science.sciencemag.org/cgi/content/full/science.abg0821/DC1

Supplementary Materials for

SARS-CoV-2 within-host diversity and transmission

Katrina A. Lythgoe[1,2]*†, Matthew Hall[1]*†, Luca Ferretti[1], Mariateresa de Cesare[1,3], George MacIntyre-Cockett[1,3], Amy Trebes[3], Monique Andersson[4,5], Newton Otecko[1], Emma L. Wise[6,7], Nathan Moore[6], Jessica Lynch[6], Stephen Kidd[6], Nicholas Cortes[6,8], Matilde Mori[9], Rebecca Williams[6], Gabrielle Vernet[6], Anita Justice[4], Angie Green[3], Samuel M. Nicholls[10], M. Azim Ansari[11], Lucie Abeler-Dörner[1], Catrin E. Moore[1], Timothy E. A. Peto[4,12], David W. Eyre[4,13], Robert Shaw[4], Peter Simmonds[11], David Buck[3], John A. Todd[3] on behalf of the Oxford Virus Sequencing Analysis Group (OVSG)‡, Thomas R. Connor[14,15], Shirin Ashraf[16], Ana da Silva Filipe[16], James Shepherd[16], Emma C. Thomson[16], The COVID-19 Genomics UK (COG-UK) Consortium§, David Bonsall[1,3,4], Christophe Fraser[1,3,17], Tanya Golubchik[1,2]†

*These authors contributed equally to this work.
†Corresponding author. Email: Tanya.Golubchik@bdi.ox.ac.uk (T.G.);
Katrina.Lythgoe@bdi.ox.ac.uk (K.A.L.); Matthew.Hall@bdi.ox.ac.uk (M.H.)
‡The full list of the OVSG members is provided in the supplementary materials.
§The full list of names and affiliations of COG-UK members is provided in the supplementary materials.

Published 9 March 2021 on *Science* First Release
DOI: 10.1126/science.abg0821

This PDF file includes:

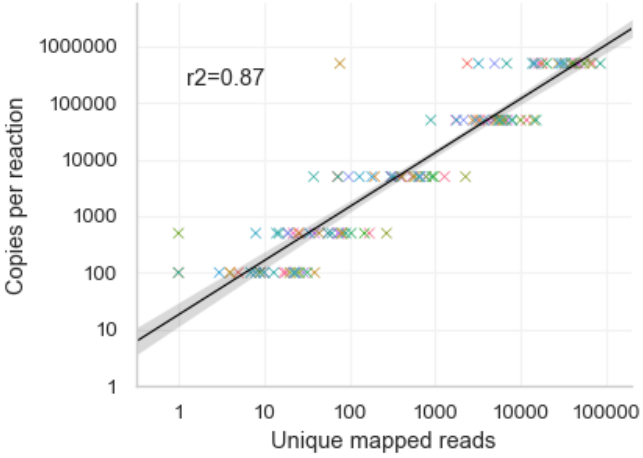
- Figs. S1 to S8
- Tables S1 to S5
- List of OVSG members
- List of COG-UK consortium names and affiliations

Other Supplementary Material for this manuscript includes the following:
(available at science.sciencemag.org/cgi/content/full/science.abg0821/DC1)

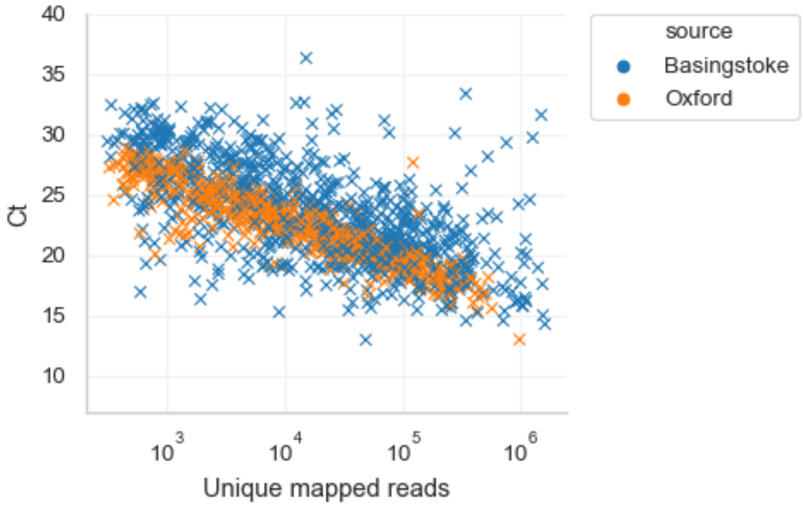
- MDAR Reproducibility Checklist (PDF)

Figures S1-S8

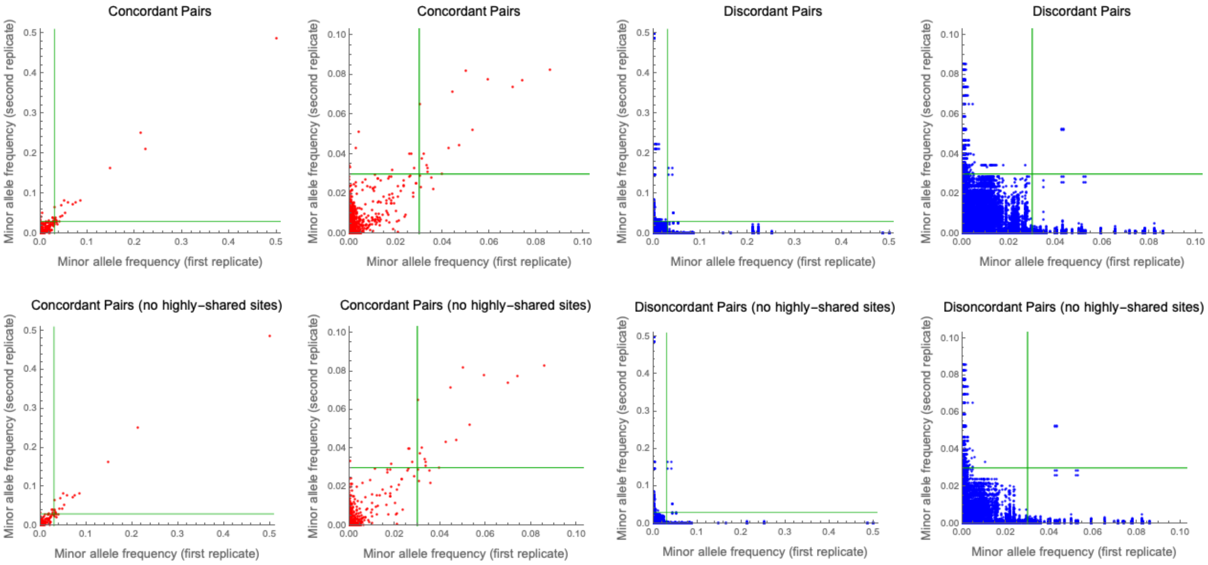
a



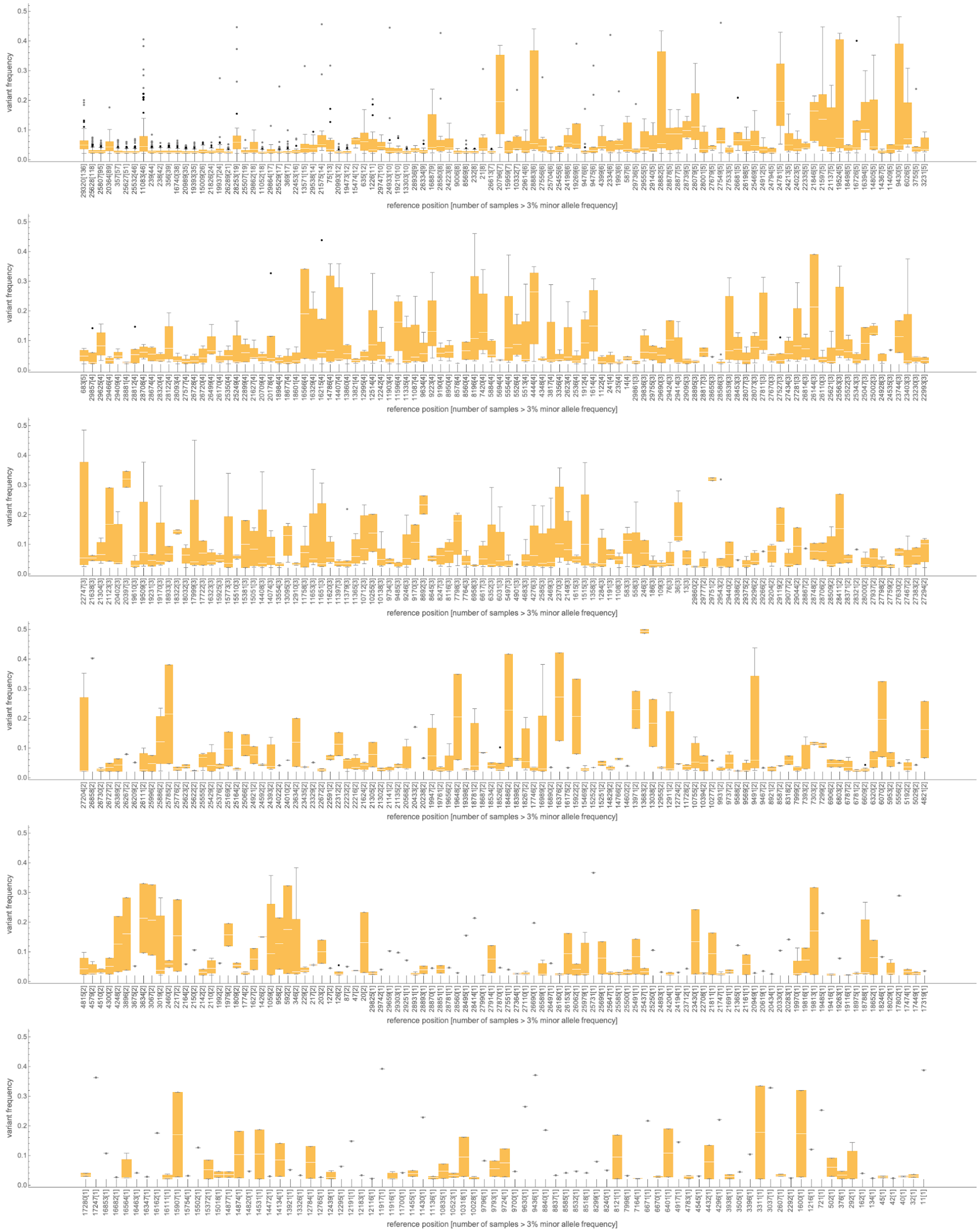
b



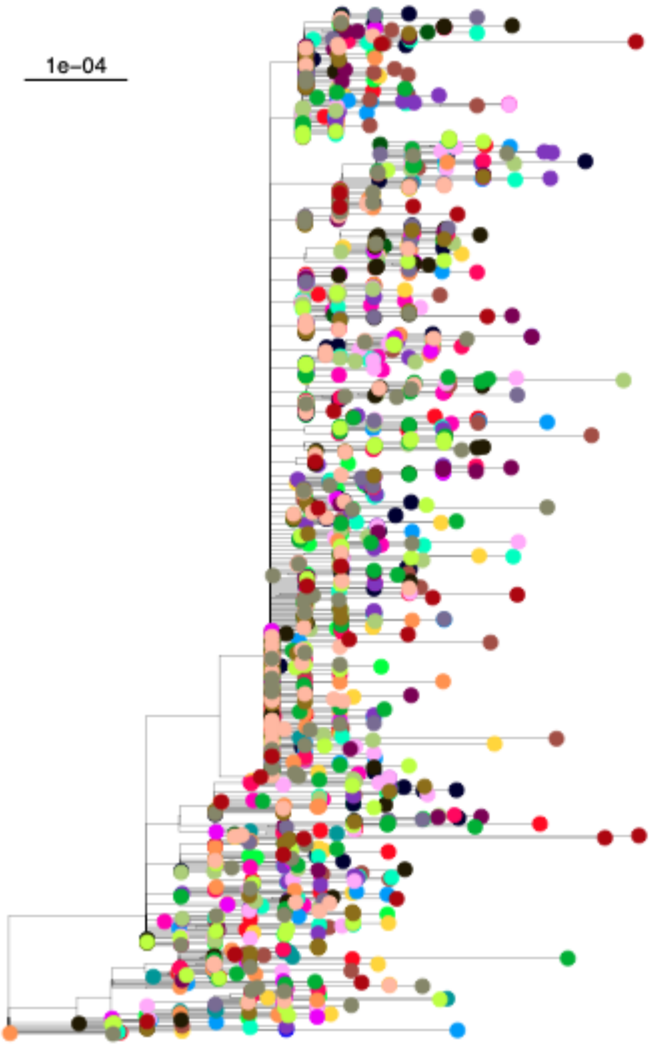
Supplementary Figure 1. (a) Correlation between number of SARS-CoV-2 unique reads and RNA copies/ml for within-batch standard curves for dilution series of positive control RNA. Colour indicates batch. Synthetic SARS-CoV-2 RNA (generated by in vitro transcription by Twist Bioscience) was serially diluted into Universal Human Reference RNA (UHRR) to a final concentration of SARS-CoV-2 RNA of 500,000, 50,000, 5,000, 500, 100 and 0 copies/reaction. Controls were processed and sequenced alongside each batch of samples (batches 3-27). Batches 1 and 2 were processed prior to controls being available and did not have a standard curve. (b) Correlation between nearest available cycle threshold (Ct) value for sequenced clinical samples, as reported by the collecting laboratory, and the number of unique mapped reads. Due to variation in qPCR methodology, Ct values varied substantially between laboratories and over time (*r2* = 0.43 Basingstoke, *r2*=0.81 Oxford, overall *r2*=0.50) Slightly lower Ct values were observed for Basingstoke samples for the same viral load (estimated from the standard curve), consistent with laboratory variation in Ct estimates.



