probability that the growth rate was greater than zero was greater than 0.975 or less than 0.025, similar to a p-value threshold of 0.05.

The growth rates of AY.4.2 and AY.4 infected individuals were estimated by fitting an exponential model to the daily weighted prevalence using all REACT-1 data (all negatives and all AY.4/AY.4.2 associated positives) for rounds 14 and 15 assuming a Binomial likelihood. Weightings for individual REACT-1 samples were calculated using rim weighting [37] by: sex, deciles of the IMD, LTLA counts and ethnic group. Growth rates were then converted to estimates of the reproduction number R assuming a gamma-distributed generation time with the shape parameter, $n = 2.29$, and rate parameter, $b = 0.36$ [38] through the equation $(1 + \frac{r}{b})^n$ [39]. The multiplicative R advantage of AY.4.2 over AY.4 was estimated using the entire posterior distribution of $R_{AY.4.2}/R_{AY.4}$ with the median and 95% credible interval reported.

For each lineage with more than 1 sample in a round the presence of clustering was assessed. The pairwise distance matrix between all n samples that were designated to a specific lineage was calculated and from this a mean pairwise distance was calculated for the lineage. Next, 10,000 random combinations of n positive individuals (n positive individuals chosen each time without replacement), for which any lineage was determined, were selected and for each combination the distance matrix and mean distance was calculated. The proportion of the 10,000 estimated mean distances below the lineage-specific mean distance was then calculated. Clustering was deemed to be significant if this proportion was less than 0.05.

For the 8 most prevalent lineages across rounds 14 and 15 Gaussian regression was performed to estimate the mean N- and E-gene Ct values for each lineage and p-values used to assess the significance of any difference to the reference lineage (AY.4). Models were run on all data

Eales *et al. BMC Infectious Diseases*   (2022) 22:647

Page 4 of 17

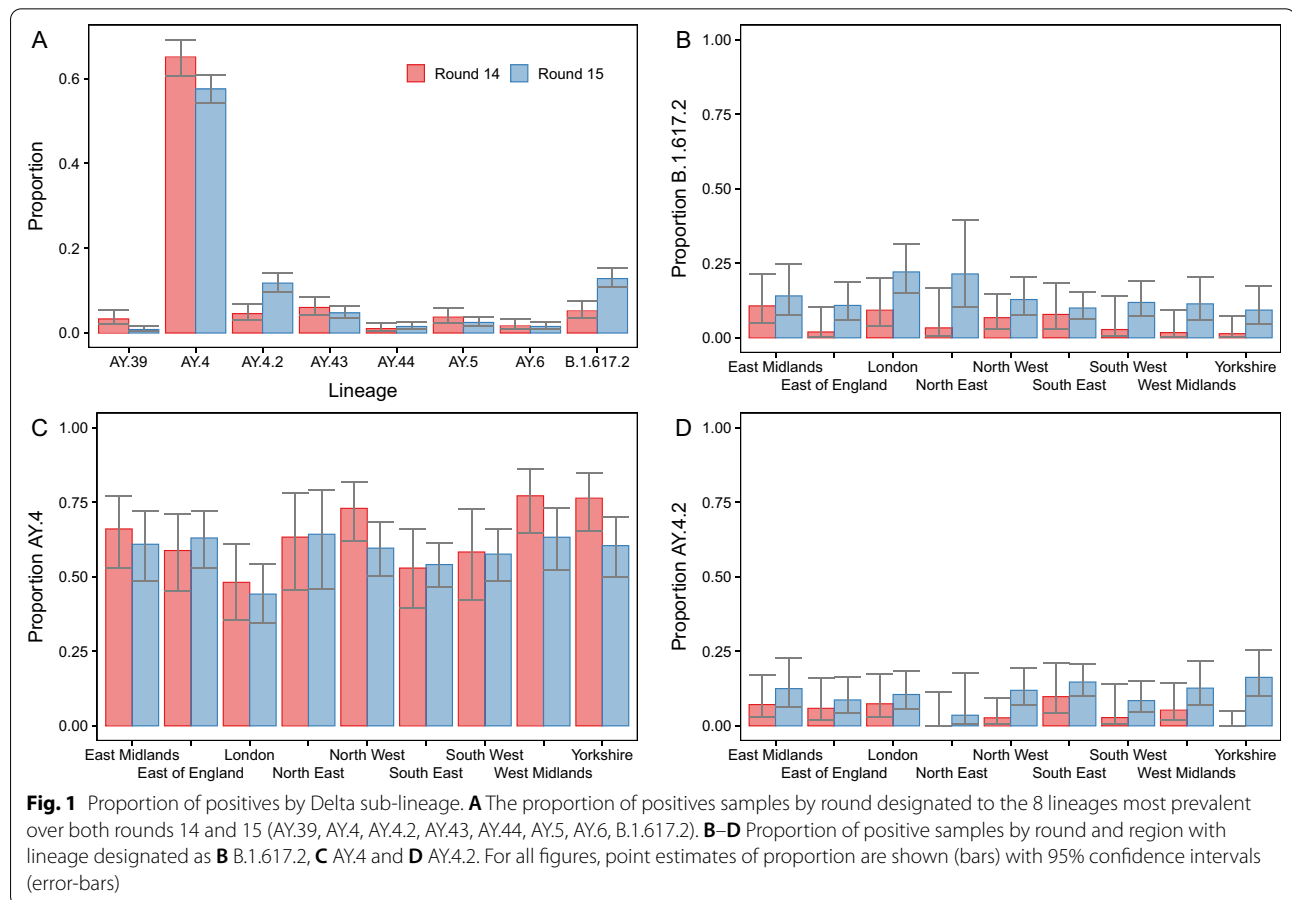(rounds 14 and 15 combined) and then run on data from each individual round as a sensitivity analysis.

The proportion of individuals reporting any symptoms in the month prior to swabbing and any of the most predictive COVID-19 symptoms in the month prior to swabbing was calculated for the 8 most prevalent lineages across rounds 14 and 15. P-values were estimated for each lineage relative to AY.4 by performing logistic regression with the symptom status as a binary variable (any symptoms vs no symptoms, and separately most predictive COVID-19 symptoms vs none of the most predictive COVID-19 symptoms). The sensitivity of the results that AY.4.2 is less likely to exhibit the most predictive COVID-19 symptoms, relative to AY.4, was assessed by fitting further logistic regression models including age, round of study and N-gene Ct value as covariates (E-gene was also investigated but was no different to using N-gene and so this was not included).

Results

Lineage diversity.

In round 14 the lineage was determined for 481 of 764 positive samples. All lineages were Delta or a Delta sub-lineage with the four most prevalent lineages being AY.4 at 65.1% (60.7%, 69.2%, n=313), AY.43 at 6.0% (4.2%, 8.5%, n=29), B.1.617.2 (original Delta variant) at 5.2% (3.6%, 7.6%, n=25) and AY.4.2 at 4.6% (3.0%, 6.8%, n=22) (Fig. 1-A, Additional file 2: Table S1). In round 15 the lineage was determined for 841 of 1399 positive samples. Again all samples were Delta or a Delta sub-lineage with the most prevalent lineages again being AY.4 at 57.6% (54.2%, 60.9%, n=484), B.1.617.2 at 12.8% (10.8%, 15.3%, n=108), AY.4.2 at 11.8% (9.8%, 14.1%, n=99) and AY.43 at 4.8% (3.5%, 6.4%, n=40). The next four most prevalent lineages over both rounds combined were AY.5, AY.6, AY.39, and AY.44. However, even a single detection of a lineage corresponded nationally to an average of 971 (95% CI [171, 5463]) individuals that would test swab-positive on any given day during round 14 and 1051 (95% CI [185, 5928]) individuals that would test swab-positive on any given day during round 15. During rounds 14 and 15 there were 33 and 31 unique lineages detected, respectively with 44 unique lineages detected overall. There was no apparent difference in genetic diversity between the two rounds as estimated by the Shannon diversity (p=0.831) (Additional file 2: Table S2).