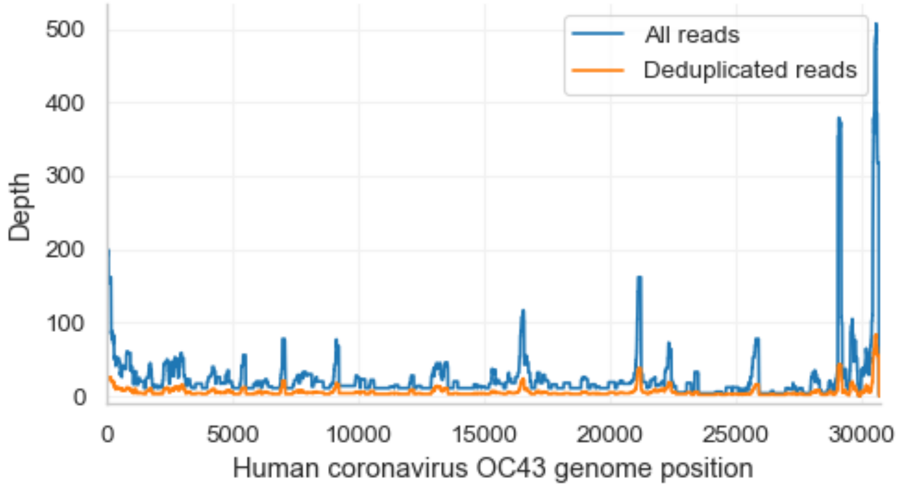
Supplementary Figure 2. Comparison of minor allele frequencies among replicate samples. Data is only included for the 27 replicate pairs where both replicates had more than 50,000 unique mapped reads, and for all sites where minor allele frequency (MAF) >=2% and depth >= 100 in at least one of the 54 replicates. For MAFs > 3%, and excluding highly-shared sites, MAFs are highly reproducible. Concordant pairs: The points represent the MAFs in each of the replicate pairs, for all 27 replicate pairs for all identified sites. If MAFs are reproducible, we expect a positive correlation. Discordant pairs: The points represent the MAFs for all pairwise permutations of replicates for all identified sites, excluding concordant replicates. Unless variants are present in multiple samples, the expectation is for points to be positioned along the axes. Top row includes all sites, whereas the bottom row exclude highly-shared sites (those observed at MAF >=3% in 20 or more samples across the entire dataset). The blue points in the upper-right quadrants represent site 28580, which is present in phylogenetically linked individuals, with two of these included in the 27 replicate pairs. The green line shows MAF 3%.



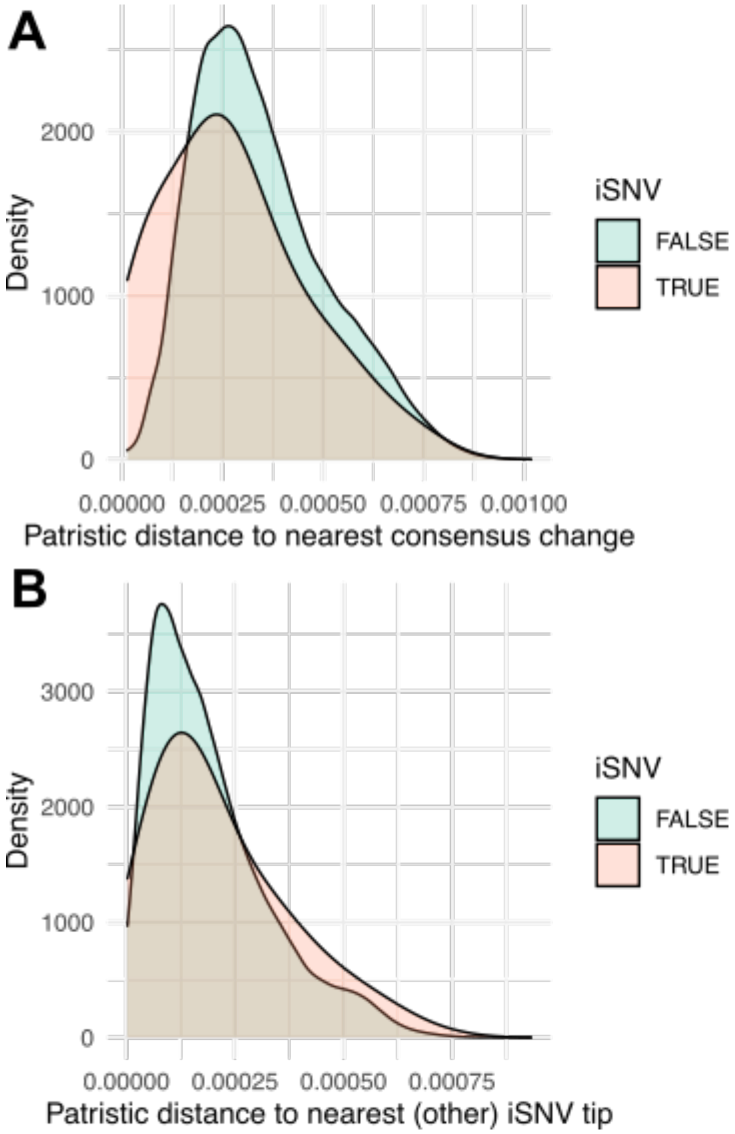
Supplementary Figure 3. Box-whisker plots showing distribution of variant frequencies. All sites variable at minor allele frequency (MAF) > 3% in at least one high-depth sample (>50,000 mapped reads) are included. All variants with MAF > 2% are plotted. The x-axis gives the nucleotide position, with the number of samples MAF > 3% in brackets. All sites are included except the poly A tail, and sites diverse in synthetic controls. The first 18 'highly-shared' sites are either ubiquitous, or only ever seen at low frequencies, suggesting these are non-genomic variants.



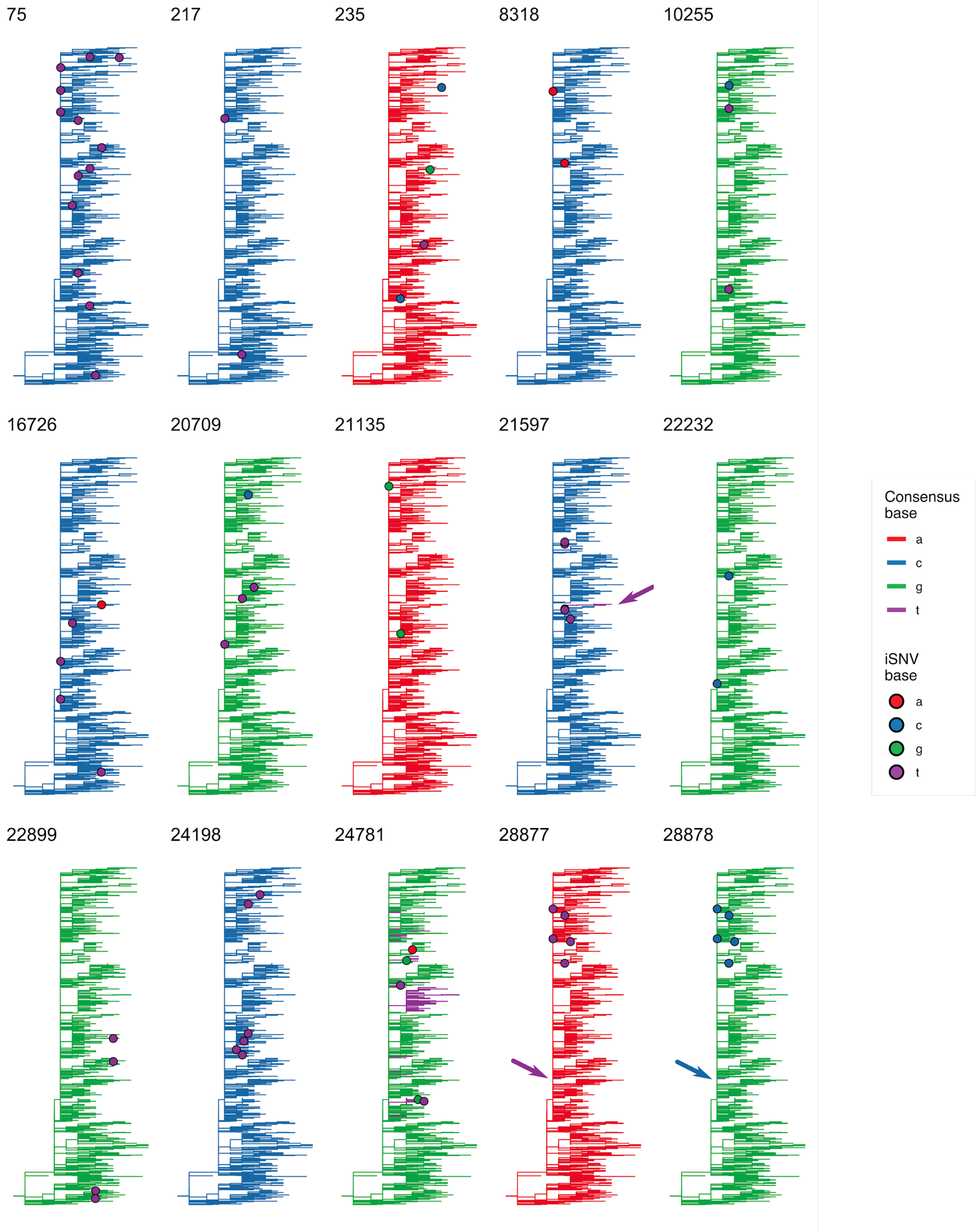
Supplemental Figure 4 - Consensus phylogeny of all 1390 Oxford and Basingstoke samples. Tips are coloured by sequencing batch.



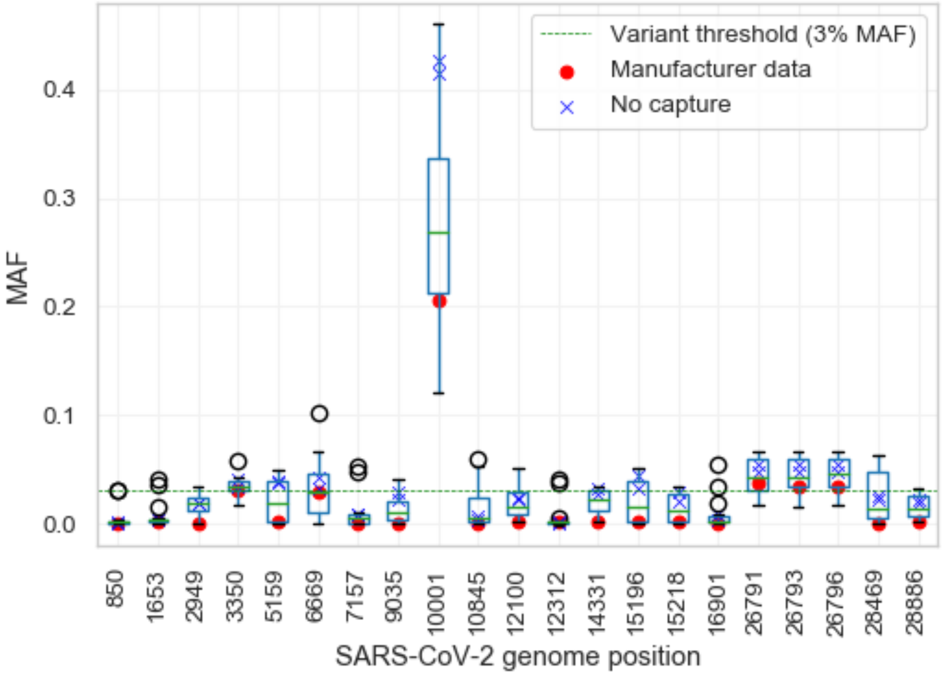
Supplementary Figure 5. Genome coverage for co-infecting human coronavirus OC43 in a SARS-CoV-2-positive sample, OXON-AEC3D. A single co-infection with a non-SARS-CoV-2 circulating coronavirus was detected among a subset of 111 samples analysed with both SARS-CoV-2-specific probes and the Castanet metagenomic respiratory probe panel. Shown in blue are positions of the 2953 proper read pairs mapping to the Castanet reference for OC43, with unique (deduplicated) read depth in orange.