Fig. 1 Proportion of positives by Delta sub-lineage. **A** The proportion of positives samples by round designated to the 8 lineages most prevalent over both rounds 14 and 15 (AY.39, AY.4, AY.4.2, AY.43, AY.44, AY.5, AY.6, B.1.617.2). **B–D** Proportion of positive samples by round and region with lineage designated as **B** B.1.617.2, **C** AY.4 and **D** AY.4.2. For all figures, point estimates of proportion are shown (bars) with 95% confidence intervals (error-bars)
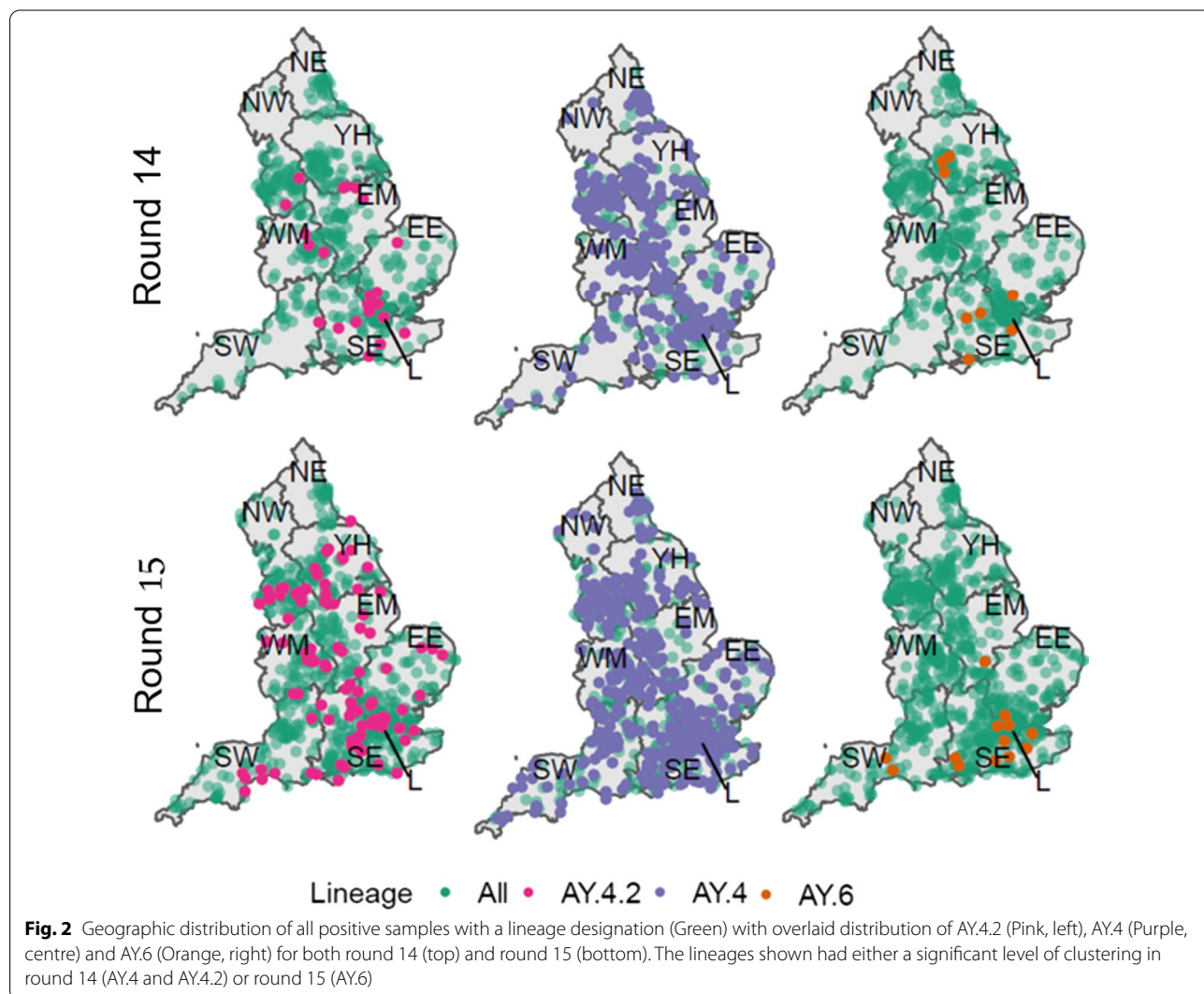
Distribution by region and age

During round 15 the proportion of B.1.617.2 was found to be highest in London at 22.1% (14.9%, 31.4%), being greater than the proportion in South East, East of England and Yorkshire and The Humber (Fig. 1B, Additional file 2: Table S3). Conversely, in round 14 and 15 the proportion of AY.4 was lowest in London at 48.1% (35.4%, 61.1%) and 44.2% (34.6%, 54.2%) respectively and was found to be higher in North West, West Midlands and Yorkshire and The Humber during both rounds (Fig. 1C, Additional file 2: Table S3). This reduced proportion of the nationally most prevalent lineage (AY.4) in London coincided with a higher level of genetic diversity in London. The Shannon diversity was highest in London during both rounds at 1.814 in round 14 and 1.809 in round 15 ($p < 0.001$ and $p = 0.002$ respectively, reference = West Midlands, Additional file 2: Table S2). Higher levels of genetic diversity were also found during both rounds in the South East and South West, relative to the West Midlands (which showed the lowest levels of genetic diversity in round 14 and the second lowest in round 15). There were no regional differences in the proportion of AY.4.2 during round 15 (Fig. 1D, Additional file 2: Table S3). Regional differences during round 14 and regional differences for other lineages could not be investigated due to small sample sizes but numbers are provided in Additional file 2: Table S4.

Sub-regional analysis was performed in order to investigate the presence of clustering in each round for each lineage (see Methods). Despite being highly geographically dispersed (Fig. 2) clustering was detected in round 14 for AY.4 ($p = 0.037$) and AY.4.2 ($p = 0.029$) (Additional file 2: Table S5). However, during round 15 clustering was no longer evident for both AY.4 ($p = 0.706$) and AY.4.2 ($p = 0.067$). The only lineage for which clustering was detected in round 15 was AY.6 ($p = 0.003$) which was

Fig. 2 Geographic distribution of all positive samples with a lineage designation (Green) with overlaid distribution of AY.4.2 (Pink, left), AY.4 (Purple, centre) and AY.6 (Orange, right) for both round 14 (top) and round 15 (bottom). The lineages shown had either a significant level of clustering in round 14 (AY.4 and AY.4.2) or round 15 (AY.6)

Eales *et al. BMC Infectious Diseases*   (2022) 22:647

Page 6 of 17

found mainly in London and towards the South coast of England.

During round 15 the proportion of B.1.617.2 was higher in individuals ages 25–34 years old at 24.2% (12.8%, 41.0%) relative to those aged 35–44 years old at 8.0% (4.1%, 15.0%) (p = 0.026) (Additional file 2: Table S6). The proportion of AY.4 was found to be lower in 5–12 year olds at 52.1% (44.6%, 59.5%) relative to 35–44 year olds in which the proportion of AY.4 was 65.0% (55.3%, 73.6%) (p = 0.042) in round 15, while it was not in round 14. There were no differences between age groups in the proportion of AY.4.2 during round 15. Differences between age groups during round 14 for AY.4.2

and other lineages could not be investigated due to small sample sizes but numbers are provided in Additional file 2: Table S7.

Detection of increasing sub-lineages

Logistic regression models were fitted to the proportion of each lineage detected in either round 14 or 15, allowing daily growth rates in proportion to be estimated (Fig. 3, Additional file 2: Table S8). Of the 44 unique lineages detected, 6 were estimated to have growth rates different to zero. AY.4, AY.39, AY.98.1 and AY.111 were decreasing in proportion, whereas AY.4.2 and B.1.617.2 were increasing in proportion. The decrease in
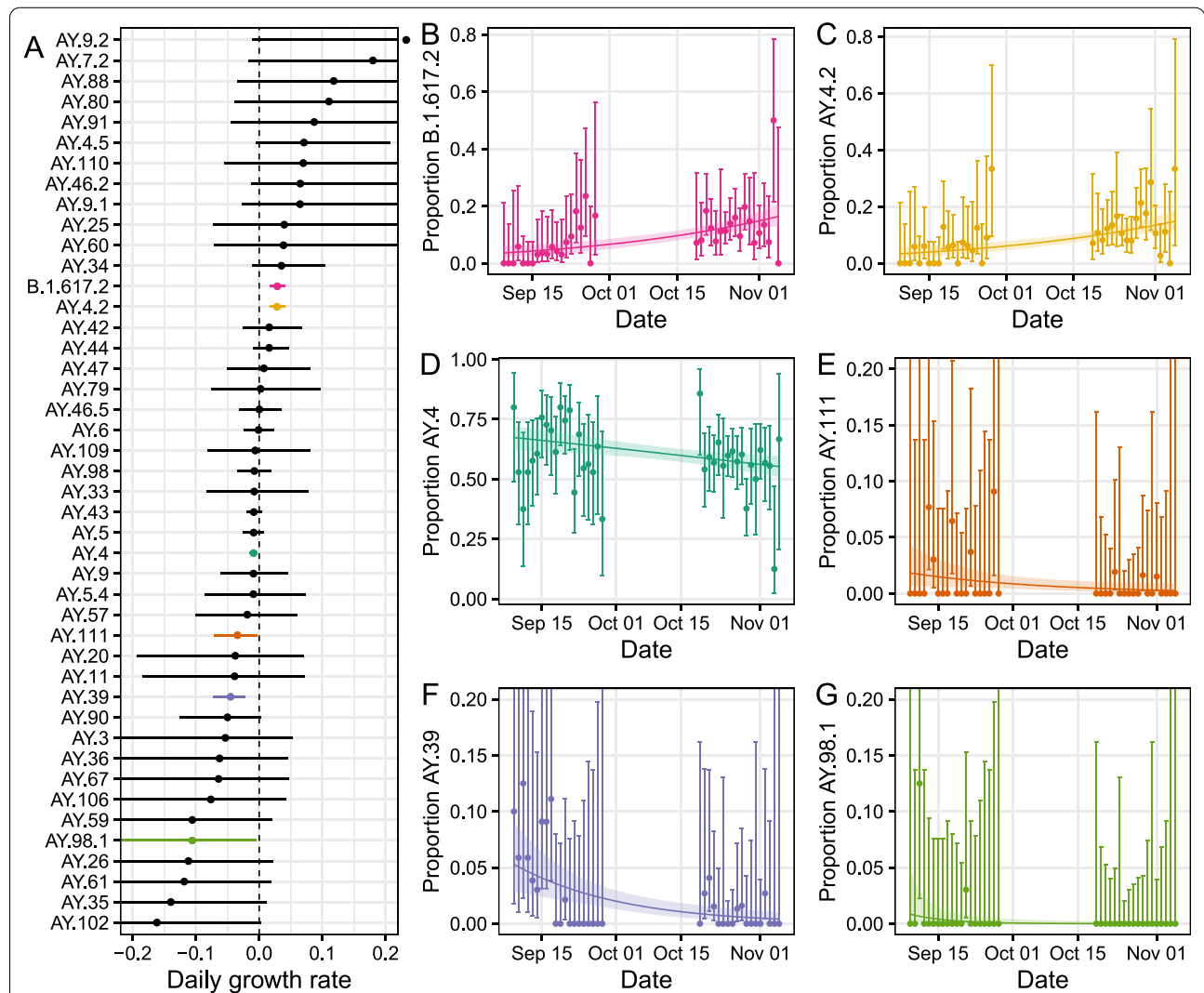
Fig. 3 A Estimated daily growth rate of the log odds of each lineage detected relative to all other lineages. Shown are both lineages with a growth rate in proportion not significantly different to zero (black) and those with a growth rate in proportion significantly different to zero (coloured). **B–G** Raw estimates of the daily proportions (points) with 95% confidence intervals (error bars) for lineages with a growth rate in proportion significantly different to zero: B.1.617.2 (**B**, pink), AY.4.2 (**C**, yellow), AY.4 (**D**, dark green), AY.111 (**E**, orange), AY.39 (**F**, purple), AY.98.1 (**G**, light green). Also shown is the best-fit Bayesian logistic regression model with central estimate (solid line) and 95% credible interval (shaded region)

Eales *et al. BMC Infectious Diseases*   (2022) 22:647

Page 7 of 17

proportion of AY.4 corresponded to a daily growth rate of -0.009 (-0.015, -0.003). The increase in proportions of B.1.617.2 and AY.4.2 corresponded to growth rates of 0.029 (0.017, 0.041) and 0.028 (0.016, 0.041) respectively.

Comparing estimates of the reproduction number R from round 14 to round 15 for AY.4 and AY.4.2 (see Methods) we estimate a multiplicative R advantage of 1.15 (1.08, 1.23), assuming no change in the generation time distribution.

Differences in cycle threshold values

There were quantitative differences between lineages in the N- and E-gene Ct values. The mean N- and E-gene Ct values were lowest for AY.6 though not materially lower than the values obtained for AY.4 (Fig. 4, Additional file 2: Table S9). Mean N-gene Ct value was 22.14 (20.30, 23.99) for AY.6 compared to 23.98 (23.68, 24.28) for AY.4 (p=0.054). Mean E-gene Ct value was 20.74 (18.90, 22.59) for AY.6 compared to 22.46 (22.16, 22.76) for AY.4 (p=0.071). Mean N- and E-gene Ct values were found to be comparable to AY.4 for both AY.4.2 and AY.5. Relative to AY.4, mean N- and E-gene Ct values for AY.43, AY.44, AY.39 and B.1.617.2 were all higher.

Differences in symptomatology

The proportion of individuals exhibiting the most predictive COVID-19 symptoms (loss or change of sense of taste, loss or change of sense of smell, new persistent cough, fever) in the month prior to swabbing was lower (p=0.029) in those infected with AY.4.2 at 42.1% (33.1%, 51.5%) relative to those infected with AY.4 at 53.4% (49.7%, 57.1%) (Fig. 5A, Additional file 2: Table S10). This difference was not explained by patterns in age, round of the study or N-gene Ct value (Fig. 5B, Additional file 2: Table S11).