Supplementary Figure 6 - A Across all iSNVs that reach consensus, kernel density plot of the patristic distances from iSNV tips (orange) and other tips (green) to the nearest consensus branch change of the nucleotides involved. **B** Across all iSNVs that do not reach consensus and occur at least twice, kernel density plot of the patristic distances from iSNV tips and other tips to the nearest iSNV tip (other than the tip itself).



Supplementary Figure 7 - The consensus phylogeny coloured by SNP and iSNV for sites mentioned in the main text. Where consensus changes are hard to see they are indicated with an appropriately coloured arrow. Sites 21597, 24781, 28877 and 28878 are the remaining positions where a statistically significant association of iSNV tips with branches with a consensus base change was identified (along with 20796 and 28580). For some of these, coloured arrows indicate the presence of branches with consensus SNPs where this is difficult to see. Sites 22899 and 24198 are Spike variants shown to exhibit reduced sensitivity to convalescent sera (44). The remaining 9 subfigures are for iSNVs which never reach consensus but show a $p<0.025$ for phylogenetic association of iSNV tips using the association index or the mean patristic distance between iSNV tips. While we lack the power to identify these once the Benjamini-Hochberg adjustment is applied, the patterns remain suggestive of transmission of iSNVs by eye.



Supplementary Figure 8. Within-sample diversity assessed in control RNA (Twist Bioscience). Within-sample diversity was assessed in RNA controls sequenced with each sequencing batch (0.5 mln copies per reaction). At all sites where at least 2 replicates had a minor variant with minimum 3% MAF (boxplot), diversity was compared against a set of NGS reads obtained from Twist Bioscience for the ancestral stock of the *in vitro* transcribed RNA used in this study (red circles). Six variants were consistently recovered from both the manufacturer data and the in-batch controls, at positions 3350, 6669, 10001, 26791, 26793, 26796. To check whether the remaining within-host variants arose during the SMARTer library prep or during probe capture, we additionally resequenced two replicates of the Twist RNA without capture (blue crosses), by diluting neat RNA 50:50 v/v in Universal Human Reference RNA (UHRR) and taking a proportion for sequencing, to yield approximately 50,000 copies of the Twist control RNA per sample. We generated SMARTer libraries from these replicates, and sequenced these alongside other samples in separate batches. The two capture-free replicates had the same range of intra-sample variants as were observed in our routinely sequenced controls, implying that any differences from the manufacturer data cannot be explained by probe capture and must be the result of the SMARTer library protocol and/or stochastic variation between our laboratory aliquot and the ancestral RNA stock sequenced by Twist.

Tables S1-S5

Table S1. Baseline characteristics of SARS-CoV-2 samples in our dataset collected by participating hospitals in Oxford and Basingstoke, UK, between 8 March and 10 June 2020. Total number of SARS-CoV-2 genomes includes replicate sequences from the same starting material (swab); number of samples corresponds to unique nasopharyngeal swabs. Participant age and sex excludes the 92 samples from anonymous participants in Oxford. Lineages were determined using Pangolin v2.1.7 (*1*) on 26 January 2021.

	Collecting laboratory	
	Oxford	**Basingstoke**
SARS-CoV-2 genomes, n(%)	552 (39.7)	838 (60.3)
Samples (swabs), n(%)	539 (41.0)	774 (58.9)
Participants, n(%)	539* (43.0)	727 (57.0)
Proportion female	0.60	0.61
Age, median	48	49
(min - max)	(0 - 98)	(0 - 100)
Sampling date, median	10-Apr-20	09-Apr-20
(min - max)	(16-Mar-2020 - 06-May-2020)	(06-Mar-2020 - 10-Jun-2020)
Ct value, median	22.2	23.2
(min - max)	(13.0 3 - 28.89)	(13.0 - 36.3)
SARS-CoV-2 lineage, n(%) - first sample per participant:		
B.1.1.119	216 (40.07)	261 (35.90)
B.1	73 (13.54)	84 (11.55)
B.1.1.194	1 (0.19)	142 (19.53)
B.1.1.1	38 (7.05)	22 (3.03)
B.1.1.220	3 (0.56)	33 (4.54)
B.1.1.175	20 (3.71)	15 (2.06)
B.40	11 (2.04)	22 (3.03)
B	15 (2.78)	12 (1.65)
B.1.1.10	4 (0.74)	20 (2.75)
B.39	2 (0.37)	21 (2.89)
B.1.1.114	15 (2.78)	2 (0.28)
B.1.1.257	15 (2.78)	1 (0.14)

B.1.1.51	1 (0.19)	14 (1.93)
B.1.13	4 (0.74)	10 (1.38)
B.1.1.289	13 (2.41)	0 (0.00)
B.28	8 (1.48)	4 (0.55)
B.1.391	9 (1.67)	3 (0.41)
B.1.1.217	7 (1.30)	5 (0.69)
B.1.1.269	4 (0.74)	6 (0.83)
B.3	4 (0.74)	5 (0.69)
B.1.1.237	7 (1.30)	0 (0.00)
B.1.321	7 (1.30)	0 (0.00)
B.1.1.270	7 (1.30)	0 (0.00)
B.1.1.104	0 (0.00)	7 (0.96)
B.1.1.164	2 (0.37)	4 (0.55)
B.1.231	6 (1.11)	0 (0.00)
B.1.1.261	5 (0.93)	0 (0.00)
B.1.1.304	5 (0.93)	0 (0.00)
B.1.1.64	4 (0.74)	1 (0.14)
B.1.249	3 (0.56)	0 (0.00)
B.1.379	3 (0.56)	0 (0.00)
B.1.93	2 (0.37)	1 (0.14)
B.1.1.277	0 (0.00)	3 (0.41)
B.1.1.123	2 (0.37)	1 (0.14)
B.31	0 (0.00)	3 (0.41)
B.1.1.90	2 (0.37)	1 (0.14)
B.29	0 (0.00)	2 (0.28)
B.1.147	2 (0.37)	0 (0.00)
B.1.104	2 (0.37)	0 (0.00)
B.1.91	0 (0.00)	2 (0.28)
B.54	2 (0.37)	0 (0.00)
B.1.1.10.2 (L.2)	0 (0.00)	2 (0.28)
B.1.1.273	2 (0.37)	0 (0.00)

B.1.1.208	1 (0.19)	1 (0.14)
B.1.1.49	0 (0.00)	1 (0.14)
B.1.1.211	0 (0.00)	1 (0.14)
B.47	0 (0.00)	1 (0.14)
B.45	1 (0.19)	0 (0.00)
B.35	0 (0.00)	1 (0.14)
B.1.1.106	1 (0.19)	0 (0.00)
B.1.1.127	0 (0.00)	1 (0.14)
B.1.1.166	1 (0.19)	0 (0.00)
B.1.1.180	0 (0.00)	1 (0.14)
B.23	1 (0.19)	0 (0.00)
B.1.98	1 (0.19)	0 (0.00)
B.1.1.183	0 (0.00)	1 (0.14)
B.1.1.216	0 (0.00)	1 (0.14)
B.1.1.71	0 (0.00)	1 (0.14)
B.1.1.256	1 (0.19)	0 (0.00)
B.1.1.264	0 (0.00)	1 (0.14)
B.1.229	1 (0.19)	0 (0.00)
B.1.222	1 (0.19)	0 (0.00)
B.1.199	0 (0.00)	1 (0.14)
B.1.182	0 (0.00)	1 (0.14)
B.1.153	0 (0.00)	1 (0.14)
B.1.1.292	0 (0.00)	1 (0.14)
B.1.1.31	1 (0.19)	0 (0.00)
B.1.1.4	1 (0.19)	0 (0.00)
B.1.1.89	0 (0.00)	1 (0.14)
B.1.1.8	1 (0.19)	0 (0.00)
A.1	0 (0.00)	1 (0.14)
A.2	1 (0.19)	0 (0.00)
A.2.2	0 (0.00)	1 (0.14)

* Includes 92 swabs from anonymous participants

Table S2. Sites masked due to sequenced diversity in Twist synthetic controls (112) or as "highly shared" sites with low-level within-host diversity in more than 20 samples (18). https://github.com/katrinalythgoe/COVIDdiversity (70).

Table S3. Identified within-host variable sites.
Sites with at least one minor allele at frequency >= 3% at depth of at least 100 reads, in a sample depth >=50,000 unique mapped reads. Throughout, "samples" refers to all sequencing runs, and therefore includes replicates in the totals. n_notPopConsensus refers to the number of samples in which the minor variant is not the population-level consensus (most common consensus allele); n_SNPs gives the number of SNPs on the tree; homoplasy is "TRUE" if a homoplasy exists on the tree; maf_median is the median minor allele frequency (MAF) for all samples with >2% MAF; maf_IQR is the inter-quartile range of minor allele frequencies for all samples with >2% MAF. https://github.com/katrinalythgoe/COVIDdiversity (70).

Table S4. Summary data for the household transmission analysis. Number of intra-host nucleotide variants sites (iSNVs) in source and recipient pairs (minor allele frequency >3%), and number of mapped reads. Stochastic sampling effects elevate the true number of iSNVs in low viral load samples (<50,000 mapped reads).

Direction inferred from collection dates (at least 7 days difference)

Household number	iSNVs (source)	iSNVs (recipient)	Mapped reads (source)	Mapped reads (recipient)
1	28	58	[10000,1e+05)	[100,1000)
3	45	0	[100,1000)	[10000,1e+05)
4	6	4	[100,1000)	[1000,10000)
8	0	3	[1e+06,1e+07)	[1e+05,1e+06)
9	1	37	[1e+05,1e+06)	[10000,1e+05)
13	3	2	[1e+05,1e+06)	[10000,1e+05)
14	2	43	[10000,1e+05)	[100,1000)

No direction inferred

Household number	iSNVs (patient 1)	iSNVs (patient 2)	Mapped reads (patient 1)	Mapped reads (patient 2)
2	2	4	[1e+05,1e+06)	[1e+05,1e+06)
5	31	35	[1000,10000)	[10000,1e+05)
6	3	0	[1e+06,1e+07)	[1e+06,1e+07)
7	34	2	[1000,10000)	[100,1000)
10	0	5	[10000,1e+05)	[100,1000)
11	19	6	[10000,1e+05)	[10000,1e+05)
12	1	1	[10000,1e+05)	[1e+05,1e+06)
16	1	4	[1e+05,1e+06)	[10000,1e+05)

Table S5. iSNVs and dn/ds by gene and over the whole genome, counting each variant only once regardless of frequency. All genome positions are relative to the Wuhan-Hu-1 reference sequence. iSNVs at the 18 “highly shared” sites and those identified from the synthetic controls are excluded, as are those in the poly-A tail (positions 29865-29903). Note that due to gene overlap and non-coding intergenic regions, the total number of iSNVs (781) cannot be obtained as the sum of any column in this table, even if the rows for nonstructural proteins in ORF1ab are excluded. * nsp12 overlaps the boundary between ORF1a and ORF1b. ** Intergenic regions are excluded from this row.

Gene	Start	End	Length	Unique iSNVs			Alternative $\mathrm{d}n/\mathrm{d}s$ (95% CI)
				Total	NS	S	
5’UTR	1	265	265	40	-	-	-
ORF1a	266	13483	13218	205	125	80	0.44 (0.33, 0.58)
nsp1	266	805	540	12	5	7	0.22 (0.06, 0.68)
nsp2	806	2719	1914	36	21	15	0.4 (0.21, 0.79)
nsp3	2720	8554	5835	74	46	28	0.45 (0.29, 0.74)
nsp4	8555	10054	1500	35	21	14	0.43 (0.22, 0.87)
nsp5A	10055	10972	918	10	9	1	2.55 (0.48, 46.97)
nsp6	10973	11842	870	17	10	7	0.39 (0.15, 1.09)
nsp7	11843	12091	249	3	1	2	0.14 (0.01, 1.48)
nsp8	12092	12685	594	7	4	3	0.36 (0.08, 1.83)
nsp9	12686	13024	339	7	4	3	0.41 (0.09, 2.08)
nsp10	13025	13441	417	4	4	0	∞ (0.46, ∞)
nsp12*	13442	16236	2795	42	23	19	0.32 (0.17, 0.59)
ORF1b	13468	21555	8088	127	73	54	0.37 (0.26, 0.52)
nsp13	16237	18039	1803	25	14	11	0.37 (0.17, 0.85)
nsp14	18040	19620	1581	30	16	14	0.3 (0.15, 0.63)
nsp15	19621	20658	1038	17	12	5	0.65 (0.24, 2.04)
nsp16	20659	21552	894	13	8	5	0.43 (0.14, 1.43)
S	21563	25384	3822	56	37	19	0.55 (0.32, 0.98)
ORF3a	25393	26220	828	33	28	5	1.6 (0.68, 4.72)
E	26245	26472	228	3	2	1	0.68 (0.07, 14.66)
M	26523	27191	669	11	7	4	0.54 (0.16, 2.05)
ORF6	27202	27387	186	5	4	1	0.97 (0.14, 18.97)
ORF7a	27394	27759	366	12	10	2	1.47 (0.39, 9.58)
ORF7b	27756	27887	132	4	4	0	∞ (0.41, ∞)
ORF8	27894	28259	366	11	7	4	0.47 (0.14, 1.78)
N	28274	29533	1260	50	33	17	0.58 (0.33, 1.06)
ORF10	29558	29674	117	3	2	1	0.53 (0.05, 11.28)
3’UTR	29675	29903	229	26	-	-	-
All coding regions**	266	29674	29256	519	331	188	0.49 (0.41, 0.59)
Full genome	1	29903	22903	781	-	-	-

Oxford Virus Sequencing Group (OVSG) Analysis Group membership

John A Todd, Tanya Golubchik, David Bonsall, Christophe Fraser, Derrick Crook, Tim Peto, Monique Andersson, Katie Jeffery, David Eyre, Timothy Walker, Robert Shaw, Peter Simmonds, Katrina Lythgoe, Luca Ferretti, Matthew Hall, Mariateresa de Cesare, Paolo Piazza, David Buck, Richard Cornall.