In addition, 68.6% (59.8%, 76.3%) of those infected with AY.4.2 reported any symptoms in the month prior to swabbing compared to 75.4% (72.2%, 78.3%) for those infected with AY.4 (p=0.119). There were no differences evident in symptom reporting between AY.4 infected individuals and the other 6 most prevalent lineages (B.1.617.2, AY.5, AY.6, AY.43, AY.44 and AY.39).
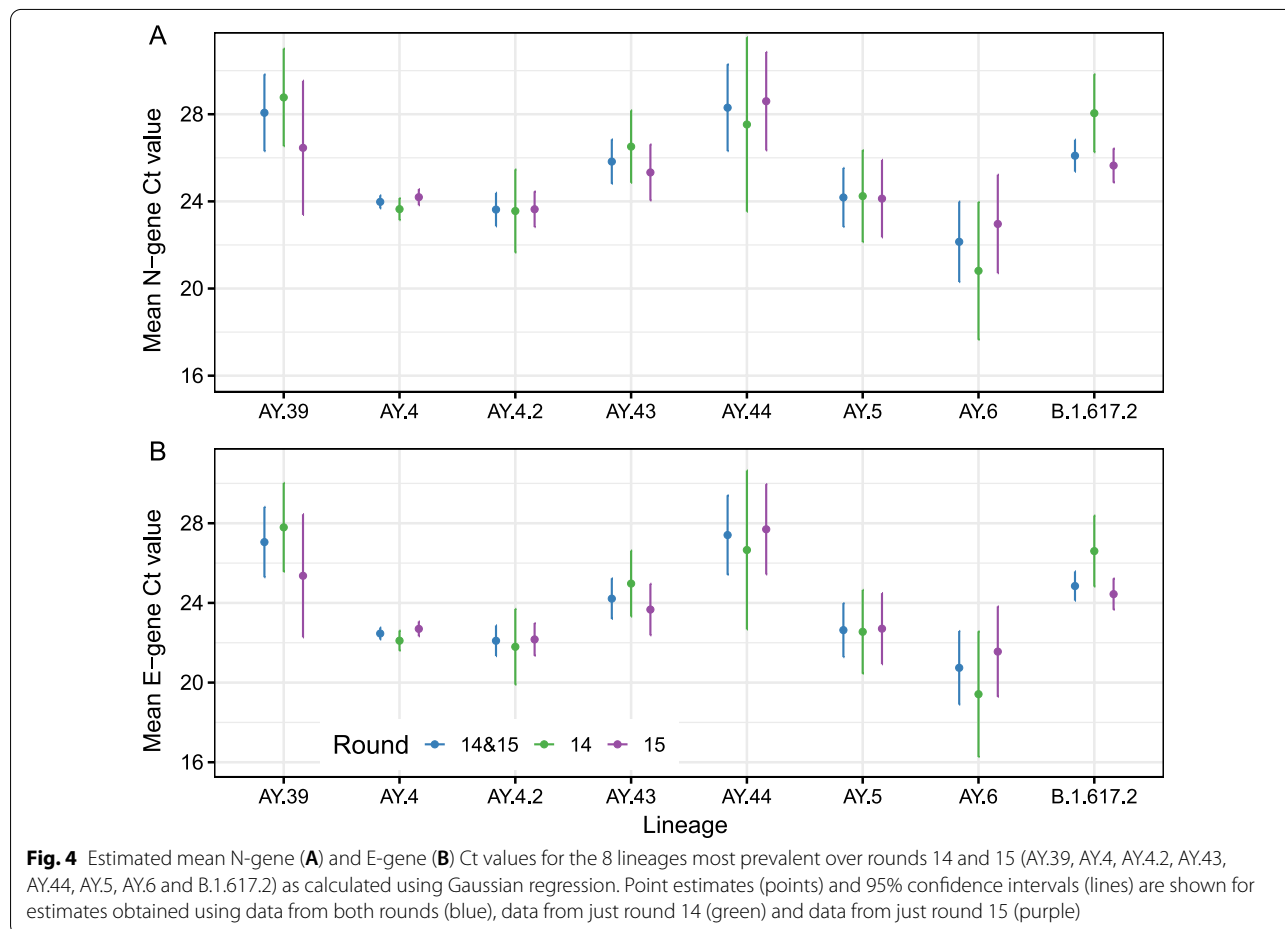
Fig. 4 Estimated mean N-gene (**A**) and E-gene (**B**) Ct values for the 8 lineages most prevalent over rounds 14 and 15 (AY.39, AY.4, AY.4.2, AY.43, AY.44, AY.5, AY.6 and B.1.617.2) as calculated using Gaussian regression. Point estimates (points) and 95% confidence intervals (lines) are shown for estimates obtained using data from both rounds (blue), data from just round 14 (green) and data from just round 15 (purple)
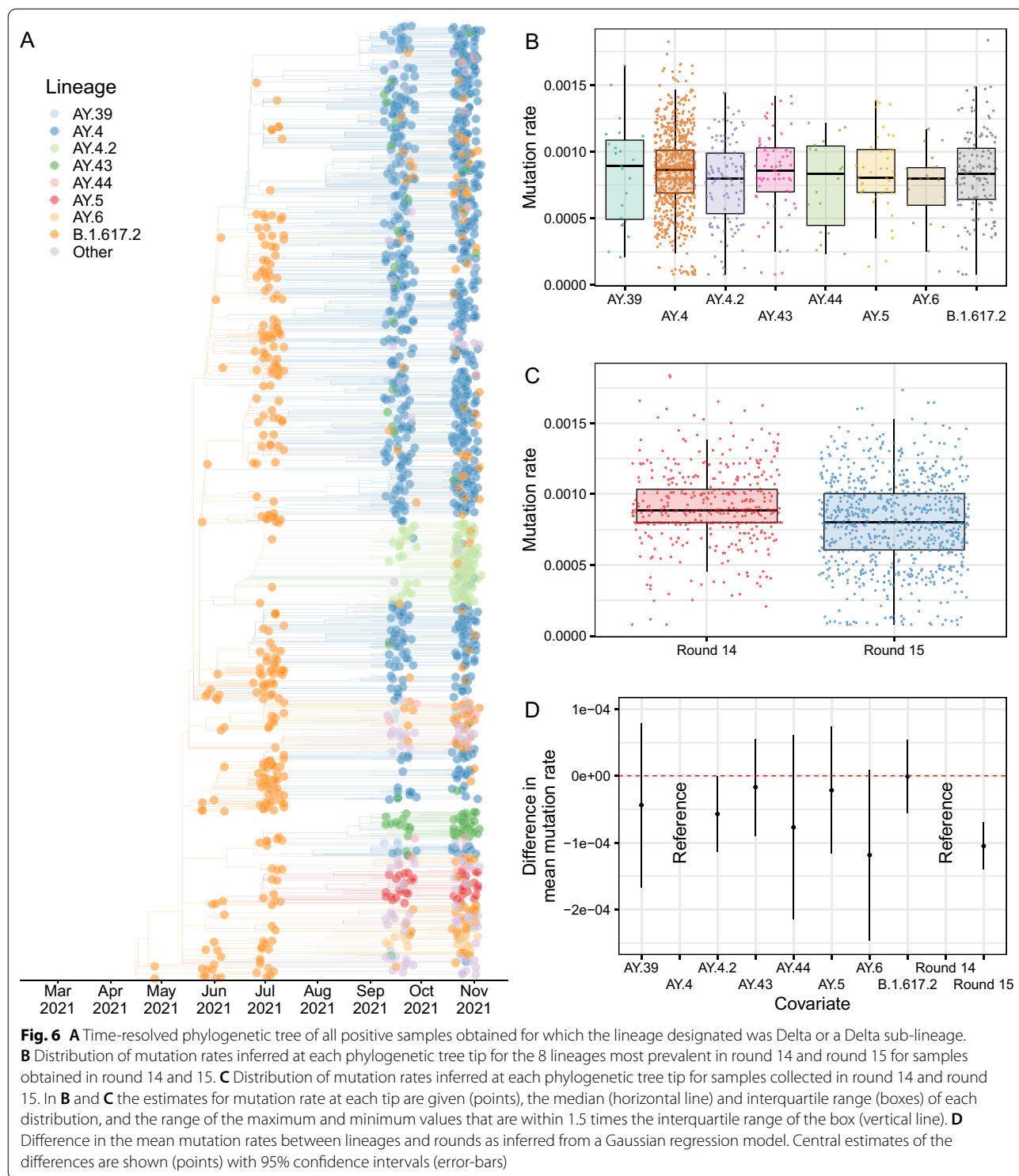
Eales *et al. BMC Infectious Diseases*   (2022) 22:647

Page 8 of 17

Fig. 5 **A** Proportion of positive individuals reporting any symptoms or reporting one of the four most predictive COVID-19 symptoms (loss or change of sense of taste, loss or change of sense of smell, new persistent cough, fever) in the last month by lineage of infection, for the 8 lineages most prevalent during rounds 14 and 15 (AY.39, AY.4, AY.4.2, AY.43, AY.44, AY.5, AY.6 and B.1.617.2). Point estimates of proportion are shown (bars) with 95% confidence intervals (error-bars). **B** Odds ratios of reporting the most predictive COVID-19 symptoms in the last months for multivariable logistic regression models including lineage (AY.4.2 with reference AY.4, red), age (relative to change of 10 years in age, blue), round of study (round 15 with reference round 14, green) and N-gene Ct value (relative to change in Ct value of 5, purple). The central estimates of odds ratios are shown (points) with 95% confidence intervals (error-bars)

Phylogeographic analysis

A relaxed molecular clock model was fit to the data and used to estimate a time-resolved phylogenetic tree (Fig. 6). AY.4.2 was found to populate two closely related clades that emerged in June/July 2021. AY.43, AY.5 and AY.6 were also observed to have distinct clade groupings having emerged around June/July 2021 as well. The mutation rates inferred at the tree's tips showed a large degree of variation in all of the 8 most prevalent lineages. The mean mutation rate for AY.4.2 was found to be 0.57 (< 0.01, 1.10)$\times 10^{-4}$ lower than the mean mutation rate of AY.4 (p$= 0.050$) (Fig. 6, Additional file 2: Table S2). The mean mutation rate inferred

for samples collected in round 15 was found to be 1.00 (0.70, 1.40)$\times 10^{-4}$ lower than the mean mutation rate for samples collected in round 14 (p< 0.001).

A mugration model was run on the time-resolved phylogenetic tree to estimate the relative virus migration rates between regions, a measure of inter-region transmission (Additional file 2: Table S13). Overall levels of inter-region transmission were lowest for the North East during round 14 and 15. The highest overall level of inter-region transmission was observed for the North West during round 14 and 15, but looking at individual rounds there were higher levels for Yorkshire and The Humber in round 14 and for the South East in round 15. High

Fig. 6 A Time-resolved phylogenetic tree of all positive samples obtained for which the lineage designated was Delta or a Delta sub-lineage. **B** Distribution of mutation rates inferred at each phylogenetic tree tip for the 8 lineages most prevalent in round 14 and round 15 for samples obtained in round 14 and 15. **C** Distribution of mutation rates inferred at each phylogenetic tree tip for samples collected in round 14 and round 15. In **B** and **C** the estimates for mutation rate at each tip are given (points), the median (horizontal line) and interquartile range (boxes) of each distribution, and the range of the maximum and minimum values that are within 1.5 times the interquartile range of the box (vertical line). **D** Difference in the mean mutation rates between lineages and rounds as inferred from a Gaussian regression model. Central estimates of the differences are shown (points) with 95% confidence intervals (error-bars)

rates of transmission during round 14 and 15 were found between the North West and Yorkshire and The Humber, the West Midlands and the South East, and also between the South East and London.

Discussion

The proportion of AY.4.2 was found to be increasing between 9 September and 5 November 2021, as also reported in the routine data surveillance for England [11]. In round 15, AY.4.2 represented 11.8% of infections

Eales *et al. BMC Infectious Diseases* (2022) 22:647

Page 10 of 17

in line with other estimates [11]. This increase in proportion corresponded to a 15% increase in transmission advantage although this assumes the generation time distribution has remained constant; a decrease of the generation time distribution for AY.4.2 would also explain the increased growth but we are unable to test for this with prevalence data. In the past, the A222V mutation, associated with AY.4.2, increased in frequency but this was eventually deemed to be due to a founder effect and not a transmission advantage [40, 41]. Given the high levels of geographic dispersion (though with some clustering) during rounds 14 and 15 it is highly unlikely that a founder effect can explain the current growth, though we can not rule out a similar effect due to higher proportions of AY.4.2 in school-aged children (prevalence increased to a greater extent in school-aged children than in adults from July to September 2021 [4, 42]). However, as the proportion AY.4.2 was approximately constant by age in round 15 this growth advantage would not be detected into the future if this was the case.

Observed distributions of N- and E-gene Ct values were similar in AY.4.2 and AY.4 and so it is unlikely that the transmission advantage observed can be attributed to a higher viral load (a Ct 1 unit lower corresponds to an approximate twofold increase in viral load [43]). However, a reduced proportion of AY.4.2 infected individuals reporting symptoms could explain the increased transmissibility in multiple ways. Higher levels of asymptomatic infection could lead to greater levels of asymptomatic transmission. Further current testing procedures and government isolation advice in England heavily focus on the most predictive COVID-19 symptoms, which are reported less often by AY.4.2 infected individuals compared with AY.4. Thus, symptom-based policies could introduce an advantage for AY.4.2 over AY.4. Finally, the reduced level of symptom reporting could be indicative of greater levels of re-infection if AY.4.2 were more successful at evading the immune response. However, studies have found that vaccines are no less effective against AY.4.2 than other Delta sub-lineages [11] and vaccine-induced antibody neutralisation titres for AY.4.2 are similar to those for AY.4 and B.1.617.2 [44]. However, any possible evasion of the immune response caused by natural infections has yet to be investigated and the numbers reporting previous infection is too small and the proportion vaccinated too large in this REACT-1 dataset to allow a meaningful comparison (715 of 817 [87.5%] individuals aged 18 and over reported having had two vaccine doses). We found a moderately reduced mutation rate of AY.4.2 relative to AY.4 which may also have introduced a fitness advantage due to a smaller number of deleterious mutations [45, 46].

Other lineages

Though we have focused on AY.4.2 we have detected a diverse set of Delta sub-lineages, with even a single detection corresponding to approximately 1000 swab-positive infections in the community at one time during the study period. The short time over which AY.4.2 went from being an undeclared lineage to a variant under investigation shows how crucial it is to have careful surveillance of all lineages irrespective of frequency. For 38 of the 44 detected lineages, it was unable to be determined whether the proportion was increasing or decreasing.

Between rounds 14 and 15 a reduction in the mean mutation rate of the virus was detected suggesting a reduction in the rate of evolution. However, despite this slowdown evolution is still occurring and we observed an increase in the proportion of B.1.617.2, an indicator that the number of undeclared B.1.617.2 sub-lineages was increasing, suggesting even further diversity of Delta sub-lineages that have yet to be given a unique lineage designation. Further, though we capture the dynamics within England, SARS-CoV-2 is a global problem and new variants of concern can arise anywhere in the world and then spread through international travel. Higher proportions of B.1.617.2 were detected in London as well as higher levels of diversity; this likely reflects the role London continues to play in the introduction of international variants [47]. Within England, the North West region played a major role in the dissemination of the virus, having the greatest inferred rate of inter-region transmission.

Analysis of N- and E-gene Ct values found decreased levels in AY.4 and AY.4.2, which is unsurprising given both have successfully disseminated across the country, but AY.5 and AY.6 were also found to have similarly low Ct values suggesting similar viral loads; the mean N- and E-gene Ct value appeared slightly lower for AY.6 compared to AY.4. Clustering was also detected in round 15 for AY.6; careful consideration of AY.6 should be given in the future in case the current lack of growth so far reported [11] has only been due to its geographic isolation.

Limitations

We have presented the inferred dynamics between Delta sub-lineages in England between 9 September and 5 November 2021. Our sample's main strength over those obtained from routine surveillance is the random nature of the testing program leading to a relatively unbiased set of positive samples. However, as the sample sizes we obtain are relatively small compared with routine national surveillance our estimates have lower precision. Lineages were only successfully determined for ~61% of positive samples, with the ability to determine a lineage heavily influenced by a sample's Ct value; this has

Eales *et al. BMC Infectious Diseases*   (2022) 22:647

Page 11 of 17

potentially led to biases with lineages with lower Ct values more heavily represented in the dataset. Detecting distinct sub-lineages is a high-dimensional problem, with often many common mutations being shared between distinct lineages with only a small number of distinguishing mutations. This is exacerbated when all the lineages are highly related, as in the current nature of the pandemic in England where all samples are descendants of Delta (B.1.617.2), and can lead to incorrect designations [48]. Further, only sub-lineages that have been defined are able to be assigned to a sample. During the emergence of a new sub-lineage there is a phase of ambiguity when numbers are small and it is unclear if the mutations present warrant the declaration of a new sub-lineage. This can be seen in the detection of AY.4.2 and AY.43; both lineages had been circulating for months by October 2021 [11] but were not yet declared sub-lineages by pangoLEARN [28] in early October 2021, and so did not appear in the publicly available technical briefings [49]. The analysis of mutation rates using Gaussian regression may also have included biases as individual measurements of mutation rates would not have independent and identically distributed normal errors, a key assumption of these linear models.

Conclusions

Since the beginning of the pandemic, selective pressure has led to rapid evolution in the spike protein [50] driving leaps in transmissibility [5]. However, as a greater proportion of the population acquires immunity through either infection or vaccination there will be a shift in evolutionary pressure towards immune escape. Even in England where there are high levels of vaccination and past infection, new variants such as AY.4.2 have emerged with advantages over previous strains. With the continued emergence of variants able to evade population immunity and undergoing transmission, SARS-CoV-2 is highly unlikely to ever undergo local extinction and is likely moving towards a state of endemicity. At the point of endemicity it is probable that adaptive evolution would more closely resemble the continual antigenic drift observed in influenza H3N2 [51, 52]. As the evolutionary phase of SARS-CoV-2 progresses towards endemicity, continued surveillance is paramount in not only detecting increased levels of transmissibility for specific lineages, but in also better characterising the mechanism behind such changes and informing policy around testing (including case definitions). Representative community studies such as REACT-1 can be useful in measuring the relative growth of lineages and in characterising differences in viral loads, symptomatology and geographic distribution.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12879-022-07628-4.

Additional file 1. REACT-1 sequence accession numbers for GISAID and the European Nucleotide Archive.