COG-UK Full list of consortium names and affiliations

Funding acquisition, leadership, supervision, metadata curation, project administration, samples, logistics, Sequencing, analysis, and Software and analysis tools:
Thomas R Connor [33, 34], and Nicholas J Loman [15].

Leadership, supervision, sequencing, analysis, funding acquisition, metadata curation, project administration, samples, logistics, and visualisation:
Samuel C Robson [68].

Leadership, supervision, project administration, visualisation, samples, logistics, metadata curation and software and analysis tools:
Tanya Golubchik [27].

Leadership, supervision, metadata curation, project administration, samples, logistics sequencing and analysis:
M. Estee Torok [8, 10].

Project administration, metadata curation, samples, logistics, sequencing, analysis, and software and analysis tools:
William L Hamilton [8, 10].

Leadership, supervision, samples logistics, project administration, funding acquisition sequencing and analysis:
David Bonsall [27].

Leadership and supervision, sequencing, analysis, funding acquisition, visualisation and software and analysis tools:
Ali R Awan [74].

Leadership and supervision, funding acquisition, sequencing, analysis, metadata curation, samples and logistics:
Sally Corden[33] .

Leadership supervision, sequencing analysis, samples, logistics, and metadata curation:
Ian Goodfellow [11].

Leadership, supervision, sequencing, analysis, samples, logistics, and Project administration:
Darren L Smith [60, 61].

Project administration, metadata curation, samples, logistics, sequencing and analysis:
Martin D Curran [14], and Surendra Parmar [14]**.**

Samples, logistics, metadata curation, project administration sequencing and analysis:
James G Shepherd [21].

Sequencing, analysis, project administration, metadata curation and software and analysis tools:
Matthew D Parker [38] and Dinesh Aggarwal [1, 2, 3].

Leadership, supervision, funding acquisition, samples, logistics, and metadata curation:
Catherine Moore [33] .

Leadership, supervision, metadata curation, samples, logistics, sequencing and analysis:
Derek J Fairley[6, 88], Matthew W Loose [54], and Joanne Watkins [33].

Metadata curation, sequencing, analysis, leadership, supervision and software and analysis tools:
Matthew Bull [33] , and Sam Nicholls [15] .

Leadership, supervision, visualisation, sequencing, analysis and software and analysis tools:
David M Aanensen [1, 30].

Sequencing, analysis, samples, logistics, metadata curation, and visualisation:
Sharon Glaysher [70] .

Metadata curation, sequencing, analysis, visualisation, software and analysis tools:
Matthew Bashton [60], and Nicole Pacchiarini [33].

Sequencing, analysis, visualisation, metadata curation, and software and analysis tools:
Anthony P Underwood [1, 30].

Funding acquisition, leadership, supervision and project administration:
Thushan I de Silva [38], and Dennis Wang [38].

Project administration, samples, logistics, leadership and supervision:
Monique Andersson[28] , Anoop J Chauhan [70], Mariateresa de Cesare [26], Catherine Ludden [1,3] , and Tabitha W Mahungu [91].

Sequencing, analysis, project administration and metadata curation:
Rebecca Dewar [20], and Martin P McHugh [20].

Samples, logistics, metadata curation and project administration:
Natasha G Jesudason [21], Kathy K Li MBBCh [21], Rajiv N Shah [21], and Yusri Taha [66].

Leadership, supervision, funding acquisition and metadata curation:
Kate E Templeton [20].

Leadership, supervision, funding acquisition, sequencing and analysis:
Simon Cottrell [33], Justin O'Grady [51], Andrew Rambaut [19], and Colin P Smith[93].

Leadership, supervision, metadata curation, sequencing and analysis:
Matthew T.G. Holden [87], and Emma C Thomson [21].

Leadership, supervision, samples, logistics and metadata curation:
Samuel Moses [81, 82].

Sequencing, analysis, leadership, supervision, samples and logistics:
Meera Chand [7], Chrystala Constantinidou [71], Alistair C Darby [46], Julian A Hiscox [46], Steve Paterson [46], and Meera Unnikrishnan [71].

Sequencing, analysis, leadership and supervision and software and analysis tools:
Andrew J Page [51], and Erik M Volz [96].

Samples, logistics, sequencing, analysis and metadata curation:
Charlotte J Houldcroft [8], Aminu S Jahun [11], James P McKenna [88], Luke W Meredith [11], Andrew Nelson [61], Sarojini Pandey [72], and Gregory R Young [60].

Sequencing, analysis, metadata curation, and software and analysis tools:
Anna Price [34], Sara Rey [33], Sunando Roy [41], Ben Temperton[49], and Matthew Wyles [38].

Sequencing, analysis, metadata curation and visualisation:
Stefan Rooke[19], and Sharif Shaaban [87].

Visualisation, sequencing, analysis and software and analysis tools:
Helen Adams [35], Yann Bourgeois [69], Katie F Loveson [68], Áine O'Toole [19], and Richard Stark [71].

Project administration, leadership and supervision:
Ewan M Harrison [1, 3], David Heyburn [33], and Sharon J Peacock [2, 3]

Project administration and funding acquisition:
David Buck [26], and Michaela John[36]

Sequencing, analysis and project administration:
Dorota Jamrozy [1], and Joshua Quick [15]

Samples, logistics, and project administration:
Rahul Batra [78], Katherine L Bellis [1, 3], Beth Blane [3], Sophia T Girgis [3], Angie Green [26], Anita Justice [28], Mark Kristiansen [41], and Rachel J Williams [41].

Project administration, software and analysis tools:
Radoslaw Poplawski[15].

Project administration and visualisation:
Garry P Scarlett [69].

Leadership, supervision, and funding acquisition:
John A Todd [26], Christophe Fraser [27], Judith Breuer [40,41], Sergi Castellano [41], Stephen L Michell [49], Dimitris Gramatopoulos [73], and Jonathan Edgeworth [78].

Leadership, supervision and metadata curation:
Gemma L Kay [51].

Leadership, supervision, sequencing and analysis:
Ana da Silva Filipe [21], Aaron R Jeffries [49], Sascha Ott [71], Oliver Pybus [24], David L Robertson [21], David A Simpson [6], and Chris Williams [33].

Samples, logistics, leadership and supervision:
Cressida Auckland [50], John Boyes [83], Samir Dervisevic [52], Sian Ellard [49, 50], Sonia Goncalves[1], Emma J Meader [51], Peter Muir [2], Husam Osman [95], Reenesh Prakash [52], Venkat Sivaprakasam [18], and Ian B Vipond [2].

Leadership, supervision and visualisation
Jane AH Masoli [49, 50].

Sequencing, analysis and metadata curation

Nabil-Fareed Alikhan[51], Matthew Carlile[54], Noel Craine[33], Sam T Haldenby[46], Nadine Holmes[54], Ronan A Lyons[37], Christopher Moore[54], Malorie Perry[33] , Ben Warne[80], and Thomas Williams[19].

Samples, logistics and metadata curation:

Lisa Berry[72], Andrew Bosworth[95] , Julianne Rose Brown[40], Sharon Campbell[67], Anna Casey[17], Gemma Clark[56], Jennifer Collins[66], Alison Cox[43, 44] , Thomas Davis[84], Gary Eltringham[66], Cariad Evans[38, 39] , Clive Graham[64], Fenella Halstead[18], Kathryn Ann Harris[40], Christopher Holmes[58], Stephanie Hutchings[2] , Miren Iturriza-Gomara[46], Kate Johnson[38, 39], Katie Jones[72], Alexander J Keeley[38], Bridget A Knight[49, 50], Cherian Koshy[90], Steven Liggett[63], Hannah Lowe[81] , Anita O Lucaci[46] , Jessica Lynch[25, 29] , Patrick C McClure[55], Nathan Moore[31] , Matilde Mori[25, 29, 32] , David G Partridge[38, 39] , Pinglawathee Madona[43, 44] , Hannah M Pymont[2], Paul Anthony Randell[43, 44] , Mohammad Raza[38, 39] , Felicity Ryan[81] , Robert Shaw[28], Tim J Sloan[57], and Emma Swindells[65] .

Sequencing, analysis, Samples and logistics:

Alexander Adams[33], Hibo Asad[33], Alec Birchley[33] , Tony Thomas Brooks[41], Giselda Bucca[93], Ethan Butcher[70], Sarah L Caddy[13], Laura G Caller[2, 3, 12] , Yasmin Chaudhry[11], Jason Coombes[33], Michelle Cronin[33], Patricia L Dyal[41], Johnathan M Evans[33], Laia Fina[33], Bree Gatica-Wilcox[33], Iliana Georgana[11], Lauren Gilbert[33] , Lee Graham[33], Danielle C Groves[38], Grant Hall[11], Ember Hilvers[33], Myra Hosmillo[11], Hannah Jones[33], Sophie Jones[33], Fahad A Khokhar[13] , Sara Kumziene-Summerhayes[33], George MacIntyre-Cockett[26], Rocio T Martinez Nunez[94], Caoimhe McKerr[33], Claire McMurray[15], Richard Myers[7], Yasmin Nicole Panchbhaya[41], Malte L Pinckert[11] , Amy Plimmer[33] , Joanne Stockton[15] , Sarah Taylor[33] , Alicia Thornton[7] , Amy Trebes[26] , Alexander J Trotter[51] ,Helena Jane Tutill[41] ,Charlotte A Williams[41] , Anna Yakovleva[11] and Wen C Yew[62].

Sequencing, analysis and software and analysis tools:

Mohammad T Alam[71], Laura Baxter[71], Olivia Boyd[96] , Fabricia F. Nascimento[96], Timothy M Freeman[38], Lily Geidelberg[96], Joseph Hughes[21], David Jorgensen[96], Benjamin B Lindsey[38], Richard J Orton[21] , Manon Ragonnet-Cronin[96] Joel Southgate[33, 34,] and Sreenu Vattipally[21].

Samples, logistics and software and analysis tools:

Igor Starinskij[23].

Visualisation and software and analysis tools:

Joshua B Singer[21] , Khalil Abudahab[1, 30], Leonardo de Oliveira Martins[51] , Thanh Le-Viet[51] ,Mirko Menegazzo[30] ,Ben EW Taylor[1, 30], and Corin A Yeats[30].

Project Administration:

Sophie Palmer [3], Carol M Churcher [3] , Alisha Davies [33], Elen De Lacy [33], Fatima Downing [33], Sue Edwards [33] , Nikki Smith [38] , Francesc Coll [97] , Nazreen F Hadjirin [3] and Frances Bolt [44, 45].

Leadership and supervision:

Alex Alderton[1], Matt Berriman[1], Ian G Charles [51], Nicholas Cortes [31], Tanya Curran [88], John Danesh[1], Sahar Eldirdiri [84], Ngozi Elumogo [52], Andrew Hattersley [49, 50], Alison Holmes [44, 45], Robin Howe [33], Rachel Jones [33], Anita Kenyon [84], Robert A Kingsley [51], Dominic Kwiatkowski [1, 9], Cordelia Langford[1], Jenifer Mason[48], Alison E Mather [51], Lizzie Meadows [51], Sian Morgan [36], James Price [44, 45], Trevor I Robinson [48], Giri Shankar [33] , John Wain [51], and Mark A Webber [51].

Metadata curation:

Declan T Bradley [5, 6], Michael R Chapman [1, 3, 4] , Derrick Crooke [28] , David Eyre [28], Martyn Guest [34] , Huw Gulliver [34], Sarah Hoosdally [28], Christine Kitchen [34], Ian Merrick [34], Siddharth Mookerjee [44, 45], Robert Munn [34] , Timothy Peto [28], Will Potter [52], Dheeraj K Sethi [52], Wendy Smith [56] , Luke B Snell [75, 94], Rachael Stanley [52] , Claire Stuart [52] and Elizabeth Wastenge[20].

Sequencing and analysis:

Erwan Acheson [6] , Safiah Afifi [36] , Elias Allara [2, 3] , Roberto Amato [1], Adrienn Angyal [38], Elihu Aranday-Cortes [21] , Cristina Ariani [1], Jordan Ashworth [19], Stephen Attwood [24], Alp Aydin [51], David J Baker [51], Carlos E Balcazar [19], Angela Beckett [68] Robert Beer [36], Gilberto Betancor [76], Emma Betteridge [1] , David Bibby [7] , Daniel Bradshaw[7] , Catherine Bresner [34], Hannah E Bridgewater [71] , Alice Broos [21], Rebecca Brown [38] , Paul E Brown [71], Kirstyn Brunker [22] , Stephen N Carmichael [21] , Jeffrey K. J. Cheng [71], Dr Rachel Colquhoun [19], Gavin Dabrera [7] , Johnny Debebe [54], Eleanor Drury [1], Louis du Plessis [24] , Richard Eccles [46], Nicholas Ellaby [7], Audrey Farbos [49], Ben Farr [1], Jacqueline Findlay [41] , Chloe L Fisher [74], Leysa Marie Forrest [41], Sarah Francois [24], Lucy R. Frost [71], William Fuller[34] , Eileen Gallagher [7], Michael D Gallagher [19] , Matthew Gemmell [46], Rachel AJ Gilroy [51], Scott Goodwin [1], Luke R Green [38], Richard Gregory [46] , Natalie Groves [7], James W Harrison [49], Hassan Hartman [7] , Andrew R Hesketh [93],Verity Hill [19], Jonathan Hubb [7], Margaret Hughes[46] , David K Jackson [1] , Ben Jackson [19], Keith James [1] ,Natasha Johnson [21] ,Ian Johnston [1], Jon-Paul Keatley [1], Moritz Kraemer [24], Angie Lackenby [7], Mara Lawniczak [1] , David Lee [7], Rich Livett [1], Stephanie Lo [1], Daniel Mair [21], Joshua Maksimovic [36], Nikos Manesis [7] , Robin Manley [49], Carmen Manso [7], Angela Marchbank [34] , Inigo Martincorena [1] , Tamyo Mbisa [7], Kathryn McCluggage [36], JT McCrone [19], Shahjahan Miah [7] , Michelle L Michelsen [49], Mari Morgan [33], Gaia Nebbia [78],Charlotte Nelson [46] ,Jenna Nichols [21] ,Paola Niola [41] , Kyriaki Nomikou [21] ,Steve Palmer [1] , Naomi Park [1], Yasmin A Parr [1] , Paul J Parsons [38] , Vineet Patel [7] , Minal Patel [1] ,Clare Pearson [2, 1], Steven Platt [7] ,Christoph Puethe [1], Mike Quail [1],Jayna Raghwani [24] , Lucille Rainbow [46] ,Shavanthi Rajatileka [1], Mary Ramsay [7] , Paola C Resende Silva [41, 42], Steven Rudder 51, Chris Ruis [3] , Christine M Sambles [49], Fei Sang [54], Ulf Schaefer[7], Emily Scher [19], Carol Scott [1] ,Lesley Shirley [1], Adrian W Signell [76], John Sillitoe [1] ,Christen Smith [1] ,Dr Katherine L Smollett [21] ,Karla Spellman [36]

,Thomas D Stanton[19], David J Studholme[49] ,Grace Taylor-Joyce[71] ,Ana P Tedim[51], Thomas Thompson[6], Nicholas M Thomson[51], Scott Thurston[1] , Lily Tong[21], Gerry Tonkin-Hill[1], Rachel M Tucker[38] , Edith E Vamos[4], Tetyana Vasylyeva[24], Joanna Warwick-Dugdale[49] , Danni Weldon[1], Mark Whitehead[46], David Williams[7], Kathleen A Williamson[19],Harry D Wilson[76],Trudy Workman[34], Muhammad Yasir[51], Xiaoyu Yu[19], and Alex Zarebski[24].

Samples and logistics:
Evelien M Adriaenssens[51], Shazaad S Y Ahmad[2,47] , Adela Alcolea-Medina[59,77], John Allan[60], Patawee Asamaphan[21], Laura Atkinson[40], Paul Baker[63], Jonathan Ball[55], Edward Barton[64], Mathew A Beale[1], Charlotte Beaver[1], Andrew Beggs[16], Andrew Bell[51], Duncan J Berger[1], Louise Berry.[56], Claire M Bewshea[49], Kelly Bicknell[70], Paul Bird[58], Chloe Bishop[7] , Tim Boswell[56], Cassie Breen[48], Sarah K Buddenborg[1], Shirelle Burton-Fanning[66] , Vicki Chalker[7], Joseph G Chappell[55], Themoula Charalampous[78,94], Claire Cormie[3], Nick Cortes[29,25], Lindsay J Coupland[52], Angela Cowell[48] , Rose K Davidson[53] , Joana Dias[3], Maria Diaz[51] , Thomas Dibling[1], Matthew J Dorman[1], Nichola Duckworth[57], Scott Elliott[70], Sarah Essex[63], Karlie Fallon[58] , Theresa Feltwell[8], Vicki M Fleming[56], Sally Forrest[3], Luke Foulser[1], Maria V Garcia-Casado[1], Artemis Gavriil[41], Ryan P George[47], Laura Gifford[33], Harmeet K Gill[3], Jane Greenaway[65], Luke Griffith[53], Ana Victoria Gutierrez[51], Antony D Hale[85], Tanzina Haque[91], Katherine L Harper[85], Ian Harrison[7] , Judith Heaney[89], Thomas Helmer[58], Ellen E Higginson[3] , Richard Hopes[2], Hannah C Howson-Wells[56], Adam D Hunter[1], Robert Impey[70], Dianne Irish-Tavares[91], David A Jackson[1] , Kathryn A Jackson[46], Amelia Joseph[56], Leanne Kane[1], Sally Kay[1], Leanne M Kermack[3], Manjinder Khakh[56], Stephen P Kidd[29,25,31], Anastasia Kolyva[51], Jack CD Lee[40], Laura Letchford[1] , Nick Levene[79], Lisa J Levett[89], Michelle M Lister[56], Allyson Lloyd[70], Joshua Loh[60] , Louissa R Macfarlane-Smith[85], Nicholas W Machin[2,47], Mailis Maes[3], Samantha McGuigan[1], Liz McMinn[1], Lamia Mestek-Boukhibar[41], Zoltan Molnar[6], Lynn Monaghan[79], Catrin Moore[27], Plamena Naydenova[3], Alexandra S Neaverson[1], Rachel Nelson[1], Marc O Niebel[21] , Elaine O'Toole[48] , Debra Padgett[64], Gaurang Patel[1] , Brendan AI Payne[66], Liam Prestwood[1], Veena Raviprakash[67], Nicola Reynolds[86], Alex Richter[16], Esther Robinson[95], Hazel A Rogers[1], Aileen Rowan[96], Garren Scott[64], Divya Shah[40], Nicola Sheriff[67], Graciela Sluga, Emily Souster[1], Michael Spencer-Chapman[1], Sushmita Sridhar[1,3], Tracey Swingler[53], Julian Tang[58], Graham P Taylor[96], Theocharis Tsoleridis[55], Lance Turtle[46], Sarah Walsh[57], Michelle Wantoch[86], Joanne Watts[48] , Sheila Waugh[66], Sam Weeks[41], Rebecca Williams[31], Iona Willingham[56], Emma L Wise[25,29,31], Victoria Wright[54], Sarah Wyllie[70], and Jamie Young[3].

Software and analysis tools
Amy Gaskin[33], Will Rowe[15], and Igor Siveroni[96].

Visualisation:
Robert Johnson[96].

1 Wellcome Sanger Institute, **2** Public Health England, **3** University of Cambridge, **4** Health Data Research UK, Cambridge, **5** Public Health Agency, Northern Ireland ,**6** Queen's University Belfast **7** Public Health England Colindale, **8** Department of Medicine, University of Cambridge, **9** University of Oxford, **10** Departments of Infectious Diseases and Microbiology, Cambridge University Hospitals NHS Foundation Trust; Cambridge, UK, **11** Division of Virology, Department of Pathology, University of Cambridge, **12** The Francis Crick Institute, **13** Cambridge Institute for Therapeutic Immunology and Infectious Disease, Department of Medicine, **14** Public Health England, Clinical Microbiology and Public Health Laboratory, Cambridge, UK, **15** Institute of Microbiology and Infection, University of Birmingham, **16** University of Birmingham, **17** Queen Elizabeth Hospital, **18** Heartlands Hospital, **19** University of Edinburgh, **20** NHS Lothian, **21** MRC-University of Glasgow Centre for Virus Research, **22** Institute of Biodiversity, Animal Health & Comparative Medicine, University of Glasgow, **23** West of Scotland Specialist Virology Centre, **24** Dept Zoology, University of Oxford, **25** University of Surrey, **26** Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, **27** Big Data Institute, Nuffield Department of Medicine, University of Oxford, **28** Oxford University Hospitals NHS Foundation Trust, **29** Basingstoke Hospital, **30** Centre for Genomic Pathogen Surveillance, University of Oxford, **31** Hampshire Hospitals NHS Foundation Trust, **32** University of Southampton, **33** Public Health Wales NHS Trust, **34** Cardiff University, **35** Betsi Cadwaladr University Health Board, **36** Cardiff and Vale University Health Board, **37** Swansea University, **38** University of Sheffield, **39** Sheffield Teaching Hospitals, **40** Great Ormond Street NHS Foundation Trust, **41** University College London, **42** Oswaldo Cruz Institute, Rio de Janeiro **43** North West London Pathology, **44** Imperial College Healthcare NHS Trust, **45** NIHR Health Protection Research Unit in HCAI and AMR, Imperial College London, **46** University of Liverpool, **47** Manchester University NHS Foundation Trust, **48** Liverpool Clinical Laboratories, **49** University of Exeter, **50** Royal Devon and Exeter NHS Foundation Trust, **51** Quadram Institute Bioscience, University of East Anglia, **52** Norfolk and Norwich University Hospital, **53** University of East Anglia, **54** Deep Seq, School of Life Sciences, Queens Medical Centre, University of Nottingham, **55** Virology, School of Life Sciences, Queens Medical Centre, University of Nottingham, **56** Clinical Microbiology Department, Queens Medical Centre, **57** PathLinks, Northern Lincolnshire & Goole NHS Foundation Trust, **58** Clinical Microbiology, University Hospitals of Leicester NHS Trust, **59** Viapath, **60** Hub for Biotechnology in the Built Environment, Northumbria University, **61** NU-OMICS Northumbria University, **62** Northumbria University, **63** South Tees Hospitals NHS Foundation Trust, **64** North Cumbria Integrated Care NHS Foundation Trust, **65** North Tees and Hartlepool NHS Foundation Trust, **66** Newcastle Hospitals NHS Foundation Trust, **67** County Durham and Darlington NHS Foundation Trust, **68** Centre for Enzyme Innovation, University of Portsmouth, **69** School of Biological Sciences, University of Portsmouth, **70** Portsmouth Hospitals NHS Trust, **71** University of Warwick, **72** University Hospitals Coventry and Warwickshire, **73** Warwick Medical School and Institute of Precision Diagnostics, Pathology, UHCW NHS Trust, **74** Genomics Innovation Unit, Guy's and St. Thomas' NHS Foundation Trust, **75** Centre for Clinical Infection & Diagnostics Research, St. Thomas' Hospital and Kings College London, **76** Department of Infectious Diseases, King's College London, **77** Guy's and St. Thomas’ Hospitals NHS Foundation Trust, **78** Centre for Clinical Infection and Diagnostics Research, Department of Infectious Diseases, Guy's and St Thomas' NHS Foundation Trust, **79** Princess Alexandra Hospital Microbiology Dept. , **80** Cambridge University Hospitals NHS Foundation Trust, **81** East Kent Hospitals University NHS Foundation Trust, **82** University of Kent, **83** Gloucestershire Hospitals NHS Foundation Trust, **84** Department of Microbiology, Kettering General Hospital, **85** National Infection Service, PHE and Leeds Teaching Hospitals Trust, **86** Cambridge Stem Cell Institute, University of Cambridge, **87** Public Health Scotland, 88 Belfast Health & Social Care Trust, **89** Health Services Laboratories, **90** Barking, Havering and Redbridge University Hospitals NHS Trust, **91** Royal Free NHS Trust, **92** Maidstone and Tunbridge Wells NHS Trust, **93** University of Brighton, **94** Kings College London, **95** PHE Heartlands, **96** Imperial College London, **97** Department of Infection Biology, London School of Hygiene and Tropical Medicine.