Additional file 2: Table S1. Lineages detected in rounds 14 and 15 of REACT-1. **Table S2.** Estimates of Shannon diversity for England, and by region for rounds 14 and 15 of REACT-1. **Table S3.** Regional distribution of AY.4 (round 14 and round 15), AY.4.2 (round 15) and B.1.617.2 (round 15). **Table S4.** Raw numbers of all lineages by region for round 14 and 15 of REACT-1. **Table S5.** Estimated P-value for the presence of clustering for all lineages with more than a single sample in an individual round, for round 14 and 15 of REACT-1. **Table S6.** Distribution of AY.4 (round 14 and round 15), AY.4.2 (round 15) and B.1.617.2 (round 15) by age group. **Table S7.** Raw numbers of all lineages by age group for round 14 and 15 of REACT-1. **Table S8.** Estimated growth rate in the log odds of every lineage detected relative to all other lineages from round 14 to 15 of REACT-1. **Table S9.** Mean N- and E-gene Ct value for the eight most prevalent lineages as inferred from Gaussian regression. **Table S10.** Symptom status by lineage for the eight most prevalent lineages in rounds 14 and 15 of REACT-1. **Table S11.** Multivariable logistic regression models to determine the effect of the lineage AY.4.2 on the odds of an individual reporting any of the most predictive COVID-19 symptoms relative to AY.4. **Table S12.** Multivariable gaussian regression model to determine the effect of lineage and round of the study on mean mutation rate. **Table S13.** Average inter-region migration rate, inferred from a mugration model run on a time-resolved phylogenetic tree, for the periods of rounds 14 and 15, round 14 and round 15

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[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[11,74] and Prof Nicholas J Loman[43]

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

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

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

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

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

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

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

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

Funding acquisition, Leadership and supervision, Metadata curation, Sequencing and analysis, and Software and analysis tools:
Dr Gaia Nebbia[12]

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

Leadership and supervision, Metadata curation, Project administration, Samples and logistics, and Sequencing and analysis:
Prof Katrina Lythgoe[5], Dr M. Estee Torok[19,20] and Prof Ian G Goodfellow[24]

Leadership and supervision, Metadata curation, Project administration, Samples and logistics, and Visualisation:
Dr Jacqui A Prieto[97,82] and Dr Kordo Saeed[97,83]

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

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

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

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

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

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

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

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

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

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

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

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

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

Leadership and supervision, Metadata curation, Samples and logistics, and Visualisation:
Dr SamuelMoses[25,106]

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

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

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

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

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

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

Metadata curation, Samples and logistics, Sequencing and analysis, and Visualisation:
Dr Sharon Glaysher[68]

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

Eales *et al. BMC Infectious Diseases*   (2022) 22:647

Page 13 of 17

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

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

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

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

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

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

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

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

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

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

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

Leadership and supervision, Sequencing and analysis, and Software and analysis tools:
Prof Oliver G Pybus[23], Dr Erik M Volz[39], Prof Daniela de Angelis[52], Prof David L Robertson[53], Dr Andrew J Page[75] and Dr Inigo Martincorena[116]

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

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

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

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

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

Metadata curation, Samples and logistics, and Sequencing and analysis:
Dr James P McKenna[3], Dr Ben Warne[9], Joshua F Taylor[22], Yasmin Chaudhry[24], Rhys Izuagbe[24], Dr Aminu S Jahun[24], Dr Gregory R Young [6,65], Dr Claire McMurray[43], Dr Clare M McCann[65,66], Dr Andrew Nelson[65,66] and Scott Elliott[68]

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

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

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

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

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

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

Sequencing and analysis, Software and analysis tools, and Visualization:
Dr Helen Adams[4], Dr Louis du Plessis[23], Dr Rob Johnson[39], Dr William T Harvey[53,42], Dr Joseph Hughes[53], Dr Richard J Orton[53], Dr Lewis GSpurgin[59], Dr Yann Bourgeois[81], Dr Chris Ruis[102], Áine O'Toole[104], Marina Gourtovaia[116] and Dr Theo Sanderson[116]

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

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

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

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

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

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

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

Eales *et al. BMC Infectious Diseases* (2022) 22:647

Page 14 of 17

Metadata curation, and Sequencing and analysis:
Matthew Carlile[18], Dr Nadine Holmes[18], Dr Christopher Moore[18], Dr Nathaniel Storey[29], Dr Stefan Rooke[73], Dr Gonzalo Yebra[73], Dr Noel Craine[74], Malorie Perry[74], Dr Nabil-Fareed Alikhan[75], Dr Stephen Bridgett[77], Kate F Cook[84], Christopher Fearn[84], Dr Salman Goudarzi[84], Prof Ronan A Lyons[88], Dr Thomas Williams[104], Dr Sam T Haldenby[107], Jillian Durham[116] and Dr Steven Leonard[116]

Metadata curation, and Software and analysis tools:
Robert M Davies[116]

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

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

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

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

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

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

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

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

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

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

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

Samples and logistics:
Dr Themoula Charalampous[12,46], Amita Patel[12], Dr Louise Berry[15], Dr Tim Boswell[15], Dr Vicki M Fleming[15], Dr Hannah C Howson-Wells[15], Dr Amelia Joseph[15], Manjinder Khakh[15], Dr Michelle M Lister[15], Paul W Bird[16], Karlie Fallon[16], Thomas Helmer[16], Dr Claire L McMurray[16], Mina Odedra[16], Jessica Shaw[16], Dr Julian W Tang[16], Nicholas J Willford[16], Victoria Blakey[17], Dr Veena Raviprakash[17], Nicola Sheriff[17], Lesley-Anne Williams[17], Theresa Feltwell[20], Dr Luke Bedford[26], Dr James S Cargill[27], Warwick Hughes[27], Dr Jonathan Moore[28], Susanne Stonehouse[28], Laura Atkinson[29], Jack CD Lee[29], Dr Divya Shah[29], Adela Alcolea-Medina[32,112], Natasha Ohemeng-Kumi[32,112], John Ramble[32,112], Jasveen Sehmi[32,112], Dr Rebecca Williams[33], Wendy Chatterton[34], Monika Pusok[34], William Everson[37], Anibolina Castigador[44], Emily Macnaughton[44], Dr Kate El Bouzidi[45], Dr Temi Lampejo[45], Dr Malur Sudhanva[45], Cassie Breen[47], Dr Graciela Sluga[48], Dr Shazaad SY Ahmad[49,70], Dr Ryan P George[49], Dr Nicholas W Machin[49,70], Debbie Binns[50], Victoria James[50], Dr Rachel Blacow[55], Dr Lindsay Coupland[58], Dr Louise Smith[59], Dr Edward Barton[60], Debra Padgett[60], Garren Scott[60], Dr Aidan Cross[61], Dr Mariyam Mirfenderesky[61], Jane Greenaway[62], Kevin Cole[64], Phillip Clarke[67], Nichola Duckworth[67], Sarah Walsh[67], Kelly Bicknell[68], Robert Impey[68], Dr Sarah Wyllie[68], Richard Hopes[70], Dr Chloe Bishop[72], Dr Vicki Chalker[72], Dr Ian Harrison[72], Laura Gifford[74], Dr Zoltan Molnar[77], Dr Cressida Auckland[79], Dr Cariad Evans[85,109], Dr Kate Johnson[85,109], Dr David G Partridge[85,109], Dr Mohammad Raza[85,109], Paul Baker[86], Prof Stephen Bonner[86], Sarah Essex[86], Leanne J Murray[86], Andrew I Lawton[87], Dr Shirelle Burton-Fanning[89], Dr Brendan Al Payne[89], Dr Sheila Waugh[89], Andrea N Gomes[91], Maimuna Kimuli[91], Darren Murray[91], Paula Ashfield[92], Dr Donald Dobie[92], Dr Fiona Ashford[93], Dr Angus Best[93], Dr Liam Crawford[93], Dr Nicola Cumley[93], Dr Megan Mayhew[93], Dr Oliver Megram[93], Dr Jeremy Mirza[93], Dr Emma Moles-Garcia[93], Dr Benita Percival[93], Megan Driscoll[96], Leah Ensell[96], Dr Helen L Lowe[96], Laurentiu Maftei[96], Matteo Mondani[96], Nicola J Chaloner[99], Benjamin J Cogger[99], Lisa J Easton[99], Hannah Huckson[99], Jonathan Lewis[99], Sarah Lowdon[99], Cassandra S Malone[99], Florence Munemo[99], Manasa Mutingwende[99], Roberto Nicodemi[99], Olga Podplomyk[99], Thomas Somassa[99], Dr Andrew Beggs[100], Dr Alex Richter[100], Claire Cormie[102], Joana Dias[102], Sally Forrest[102], Dr Ellen E Higginson[102], Mailis Maes[102], Jamie Young[102], Dr Rose K Davidson[103], Kathryn A Jackson[107], Dr Lance Turtle[107], Dr Alexander J Keeley[109], Prof Jonathan Ball[113], Timothy Byaruhanga[113], Dr Joseph G Chappell[113], Jayasree Dey[113], Jack D Hill[113], Emily J Park[113], Arezou Fanaie[114], Rachel A Hilson[114], Geraldine Yaze[114] and Stephanie Lo[116]