References and Notes

1. A. Rambaut, E. C. Holmes, Á. O'Toole, V. Hill, J. T. McCrone, C. Ruis, L. du Plessis, O. G. Pybus, A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **5**, 1403–1407 (2020). doi:10.1038/s41564-020-0770-5 Medline

2. J. Hadfield, C. Megill, S. M. Bell, J. Huddleston, B. Potter, C. Callender, P. Sagulenko, T. Bedford, R. A. Neher, Nextstrain: Real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018). doi:10.1093/bioinformatics/bty407 Medline

3. J. Lu, L. du Plessis, Z. Liu, V. Hill, M. Kang, H. Lin, J. Sun, S. François, M. U. G. Kraemer, N. R. Faria, J. T. McCrone, J. Peng, Q. Xiong, R. Yuan, L. Zeng, P. Zhou, C. Liang, L. Yi, J. Liu, J. Xiao, J. Hu, T. Liu, W. Ma, W. Li, J. Su, H. Zheng, B. Peng, S. Fang, W. Su, K. Li, R. Sun, R. Bai, X. Tang, M. Liang, J. Quick, T. Song, A. Rambaut, N. Loman, J. Raghwani, O. G. Pybus, C. Ke, Genomic epidemiology of SARS-CoV-2 in Guangdong Province, China. *Cell* **181**, 997–1003.e9 (2020). doi:10.1016/j.cell.2020.04.023 Medline

4. B. Korber, W. M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, N. Hengartner, E. E. Giorgi, T. Bhattacharya, B. Foley, K. M. Hastie, M. D. Parker, D. G. Partridge, C. M. Evans, T. M. Freeman, T. I. de Silva, C. McDanal, L. G. Perez, H. Tang, A. Moon-Walker, S. P. Whelan, C. C. LaBranche, E. O. Saphire, D. C. Montefiori, A. Angyal, R. L. Brown, L. Carrilero, L. R. Green, D. C. Groves, K. J. Johnson, A. J. Keeley, B. B. Lindsey, P. J. Parsons, M. Raza, S. Rowland-Jones, N. Smith, R. M. Tucker, D. Wang, M. D. Wyles; Sheffield COVID-19 Genomics Group, Tracking changes in SARS-CoV-2 spike: Evidence that D614G increases infectivity of the COVID-19 virus. *Cell* **182**, 812–827.e19 (2020). doi:10.1016/j.cell.2020.06.043 Medline

5. E. Volz, V. Hill, J. T. McCrone, A. Price, D. Jorgensen, Á. O'Toole, J. Southgate, R. Johnson, B. Jackson, F. F. Nascimento, S. M. Rey, S. M. Nicholls, R. M. Colquhoun, A. da Silva Filipe, J. Shepherd, D. J. Pascall, R. Shah, N. Jesudason, K. Li, R. Jarrett, N. Pacchiarini, M. Bull, L. Geidelberg, I. Siveroni, I. Goodfellow, N. J. Loman, O. G. Pybus, D. L. Robertson, E. C. Thomson, A. Rambaut, T. R. Connor, C. Koshy, E. Wise, N. Cortes, J. Lynch, S. Kidd, M. Mori, D. J. Fairley, T. Curran, J. P. McKenna, H. Adams, C. Fraser, T. Golubchik, D. Bonsall, C. Moore, S. L. Caddy, F. A. Khokhar, M. Wantoch, N. Reynolds, B. Warne, J. Maksimovic, K. Spellman, K. McCluggage, M. John, R. Beer, S. Afifi, S. Morgan, A. Marchbank, A. Price, C. Kitchen, H. Gulliver, I. Merrick, J. Southgate, M. Guest, R. Munn, T. Workman, T. R. Connor, W. Fuller, C. Bresner, L. B. Snell, T. Charalampous, G. Nebbia, R. Batra, J. Edgeworth, S. C. Robson, A. Beckett, K. F. Loveson, D. M. Aanensen, A. P. Underwood, C. A. Yeats, K. Abudahab, B. E. W. Taylor, M. Menegazzo, G. Clark, W. Smith, M. Khakh, V. M. Fleming, M. M. Lister, H. C. Howson-Wells, L. Berry, T. Boswell, A. Joseph, I. Willingham, P. Bird, T. Helmer, K. Fallon, C. Holmes, J. Tang, V. Raviprakash, S. Campbell, N. Sheriff, M. W. Loose, N. Holmes, C. Moore, M. Carlile, V. Wright, F. Sang, J. Debebe, F. Coll, A. W. Signell, G. Betancor, H. D. Wilson, T. Feltwell, C. J. Houldcroft, S. Eldirdiri, A. Kenyon, T. Davis, O. Pybus, L. du Plessis, A. Zarebski, J. Raghwani, M. Kraemer, S. Francois, S. Attwood, T. Vasylyeva, M. E. Torok, W. L. Hamilton, I. G. Goodfellow, G. Hall, A. S. Jahun, Y. Chaudhry, M. Hosmillo, M. L. Pinckert, I. Georgana, A. Yakovleva, L. W. Meredith, S. Moses, H. Lowe, F. Ryan, C. L. Fisher, A. R. Awan, J. Boyes, J. Breuer, K. A.

Harris, J. R. Brown, D. Shah, L. Atkinson, J. C. D. Lee, A. Alcolea-Medina, N. Moore, N. Cortes, R. Williams, M. R. Chapman, L. J. Levett, J. Heaney, D. L. Smith, M. Bashton, G. R. Young, J. Allan, J. Loh, P. A. Randell, A. Cox, P. Madona, A. Holmes, F. Bolt, J. Price, S. Mookerjee, A. Rowan, G. P. Taylor, M. Ragonnet-Cronin, F. F. Nascimento, D. Jorgensen, I. Siveroni, R. Johnson, O. Boyd, L. Geidelberg, E. M. Volz, K. Brunker, K. L. Smollett, N. J. Loman, J. Quick, C. McMurray, J. Stockton, S. Nicholls, W. Rowe, R. Poplawski, R. T. Martinez-Nunez, J. Mason, T. I. Robinson, E. O'Toole, J. Watts, C. Breen, A. Cowell, C. Ludden, G. Sluga, N. W. Machin, S. S. Y. Ahmad, R. P. George, F. Halstead, V. Sivaprakasam, E. C. Thomson, J. G. Shepherd, P. Asamaphan, M. O. Niebel, K. K. Li, R. N. Shah, N. G. Jesudason, Y. A. Parr, L. Tong, A. Broos, D. Mair, J. Nichols, S. N. Carmichael, K. Nomikou, E. Aranday-Cortes, N. Johnson, I. Starinskij, A. da Silva Filipe, D. L. Robertson, R. J. Orton, J. Hughes, S. Vattipally, J. B. Singer, A. D. Hale, L. R. Macfarlane-Smith, K. L. Harper, Y. Taha, B. A. I. Payne, S. Burton-Fanning, S. Waugh, J. Collins, G. Eltringham, K. E. Templeton, M. P. McHugh, R. Dewar, E. Wastenge, S. Dervisevic, R. Stanley, R. Prakash, C. Stuart, N. Elumogo, D. K. Sethi, E. J. Meader, L. J. Coupland, W. Potter, C. Graham, E. Barton, D. Padgett, G. Scott, E. Swindells, J. Greenaway, A. Nelson, W. C. Yew, P. C. Resende Silva, M. Andersson, R. Shaw, T. Peto, A. Justice, D. Eyre, D. Crooke, S. Hoosdally, T. J. Sloan, N. Duckworth, S. Walsh, A. J. Chauhan, S. Glaysher, K. Bicknell, S. Wyllie, E. Butcher, S. Elliott, A. Lloyd, R. Impey, N. Levene, L. Monaghan, D. T. Bradley, E. Allara, C. Pearson, P. Muir, I. B. Vipond, R. Hopes, H. M. Pymont, S. Hutchings, M. D. Curran, S. Parmar, A. Lackenby, T. Mbisa, S. Platt, S. Miah, D. Bibby, C. Manso, J. Hubb, M. Chand, G. Dabrera, M. Ramsay, D. Bradshaw, A. Thornton, R. Myers, U. Schaefer, N. Groves, E. Gallagher, D. Lee, D. Williams, N. Ellaby, I. Harrison, H. Hartman, N. Manesis, V. Patel, C. Bishop, V. Chalker, H. Osman, A. Bosworth, E. Robinson, M. T. G. Holden, S. Shaaban, A. Birchley, A. Adams, A. Davies, A. Gaskin, A. Plimmer, B. Gatica-Wilcox, C. McKerr, C. Moore, C. Williams, D. Heyburn, E. De Lacy, E. Hilvers, F. Downing, G. Shankar, H. Jones, H. Asad, J. Coombes, J. Watkins, J. M. Evans, L. Fina, L. Gifford, L. Gilbert, L. Graham, M. Perry, M. Morgan, M. Bull, M. Cronin, N. Pacchiarini, N. Craine, R. Jones, R. Howe, S. Corden, S. Rey, S. Kumziene-Summerhayes, S. Taylor, S. Cottrell, S. Jones, S. Edwards, J. O'Grady, A. J. Page, J. Wain, M. A. Webber, A. E. Mather, D. J. Baker, S. Rudder, M. Yasir, N. M. Thomson, A. Aydin, A. P. Tedim, G. L. Kay, A. J. Trotter, R. A. J. Gilroy, N.-F. Alikhan, L. de Oliveira Martins, T. Le-Viet, L. Meadows, A. Kolyva, M. Diaz, A. Bell, A. V. Gutierrez, I. G. Charles, E. M. Adriaenssens, R. A. Kingsley, A. Casey, D. A. Simpson, Z. Molnar, T. Thompson, E. Acheson, J. A. H. Masoli, B. A. Knight, A. Hattersley, S. Ellard, C. Auckland, T. W. Mahungu, D. Irish-Tavares, T. Haque, Y. Bourgeois, G. P. Scarlett, D. G. Partridge, M. Raza, C. Evans, K. Johnson, S. Liggett, P. Baker, S. Essex, R. A. Lyons, L. G. Caller, S. Castellano, R. J. Williams, M. Kristiansen, S. Roy, C. A. Williams, P. L. Dyal, H. J. Tutill, Y. N. Panchbhaya, L. M. Forrest, P. Niola, J. Findlay, T. T. Brooks, A. Gavriil, L. Mestek-Boukhibar, S. Weeks, S. Pandey, L. Berry, K. Jones, A. Richter, A. Beggs, C. P. Smith, G. Bucca, A. R. Hesketh, E. M. Harrison, S. J. Peacock, S. Palmer, C. M. Churcher, K. L. Bellis, S. T. Girgis, P. Naydenova, B. Blane, S. Sridhar, C. Ruis, S. Forrest, C. Cormie, H. K. Gill, J. Dias, E. E. Higginson, M. Maes, J. Young, L. M. Kermack, N. F. Hadjirin, D. Aggarwal, L. Griffith, T. Swingler, R. K. Davidson, A. Rambaut, T. Williams, C. E. Balcazar, M. D. Gallagher, Á. O'Toole, S. Rooke, B. Jackson, R. Colquhoun, J. Ashworth, V. Hill, J. T. McCrone, E. Scher, X. Yu, K. A. Williamson, T. D. Stanton, S. L. Michell, C. M.

Bewshea, B. Temperton, M. L. Michelsen, J. Warwick-Dugdale, R. Manley, A. Farbos, J. W. Harrison, C. M. Sambles, D. J. Studholme, A. R. Jeffries, A. C. Darby, J. A. Hiscox, S. Paterson, M. Iturriza-Gomara, K. A. Jackson, A. O. Lucaci, E. E. Vamos, M. Hughes, L. Rainbow, R. Eccles, C. Nelson, M. Whitehead, L. Turtle, S. T. Haldenby, R. Gregory, M. Gemmell, D. Kwiatkowski, T. I. de Silva, N. Smith, A. Angyal, B. B. Lindsey, D. C. Groves, L. R. Green, D. Wang, T. M. Freeman, M. D. Parker, A. J. Keeley, P. J. Parsons, R. M. Tucker, R. Brown, M. Wyles, C. Constantinidou, M. Unnikrishnan, S. Ott, J. K. J. Cheng, H. E. Bridgewater, L. R. Frost, G. Taylor-Joyce, R. Stark, L. Baxter, M. T. Alam, P. E. Brown, P. C. McClure, J. G. Chappell, T. Tsoleridis, J. Ball, D. Gramatopoulos, D. Buck, J. A. Todd, A. Green, A. Trebes, G. MacIntyre-Cockett, M. de Cesare, C. Langford, A. Alderton, R. Amato, S. Goncalves, D. K. Jackson, I. Johnston, J. Sillitoe, S. Palmer, M. Lawniczak, M. Berriman, J. Danesh, R. Livett, L. Shirley, B. Farr, M. Quail, S. Thurston, N. Park, E. Betteridge, D. Weldon, S. Goodwin, R. Nelson, C. Beaver, L. Letchford, D. A. Jackson, L. Foulser, L. McMinn, L. Prestwood, S. Kay, L. Kane, M. J. Dorman, I. Martincorena, C. Puethe, J.-P. Keatley, G. Tonkin-Hill, C. Smith, D. Jamrozy, M. A. Beale, M. Patel, C. Ariani, M. Spencer-Chapman, E. Drury, S. Lo, S. Rajatileka, C. Scott, K. James, S. K. Buddenborg, D. J. Berger, G. Patel, M. V. Garcia-Casado, T. Dibling, S. McGuigan, H. A. Rogers, A. D. Hunter, E. Souster, A. S. Neaverson; COG-UK Consortium, Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell* **184**, 64–75.e11 (2021). doi:10.1016/j.cell.2020.11.020 Medline

6. Y. J. Hou, S. Chiba, P. Halfmann, C. Ehre, M. Kuroda, K. H. Dinnon 3rd, S. R. Leist, A. Schäfer, N. Nakajima, K. Takahashi, R. E. Lee, T. M. Mascenik, R. Graham, C. E. Edwards, L. V. Tse, K. Okuda, A. J. Markmann, L. Bartelt, A. de Silva, D. M. Margolis, R. C. Boucher, S. H. Randell, T. Suzuki, L. E. Gralinski, Y. Kawaoka, R. S. Baric, SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* **370**, 1464–1468 (2020). Medline

7. A. Rambaut, N. Loman, O. Pybus, W. Barclay, J. Barrett, A. Carabelli, T. Connor, T. Peacock, D. L. Robertson, E. Volz, on behalf of COVID-19 Genomics Consortium UK (CoG-UK), Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations (2020); 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.

8. S. A. Kemp, W. T. Harvey, S. Lytras, The COVID-19 Genomics UK (COG-UK) consortium, A. M. Carabelli, D. L. Robertson, R. K. Gupta, Recurrent emergence and transmission of a SARS-CoV-2 Spike deletion H69/V70. bioRxiv 422555 [Preprint] 9 February 2021. https://doi.org/10.1101/2020.12.14.422555.

9. E. Volz, S. Mishra, M. Chand, J. C. Barrett, R. Johnson, L. Geidelberg, W. R. Hinsley, D. J. Laydon, G. Dabrera, Á. O'Toole, R. Amato, M. Ragonnet-Cronin, I. Harrison, B. Jackson, C. V. Ariani, O. Boyd, N. J. Loman, J. T. McCrone, S. Gonçalves, D. Jorgensen, R. Myers, V. Hill, D. K. Jackson, K. Gaythorpe, N. Groves, J. Sillitoe, D. P. Kwiatkowski, The COVID-19 Genomics UK (COG-UK) consortium, S. Flaxman, O. Ratmann, S. Bhatt, S. Hopkins, A. Gandy, A. Rambaut, N. M. Ferguson, Transmission of SARS-CoV-2 lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. medRxiv 20249034 [Preprint]. 4 January 2021. https://doi.org/10.1101/2020.12.30.20249034.

10. H. Tegally, E. Wilkinson, M. Giovanetti, A. Iranzadeh, V. Fonseca, J. Giandhari, D.

Doolabh, S. Pillay, E. J. San, N. Msomi, K. Mlisana, A. von Gottberg, S. Walaza, M. Allam, A. Ismail, T. Mohale, A. J. Glass, S. Engelbrecht, G. Van Zyl, W. Preiser, F. Petruccione, A. Sigal, D. Hardie, G. Marais, M. Hsiao, S. Korsman, M.-A. Davies, L. Tyers, I. Mudau, D. York, C. Maslo, D. Goedhals, S. Abrahams, O. Laguda-Akingba, A. Alisoltani-Dehkordi, A. Godzik, C. K. Wibmer, B. T. Sewell, J. Lourenço, L. C. J. Alcantara, S. L. Kosakovsky Pond, S. Weaver, D. Martin, R. J. Lessells, J. N. Bhiman, C. Williamson, T. de Oliveira, Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. medRxiv 20248640 [Preprint]. 22 December 2020. https://doi.org/10.1101/2020.12.21.20248640.

11. N. R. Faria, I. M. Claro, D. Candido, L. A. Moyses Franco, P. S. Andrade, T. M. Coletti, C. A. Silva, F. C. Sales, E. R. Manuli, R. S. Aguiar, N. Gaburo, C. da C. Camilo, N. A. Fraiji, M. A. Esashika Crispin, M. do Perpetuao, A. Rambaut, N. Loman, O. G. Pybus, E. C. Sabino, on behalf of CADDE Genomic Network, Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings (2021); https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586).

12. C. K. Wibmer, F. Ayres, T. Hermanus, M. Madzivhandila, P. Kgagudi, B. E. Lambson, M. Vermeulen, K. van den Berg, T. Rossouw, M. Boswell, V. Ueckermann, S. Meiring, A. von Gottberg, C. Cohen, L. Morris, J. N. Bhiman, P. L. Moore, SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. bioRxiv 427166 [Preprint]. 1 March 2021. https://doi.org/10.1101/2021.01.18.427166.

13. S. Cele, I. Gazy, L. Jackson, S.-H. Hwa, H. Tegally, G. Lustig, J. Giandhari, S. Pillay, E. Wilkinson, Y. Naidoo, F. Karim, Y. Ganga, K. Khan, A. B. Balazs, B. I. Gosnell, W. Hanekom, M.-Y. S. Moosa, NGS-SA, COMMIT-KZN Team, R. J. Lessells, T. de Oliveira, A. Sigal, Escape of SARS-CoV-2 501Y.V2 variants from neutralization by convalescent plasma. medRxiv 21250224 [Preprint]. 27 February 2021. https://doi.org/10.1101/2021.01.26.21250224.

14. K. Wu, A. P. Werner, J. I. Moliva, M. Koch, A. Choi, G. B. E. Stewart-Jones, H. Bennett, S. Boyoglu-Barnum, W. Shi, B. S. Graham, A. Carfi, K. S. Corbett, R. A. Seder, D. K. Edwards, mRNA-1273 vaccine induces neutralizing antibodies against spike mutants from global SARS-CoV-2 variants. bioRxiv 427948 [Preprint]. 25 January 2021. https://doi.org/10.1101/2021.01.25.427948.

15. L. du Plessis, J. T. McCrone, A. E. Zarebski, V. Hill, C. Ruis, B. Gutierrez, J. Raghwani, J. Ashworth, R. Colquhoun, T. R. Connor, N. R. Faria, B. Jackson, N. J. Loman, Á. O'Toole, S. M. Nicholls, K. V. Parag, E. Scher, T. I. Vasylyeva, E. M. Volz, A. Watts, I. I. Bogoch, K. Khan, COVID-19 Genomics UK (COG-UK) Consortium, D. M. Aanensen, M. U. G. Kraemer, A. Rambaut, O. G. Pybus, Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science* 10.1126/science.abf2946 (2021).

16. D. Bonsall, T. Golubchik, M. de Cesare, M. Limbada, B. Kosloff, G. MacIntyre-Cockett, M. Hall, C. Wymant, M. A. Ansari, L. Abeler-Dörner, A. Schaap, A. Brown, E. Barnes, E. Piwowar-Manning, S. Eshleman, E. Wilson, L. Emel, R. Hayes, S. Fidler, H. Ayles, R. Bowden, C. Fraser; HPTN 071 (PopART) Team, A comprehensive genomics solution for HIV surveillance and clinical monitoring in low-income settings. *J. Clin. Microbiol.* **58**, e00382-20 (2020). doi:10.1128/JCM.00382-20

Medline

17. C. Goh, T. Golubchik, A. Anzari, M. de Cesare, A. Trebes, I. Elliott, D. Bonsall, P. Piazza, A. Brown, H. Slawinski, N. Martin, S. Defres, M. J. Griffiths, J. E. Bray, M. C. Maiden, P. Hutton, C. J. Hinds, T. Solomon, E. Barnes, A. J. Pollard, M. Sadarangani, J. C. Knight, R. Bowden, Targeted metagenomic sequencing enhances the identification of pathogens associated with acute infection. bioRxiv 716902 [Preprint]. 28 July 2019. https://doi.org/10.1101/716902.
18. G.-L. Lin, T. Golubchik, S. Drysdale, D. O'Connor, K. Jefferies, A. Brown, M. de Cesare, D. Bonsall, M. A. Ansari, J. Aerssens, L. Bont, P. Openshaw, F. Martinón-Torres, R. Bowden, A. J. Pollard, H. Nair, H. Campbell, S. Cunningham, P. Beutels, L. Bont, J. Wildenbeest, A. Pollard, C. Butler, M. Snape, S. Drysdale, G.-L. Lin, D. O'Connor, E. Clutterbuck, K. Jefferies, J. McGinley, P. Openshaw, R. Thwaites, D. Wiseman, F. Martinon-Torres, A. Gómez-Carballa, T. Heikkinen, A. Meijer, T. K. Fischer, M. van den Berge, C. Giaquinto, M. Abram, K. Swanson, A. Leach, C. Demont, S. Gallichan, J. Aerssens, D. Öner, B. Rosen, E. Molero, H. Nair, H. Campbell, S. Cunningham, P. Beutels, L. Bont, J. Wildenbeest, A. Pollard, C. Butler, M. Snape, S. Drysdale, G.-L. Lin, D. O'Connor, E. Clutterbuck, K. Jefferies, J. McGinley, P. Openshaw, R. Thwaites, D. Wiseman, F. Martinon-Torres, A. Gómez-Carballa, T. Heikkinen, A. Meijer, T. K. Fischer, M. van den Berge, C. Giaquinto, M. Abram, K. Swanson, A. Leach, C. Demont, S. Gallichan, J. Aerssens, D. Öner, B. Rosen, E. Molero; RESCEU Investigators, Simultaneous viral whole-genome sequencing and differential expression profiling in respiratory syncytial virus infection of infants. *J. Infect. Dis.* **222**, S666–S671 (2020). doi:10.1093/infdis/jiaa448 Medline
19. D. Bonsall, M. A. Ansari, C. Ip, A. Trebes, A. Brown, P. Klenerman, D. Buck, P. Piazza, E. Barnes, R. Bowden; STOP-HCV Consortium, ve-SEQ: Robust, unbiased enrichment for streamlined detection and whole-genome sequencing of HCV and other highly diverse pathogens. *F1000Res.* **4**, 1062 (2015). doi:10.12688/f1000research.7111.1 Medline
20. S. M. Kissler, J. R. Fauver, C. Mack, S. W. Olesen, C. Tai, K. Y. Shiue, C. C. Kalinich, S. Jednak, I. M. Ott, C. B. F. Vogels, J. Wohlgemuth, J. Weisberger, J. DiFiori, D. J. Anderson, J. Mancell, D. D. Ho, N. D. Grubaugh, Y. H. Grad, SARS-CoV-2 viral dynamics in acute infections. medRxiv 20217042 [Preprint]. 1 December 2020. https://doi.org/10.1101/2020.10.21.20217042.
21. L. Ferretti, A. Ledda, C. Wymant, L. Zhao, V. Ledda, L. A. Dorner, M. Kendall, A. Nurtay, H.-Y. Cheng, T.-C. Ng, H.-H. Lin, R. Hinch, J. Masel, A. Marm Kilpatrick, C. Fraser, The timing of COVID-19 transmission. medRxiv 20188516 [Preprint]. 16 September 2020. https://doi.org/10.1101/2020.09.04.20188516.
22. M. Marks, P. Millat-Martinez, D. Ouchi, C. Roberts, A. Alemany, M. Corbacho-Monné, M. Ubals, A. Tobias, C. Tebé, E. Ballana, Q. Bassat, B. Baro, M. Vall-Mayans, C. G-Beiras, N. Prat, J. Ara, B. Clotet, O. Mitjà, C. G. Beiras, N. Prat, J. Ara, B. Clotet, O. Mitjà, Transmission of COVID-19 in 282 clusters in Catalonia, Spain: A cohort study. *Lancet Infect. Dis.* (2021). doi:10.1016/S1473-3099(20)30985-3
23. J. Raghwani, A. D. Redd, A. F. Longosz, C.-H. Wu, D. Serwadda, C. Martens, J. Kagaayi, N. Sewankambo, S. F. Porcella, M. K. Grabowski, T. C. Quinn, M. A. Eller, L. A. Eller, F. Wabwire-Mangen, M. L. Robb, C. Fraser, K. A. Lythgoe, Evolution of HIV-1 within untreated individuals and at the population scale in Uganda. *PLOS*

Pathog. **14**, e1007167 (2018). doi:10.1371/journal.ppat.1007167 Medline

24. G. Tonkin-Hill, I. Martincorena, R. Amato, A. R. J. Lawson, M. Gerstung, I. Johnston, D. K. Jackson, N. R. Park, S. V. Lensing, M. A. Quail, S. Gonçalves, C. Ariani, M. S. Chapman, W. L. Hamilton, L. W. Meredith, G. Hall, A. S. Jahun, Y. Chaudhry, M. Hosmillo, M. L. Pinckert, I. Georgana, A. Yakovleva, L. G. Caller, S. L. Caddy, T. Feltwell, F. A. Khokhar, C. J. Houldcroft, M. D. Curran, S. Parmar, The COVID-19 Genomics UK (COG-UK) Consortium, A. Alderton, R. Nelson, E. Harrison, J. Sillitoe, S. D. Bentley, J. C. Barrett, M. Estee Torok, I. G. Goodfellow, C. Langford, D. Kwiatkowski, Wellcome Sanger Institute COVID-19 Surveillance Team, Patterns of within-host genetic diversity in SARS-CoV-2. bioRxiv 424229 [Preprint]. 25 December 2020. https://doi.org/10.1101/2020.12.23.424229.