Sequencing and analysis:
Safiah Afifi[10], Robert Beer[10], Joshua Maksimovic[10], Kathryn McCluggage[10], Karla Spellman[10], Catherine Bresner[11], William Fuller[11], Dr Angela Marchbank[11], Trudy Workman[11], Dr Ekaterina Shelest[13,81], Dr Johnny Debebe[18], Dr Fei Sang[18], Dr Marina Escalera Zamudio[23], Dr Sarah Francois[23], Bernardo Gutierrez[23], Dr Tetyana I Vasylyeva[23], Dr Flavia Flaviani[31], Dr Manon Ragonnet-Cronin[39], Dr Katherine L Smollett[42], Alice Broos[53], Daniel Mair[53], Jenna Nichols[53], Dr Kyriaki Nomikou[53], Dr Lily Tong[53], Ioulia Tsatsani[53], Prof Sarah O'Brien[54], Prof Steven Rushton[54], Dr Roy Sanderson[54], Dr Jon Perkins[55], Seb Cotton[56], Abbie Gallagher[56], Dr Elias Allara[70,102], Clare Pearson[70,102], Dr David Bibby[72], Dr Gavin Dabrera[72], Dr Nicholas Ellaby[72], Dr Eileen Gallagher[72], Dr Jonathan Hubb[72], Dr Angie Lackenby[72], Dr David Lee[72], Nikos Manesis[72], Dr Tamyo Mbisa[72], Dr Steven Platt[72], Katherine A Twohig[72], Dr Mari Morgan[74], Alp Aydin[75], David J Baker[75], Dr Ebenezer Foster-Nyarko[75], Dr Sophie J Prosolek[75], Steven Rudder[75], Chris Baxter[77], Sílvia F Carvalho[77], Dr Deborah Lavin[77], Dr Arun Mariappan[77], Dr Clara Radulescu[77], Dr Aditi Singh[77], Miao Tang[77], Helen Morcrette[79], Nadua Bayzid[96], Marius Cotic[96], Dr Carlos E Balcazar[104], Dr Michael D Gallagher[104], Dr Daniel Maloney[104], Thomas D Stanton[104], Dr Kathleen A Williamson[104], Dr Robin Manley[105], Michelle L Michelsen[105], Dr Christine M Sambles[105], Dr David J Studholme[105], Joanna Warwick-Dugdale[105], Richard Eccles[107], Matthew Gemmell[107], Dr Richard Gregory[107], Dr Margaret Hughes[107], Charlotte Nelson[107], Dr Lucille Rainbow[107], Dr Edith E Vamos[107], Hermione J Webster[107], Dr Mark Whitehead[107], Claudia Wierzbicki[107], Dr Adrienn Angyal[109], Dr Luke R Green[109], Dr Max Whiteley[109], Emma Betteridge[116], Dr Iraad F Bronner[116], Ben

Eales *et al. BMC Infectious Diseases*   (2022) 22:647

Page 15 of 17

W Farr[116], Scott Goodwin[116], Dr Stefanie V Lensing[116], Shane A McCarthy[116,102], Dr Michael A Quail[116], Diana Rajan[116], Dr Nicholas M Redshaw[116], Carol Scott[116], Lesley Shirley[116] and Scott AJ Thurston[116]

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

Author contributions

PE and CAD are joint corresponding authors. OE, SR, MC-H, CAD and PE conceived the study and the analytical plan. OE and LdOM performed the statistical analyses. OE, HWang, DH, BB and JJ curated the data. CA, DA, WB, GT, GC, HW, AD provided insights into the study design and results interpretation. AJP generated the sequencing data. AD and PE obtained funding. All authors revised the manuscript for important intellectual content and approved the submission of the manuscript. PE had full access to the data and takes responsibility for the integrity of the data and the accuracy of the data analysis and for the decision to submit for publication. All authors read and approved the final manuscript.

Availability of data and materials

Access to REACT-1 data is restricted due to ethical and security considerations. Summary statistics and descriptive tables from the current REACT-1 study are available in the Additional file 2. Additional summary statistics and results from the REACT-1 programme are also available at https://www.imperial.ac.uk/medicine/research-and-impact/groups/react-study/real-time-assessment-of-community-transmission-findings/ and https://github.com/mrc-ide/reactidd/tree/master/inst/extdata REACT-1 Study Materials are available for each round at https://www.imperial.ac.uk/medicine/research-and-impact/groups/react-study/react-1-study-materials/. Sequence read data are available without restriction from the European Nucleotide Archive at https://www.ebi.ac.uk/ena/browser/view/PRJEB37886, and consensus genome sequences are available from the Global initiative on sharing all influenza data at https://www.gisaid.org. Accession numbers are provided in the Additional file 1.

Declarations

Ethics approval and consent to participate

We obtained research ethics approval from the South Central-Berkshire B Research Ethics Committee (IRAS ID: 283787). All methods were carried out in accordance with relevant guidelines and regulations. Informed consent was obtained from all participants or their parent/guardian for minors. During initial registration for the study participants are asked "Are you willing to take part in this study?/Are you willing for your child to take part in this study?" with possible answers being "1. Yes, I want to take part in this study" or "2. No, I do not want to take part". Those who answered "2. No, I do not want to take part." were not sent testing kits and did not participate further in the study. Full registration forms for all rounds of REACT-1 are available here: https://www.imperial.ac.uk/medicine/research-and-impact/groups/react-study/for-researchers/react-1-study-materials/.

Consent for publication

Not applicable.

Competing interests

We declare no competing interests.

Author details
[1]School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, UK. [2]MRC Centre for Global Infectious Disease Analysis and Jameel Institute, Imperial College London, London, UK. [3]Quadram Institute, Norwich, UK. [4]MRC Centre for Environment and Health, School of Public Health, Imperial College London, London, UK. [5]Department of Infectious Disease, Imperial College London, London, UK. [6]Imperial College Healthcare NHS Trust, London, UK. [7]National Institute for Health Research Imperial Biomedical Research Centre, London, UK. [8]Institute of Global Health Innovation, Imperial College London, London, UK. [9]Department of Statistics, University of Oxford, Oxford, UK. [10]Health Data Research (HDR) UK, Imperial College London, London, UK. [11]UK Dementia Research Institute Centre at Imperial, Imperial College London, London, UK.

References
1. Lineage B.1.617.2 Pangolin report. https://cov-lineages.org/global_report_B.1.617.2.html. Accessed 26 Nov 2021.
2. World Health Organisation. COVID-19 Weekly Epidemiological Update - Edition 66. 2021. https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---16-november-2021. Accessed 23 Nov 2021.
3. PHE Genomics Cell, PHE Outbreak Surveillance Team, PHE Epidemiology Cell, PHE Contact Tracing Data Team, PHE Health, Protection Data Science Team, PHE Joint Modelling Team, NHS Test and Trace Joint Biosecurity Centre, Public Health Scotland and EAVE group, Contributions from the Variant Technical Group Members. SARS-CoV-2 variants of concern and variants under investigation in England - Technical briefing 15, 11 June 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993879/Variants_of_Concern_VOC_Technical_Briefing_15.pdf
4. 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;374:eabl9551.
5. Obermeyer F, Schaffner SF, Jankowiak M, Barkas N, Pyle JD, Park DJ, et al. Analysis of 2.1 million SARS-CoV-2 genomes identifies mutations associated with transmissibility. https://doi.org/10.1101/2021.09.07.21263228
6. Bernal JL, Andrews N, Gower C, Gallagher E, Simmons R, Thelwall S, et al. Effectiveness of COVID-19 vaccines against the B.1.617.2 variant. medRxiv. 2021; 2021.05.22.21257658.
7. Baj A, Novazzi F, Drago Ferrante F, Genoni A, Tettamanzi E, Catanoso G, et al. Spike protein evolution in the SARS-CoV-2 Delta variant of concern: a case series from Northern Lombardy. Emerg Microbes Infect. 2021;10:2010–5.
8. Kistler KE, Huddleston J, Bedford T. Rapid and parallel adaptive mutations in spike S1 drive clade success in SARS-CoV-2. BioRxiv. 2021. https://doi.org/10.1101/2021.09.11.459844.
9. 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. Nat Microbiol. 2020;5:1403–7.
10. UKHSA Genomics Cell UKHSA Outbreak Surveillance Team UKHSA Epidemiology Cell UKHSA Contact Tracing Data Team UKHSA International Cell UKHSA Environmental Monitoring for Health Protection Team. SARS-CoV-2 variants of concern and variants under investigation in England - Technical briefing 27, 29 October 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1029715/technical-briefing-27.pdf
11. UKHSA Genomics Cell UKHSA Outbreak Surveillance Team UKHSA Epidemiology Cell UKHSA Contact Tracing Data Team UKHSA International Cell UKHSA Environmental Monitoring for Health Protection Team. SARS-CoV-2 variants of concern and variants under investigation in England - Technical briefing 28, 12 November 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1033101/Technical_Briefing_28_12_Nov_2021.pdf
12. UKHSA Genomics Cell UKHSA Outbreak Surveillance Team UKHSA Epidemiology Cell UKHSA Contact Tracing Data Team UKHSA International Cell UKHSA Environmental Monitoring for Health Protection Team. SARS-CoV-2 variants of concern and variants under investigation in England - Technical briefing 26, 22 October 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1028113/Technical_Briefing_26.pdf
13. Lineage AY.4.2 Pangolin report. https://cov-lineages.org/lineage.html?lineage=AY.4.2. Accessed 23 Nov 2021.
14. Alaa Abdel Latif, Julia L. Mullen, Manar Alkuzweny, Ginger Tsueng, Marco Cano, Emily Haag, Jerry Zhou, Mark Zeller, Emory Hufbauer, Nate Matteson, Chunlei Wu, Kristian G. Andersen, Andrew I. Su, Karthik Gangavarapu, Laura D. Hughes, and the Center for Viral Systems Biology. AY.4 lineage report, outbreak.info. https://outbreak.info/situation-reports?pango=AY.4.2. Accessed 23 Nov 2021.
15. Lineage AY.4 Pangolin report. https://cov-lineages.org/lineage.html?lineage=AY.4. Accessed 23 Nov 2021.
16. Umair M, Ikram A, Rehman Z, Haider A, Badar N, Ammar M, et al. Genomic diversity of SARS-CoV-2 in Pakistan during fourth wave of pandemic. BioRxiv. 2021. https://doi.org/10.1101/2021.09.30.21264343.
17. Danish Covid-19 Genome Consortium. Genomic overview of SARS-CoV-2 in Denmark, 19 November 2021. https://www.covid19genomics.dk/statistics. Accessed 23 Nov 2021.
18. Official UK Coronavirus Dashboard. https://coronavirus.data.gov.uk/. Accessed 17 May 2021.
19. Latest insights team. Coronavirus (COVID-19) latest insights - Office for National Statistics. Office for National Statistics; 23 Nov 2021. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19latestinsights/antibodies. Accessed 25 Nov 2021.
20. Riley S, Ainslie KEC, Eales O, Walters CE, Wang H, Atchison C, et al. Resurgence of SARS-CoV-2: detection by community viral surveillance. Science. 2021. https://doi.org/10.1126/science.abf0874.
21. Ricoca Peixoto V, Nunes C, Abrantes A. Epidemic surveillance of Covid-19: considering uncertainty and under-ascertainment. Portuguese J Public Health. 2020;38:23–9.
22. Feng S, Phillips DJ, White T, Sayal H, Aley PK, Bibi S, et al. Correlates of protection against symptomatic and asymptomatic SARS-CoV-2 infection. Nat Med. 2021;27:2032–40.
23. Riley S, Atchison C, Ashby D, Donnelly CA, Barclay W, Cooke G, et al. REal-time Assessment of Community Transmission (REACT) of SARS-CoV-2 virus: study protocol. Wellcome Open Res. 2020;5:200.
24. Quick J. nCoV-2019 sequencing protocol v3 (LoCost). 2020. https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye. Accessed 4 May 2021.
25. Baker DJ, Aydin A, Le-Viet T, Kay GL, Rudder S, de Oliveira ML, et al. CoronaHiT: high-throughput sequencing of SARS-CoV-2 genomes. Genome Med. 2021;13:21.
26. A Nextflow pipeline for running the ARTIC network's field bioinformatics tools. Github. https://github.com/connor-lab/ncov2019-artic-nf
27. Connor TR, Loman NJ, Thompson S, Smith A, Southgate J, Poplawski R, et al. CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community. Microb Genom. 2016;2: e000086.
28. Phylogenetic Assignment of Named Global Outbreak LINeages (PANGOLIN). Github; https://github.com/cov-lineages/pangolin
29. 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. 2015;32:268–74.
30. Sagulenko P, Puller V, Neher RA. TreeTime: maximum-likelihood phylodynamic analysis. Virus Evol. 2018;4:vex042.
31. Wilson EB. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc. 1927;22:209–12.
32. Brown LD, Tony Cai T, DasGupta A. Interval estimation for a binomial proportion. SSO Schweiz Monatsschr Zahnheilkd. 2001;16:101–33.
33. Chadeau-Hyam M, Eales O, Bodinier B, Wang H, Haw D, Whitaker M, et al. REACT-1 round 15 final report: Increased breakthrough SARS-CoV-2 infections among adults who had received two doses of vaccine, but booster doses and first doses in children are providing important protection. 2021. http://spiral.imperial.ac.uk/handle/10044/1/92501. Accessed 26 Nov 2021.
34. Park N. Population estimates for the UK, England and Wales, Scotland and Northern Ireland - Office for National Statistics. Office for National Statistics; 24 Jun 2021. https://www.ons.gov.uk/peoplepopulationandc

Eales *et al. BMC Infectious Diseases*   (2022) 22:647

Page 17 of 17

ommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2020. Accessed 11 Aug 2021.

35. Spellerberg IF, Fedor PJ. A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the "Shannon-Wiener" Index. Glob Ecol Biogeogr. 2003;12:177–9.

36. Hutcheson K. A test for comparing diversities based on the Shannon formula. J Theor Biol. 1970;29:151–4.

37. Sharot T. Weighting survey results. J Mark Res Soc. 1986;28:269–84.

38. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and Transmission of COVID-19 in Shenzhen China: Analysis of 391 cases and 1,286 of their close contacts. MedRxiv. 2020. https://www.medrxiv.org/content/medrxiv/early/2020/03/19/2020.03.03.20028423.full.pdf

39. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. Proc Biol Sci. 2007;274:599–604.

40. Díez-Fuertes F, Iglesias-Caballero M, García-Pérez J, Monzón S, Jiménez P, Varona S, et al. A founder effect led early SARS-CoV-2 transmission in Spain. J Virol. 2021. https://doi.org/10.1128/JVI.01583-20.

41. 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;595:707–12.

42. Chadeau-Hyam M, Wang H, Eales O, Haw D, Bodinier B, Whitaker M, et al. REACT-1 study round 14: High and increasing prevalence of SARS-CoV-2 infection among school-aged children during September 2021 and vaccine effectiveness against infection in England. medRxiv. 2021; 2021.10.14.21264965.

43. Yelin I, Aharony N, Tamar ES, Argoetti A, Messer E, Berenbaum D, et al. Evaluation of COVID-19 RT-qPCR Test in Multi sample Pools. Clin Infect Dis. 2020;71:2073–8.

44. Lassaunière R, Polacek C, Fonager J, Bennedbæk M, Boding L, Rasmussen M, et al. Neutralisation of the SARS-CoV-2 Delta sub-lineage AY.4.2 and B.1.617.2 + E484K by BNT162b2 mRNA vaccine-elicited sera. bioRxiv. 2021. https://doi.org/10.1101/2021.11.08.21266075.

45. Peck KM, Lauring AS. Complexities of Viral Mutation Rates. J Virol. 2018. https://doi.org/10.1128/JVI.01031-17.

46. Koelle K, Rasmussen DA. The effects of a deleterious mutation load on patterns of influenza A/H3N2's antigenic evolution in humans. Elife. 2015;4: e07361.

47. 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. https://doi.org/10.1101/2021.05.08.21256867.

48. 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:vea0b64.

49. UKHSA Genomics Cell UKHSA Outbreak Surveillance Team UKHSA Epidemiology Cell UKHSA Contact Tracing Data Team UKHSA International Cell UKHSA Environmental Monitoring for Health Protection Team. SARS-CoV-2 variants of concern and variants under investigation in England - Technical briefing 25, 1 October 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1025827/Technical_Briefing_25.pdf

50. Saputri DS, Li S, van Eerden FJ, Rozewicki J, Xu Z, Ismanto HS, et al. Flexible, functional, and familiar: characteristics of SARS-CoV-2 spike protein evolution. Front Microbiol. 2020;11:2112.

51. Bedford T, Riley S, Barr IG, Broor S, Chadha M, Cox NJ, et al. Global circulation patterns of seasonal influenza viruses vary with antigenic drift. Nature. 2015;523:217–20.

52. Andrew Rambaut, Oliver G. Pybus, Martha I. Nelson, Cecile Viboud, Jeffery K. Taubenberger, Edward C. Holmes. The genomic and epidemiological dynamics of human influenza A virus. Nature. 2008. https://www.nature.com/articles/nature06945

Publisher's Note