25. A. Popa, J.-W. Genger, M. D. Nicholson, T. Penz, D. Schmid, S. W. Aberle, B. Agerer, A. Lercher, L. Endler, H. Colaço, M. Smyth, M. Schuster, M. L. Grau, F. Martínez-Jiménez, O. Pich, W. Borena, E. Pawelka, Z. Keszei, M. Senekowitsch, J. Laine, J. H. Aberle, M. Redlberger-Fritz, M. Karolyi, A. Zoufaly, S. Maritschnik, M. Borkovec, P. Hufnagl, M. Nairz, G. Weiss, M. T. Wolfinger, D. von Laer, G. Superti-Furga, N. Lopez-Bigas, E. Puchhammer-Stöckl, F. Allerberger, F. Michor, C. Bock, A. Bergthaler, Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. *Sci. Transl. Med.* **12**, eabe2555 (2020). doi:10.1126/scitranslmed.abe2555 Medline

26. A. L. Valesano, K. E. Rumfelt, D. E. Dimcheff, C. N. Blair, W. J. Fitzsimmons, J. G. Petrie, E. T. Martin, A. S. Lauring, Temporal dynamics of SARS-CoV-2 mutation accumulation within and across infected hosts. bioRxiv 2021.01.19.427330 (2021). 10.1101/2021.01.19.427330 Medline

27. M. P. Zwart, S. F. Elena, Matters of size: Genetic bottlenecks in virus infection and their potential impact on evolution. *Annu. Rev. Virol.* **2**, 161–179 (2015). doi:10.1146/annurev-virology-100114-055135 Medline

28. A. Sobel Leonard, D. B. Weissman, B. Greenbaum, E. Ghedin, K. Koelle, Transmission bottleneck size estimation from pathogen deep-sequencing data, with an application to human influenza A virus. *J. Virol.* **91**, e00171-17 (2017). doi:10.1128/JVI.00171-17 Medline

29. A. Sobel Leonard, D. B. Weissman, B. Greenbaum, E. Ghedin, K. Koelle, Correction for Sobel Leonard et al., “Transmission bottleneck size estimation from pathogen deep-sequencing data, with an application to human influenza A virus”. *J. Virol.* **93**, e00936-19 (2019). doi:10.1128/JVI.00936-19 Medline

30. M. Ghafari, C. K. Lumby, D. B. Weissman, C. J. R. Illingworth, Inferring transmission bottleneck size from viral sequence data using a novel haplotype reconstruction method. *J. Virol.* **94**, e00014-20 (2020). doi:10.1128/JVI.00014-20 Medline

31. J. T. McCrone, R. J. Woods, E. T. Martin, R. E. Malosh, A. S. Monto, A. S. Lauring, Stochastic processes constrain the within and between host evolution of influenza virus. *eLife* **7**, e35962 (2018). doi:10.7554/eLife.35962 Medline

32. A. Varble, R. A. Albrecht, S. Backes, M. Crumiller, N. M. Bouvier, D. Sachs, A. García-Sastre, B. R. tenOever, Influenza A virus transmission bottlenecks are defined by infection route and recipient host. *Cell Host Microbe* **16**, 691–700 (2014). doi:10.1016/j.chom.2014.09.020 Medline

33. Z. Shen, Y. Xiao, L. Kang, W. Ma, L. Shi, L. Zhang, Z. Zhou, J. Yang, J. Zhong, D.

Yang, L. Guo, G. Zhang, H. Li, Y. Xu, M. Chen, Z. Gao, J. Wang, L. Ren, M. Li, Genomic diversity of severe acute respiratory syndrome–coronavirus 2 in patients with coronavirus disease 2019. *Clin. Infect. Dis.* (2020). doi:10.1093/cid/ciaa203

34. T. H. Wang, Y. K. Donaldson, R. P. Brettle, J. E. Bell, P. Simmonds, Identification of shared populations of human immunodeficiency virus type 1 infecting microglia and tissue macrophages outside the central nervous system. *J. Virol.* **75**, 11686–11699 (2001). doi:10.1128/JVI.75.23.11686-11699.2001 Medline

35. K. M. Braun, G. K. Moreno, P. J. Halfmann, E. B. Hodcroft, D. A. Baker, E. C. Boehm, A. M. Weiler, A. K. Haj, M. Hatta, S. Chiba, T. Maemura, Y. Kawaoka, K. Koelle, D. H. O'Connor, T. C. Friedrich, Transmission of SARS-CoV-2 in domestic cats imposes a narrow bottleneck. bioRxiv 384917 [Preprint]. 4 January 2021. https://doi.org/10.1101/2020.11.16.384917.

36. S. E. James, S. Ngcapu, A. M. Kanzi, H. Tegally, V. Fonseca, J. Giandhari, E. Wilkinson, B. Chimukangara, S. Pillay, L. Singh, M. Fish, I. Gazy, K. Khanyile, R. Lessells, T. de Oliveira, High resolution analysis of transmission dynamics of Sars-Cov-2 in two major hospital outbreaks in South Africa leveraging intrahost diversity. medRxiv 20231993 [Preprint]. 16 November 2020. https://doi.org/10.1101/2020.11.15.20231993.

37. M. A. Martin, K. Koelle, Reanalysis of deep-sequencing data from Austria points towards a small SARS-COV-2 transmission bottleneck on the order of one to three virions. bioRxiv 432096 [Preprint]. 22 February 2021. https://doi.org/10.1101/2021.02.22.432096.

38. M. Gelbart, S. Harari, Y. Ben-Ari, T. Kustin, D. Wolf, M. Mandelboim, O. Mor, P. S. Pennings, A. Stern, Drivers of within-host genetic diversity in acute infections of viruses. *PLOS Pathog.* **16**, e1009029 (2020). doi:10.1371/journal.ppat.1009029 Medline

39. M. D. Nowak, E. M. Sordillo, M. R. Gitman, A. E. Paniz Mondolfi, Coinfection in SARS-CoV-2 infected patients: Where are influenza virus and rhinovirus/enterovirus? *J. Med. Virol.* **92**, 1699–1700 (2020). doi:10.1002/jmv.25953 Medline

40. D. Kim, J. Quinn, B. Pinsky, N. H. Shah, I. Brown, Rates of co-infection between SARS-CoV-2 and other respiratory pathogens. *JAMA* **323**, 2085–2086 (2020). doi:10.1001/jama.2020.6266 Medline

41. B. Dearlove, E. Lewitus, H. Bai, Y. Li, D. B. Reeves, M. G. Joyce, P. T. Scott, M. F. Amare, S. Vasan, N. L. Michael, K. Modjarrad, M. Rolland, A SARS-CoV-2 vaccine candidate would likely match all currently circulating variants. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 23652–23662 (2020). doi:10.1073/pnas.2008281117 Medline

42. B. Morel, P. Barbera, L. Czech, B. Bettisworth, L. Hübner, S. Lutteropp, D. Serdari, E.-G. Kostaki, I. Mamais, A. M. Kozlov, P. Pavlidis, D. Paraskevis, A. Stamatakis, Phylogenetic analysis of SARS-CoV-2 data is difficult. *Mol. Biol. Evol.* msaa314 (2020). doi:10.1093/molbev/msaa314 Medline

43. L. van Dorp, M. Acman, D. Richard, L. P. Shaw, C. E. Ford, L. Ormond, C. J. Owen, J. Pang, C. C. S. Tan, F. A. T. Boshier, A. T. Ortiz, F. Balloux, Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infect. Genet. Evol.* **83**, 104351 (2020). doi:10.1016/j.meegid.2020.104351 Medline

44. Q. Li, J. Wu, J. Nie, L. Zhang, H. Hao, S. Liu, C. Zhao, Q. Zhang, H. Liu, L. Nie, H. Qin,

M. Wang, Q. Lu, X. Li, Q. Sun, J. Liu, L. Zhang, X. Li, W. Huang, Y. Wang, The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell* **182**, 1284–1294.e9 (2020). doi:10.1016/j.cell.2020.07.012 Medline

45. S. Pond. "Natural selection analysis of global SARS-CoV-2/COVID-19 enabled by data from GISAID" (Datamonkey, 2021); https://observablehq.com/@spond/revised-sars-cov-2-analytics-page.

46. J. T. Herbeck, D. C. Nickle, G. H. Learn, G. S. Gottlieb, M. E. Curlin, L. Heath, J. I. Mullins, Human immunodeficiency virus type 1 env evolves toward ancestral states upon transmission to a new host. *J. Virol.* **80**, 1637–1644 (2006). doi:10.1128/JVI.80.4.1637-1644.2006 Medline

47. T. N. Starr, A. J. Greaney, A. Addetia, W. W. Hannon, M. C. Choudhary, A. S. Dingens, J. Z. Li, J. D. Bloom, Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science* **371**, 850–854 (2021). doi:10.1126/science.abf9302 Medline

48. A. J. Greaney, A. N. Loes, K. H. D. Crawford, T. N. Starr, K. D. Malone, H. Y. Chu, J. D. Bloom, Comprehensive mapping of mutations to the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human serum antibodies. bioRxiv 425021 [Preprint]. 4 January 2021. https://doi.org/10.1101/2020.12.31.425021.

49. K. R. McCarthy, L. J. Rennick, S. Nambulli, L. R. Robinson-McCarthy, W. G. Bain, G. Haidar, W. P. Duprex, Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *Science* eabf6950 (2021). Medline

50. E. M. Volz, V. Hill, J. T. McCrone, A. Price, D. Jorgensen, A. O'Toole, J. A. Southgate, R. Johnson, B. Jackson, F. F. Nascimento, S. M. Rey, S. M. Nicholls, R. M. Colquhoun, A. da Silva Filipe, J. G. Shepherd, D. J. Pascall, R. Shah, N. Jesudason, K. Li, R. Jarrett, N. Pacchiarini, M. Bull, L. Geidelberg, I. Siveroni, I. G. Goodfellow, N. J. Loman, O. Pybus, D. L. Robertson, E. C. Thomson, A. Rambaut, T. R. Connor, Evaluating the effects of SARS-CoV-2 Spike mutation D614G on transmissibility and pathogenicity. medRxiv 20166082 [Preprint]. 1 September 2020. https://doi.org/10.1101/2020.07.31.20166082.

51. E. C. Thomson, L. E. Rosen, J. G. Shepherd, R. Spreafico, A. da Silva Filipe, J. A. Wojcechowskyj, C. Davis, L. Piccoli, D. J. Pascall, J. Dillen, S. Lytras, N. Czudnochowski, R. Shah, M. Meury, N. Jesudason, A. De Marco, K. Li, J. Bassi, A. O'Toole, D. Pinto, R. M. Colquhoun, K. Culap, B. Jackson, F. Zatta, A. Rambaut, S. Jaconi, V. B. Sreenu, J. Nix, I. Zhang, R. F. Jarrett, W. G. Glass, M. Beltramello, K. Nomikou, M. Pizzuto, L. Tong, E. Cameroni, T. I. Croll, N. Johnson, J. Di Iulio, A. Wickenhagen, A. Ceschi, A. M. Harbison, D. Mair, P. Ferrari, K. Smollett, F. Sallusto, S. Carmichael, C. Garzoni, J. Nichols, M. Galli, J. Hughes, A. Riva, A. Ho, M. Schiuma, M. G. Semple, P. J. M. Openshaw, E. Fadda, J. K. Baillie, J. D. Chodera, S. J. Rihn, S. J. Lycett, H. W. Virgin, A. Telenti, D. Corti, D. L. Robertson, G. Snell; ISARIC4C Investigators; COVID-19 Genomics UK (COG-UK) Consortium, Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell* **184**, 1171–1187.e20 (2021). doi:10.1016/j.cell.2021.01.037 Medline

52. M. R. Zambenedetti, D. P. Pavoni, A. C. Dallabona, A. C. Dominguez, C. O. Poersch, S. P. Fragoso, M. A. Krieger, Internal control for real-time polymerase chain reaction based on MS2 bacteriophage for RNA viruses diagnostics. *Mem. Inst. Oswaldo Cruz*

112, 339–347 (2017). doi:10.1590/0074-02760160380 Medline

53. F. Gao, D. L. Robertson, C. D. Carruthers, S. G. Morrison, B. Jian, Y. Chen, F. Barré-Sinoussi, M. Girard, A. Srinivasan, A. G. Abimiku, G. M. Shaw, P. M. Sharp, B. H. Hahn, A comprehensive panel of near-full-length clones and reference sequences for non-subtype B isolates of human immunodeficiency virus type 1. *J. Virol.* **72**, 5680–5698 (1998). doi:10.1128/JVI.72.7.5680-5698.1998 Medline

54. D. E. Wood, J. Lu, B. Langmead, Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019). doi:10.1186/s13059-019-1891-0 Medline

55. T. Golubch, “Castanet” (Github, 2020); https://github.com/tgolubch/castanet.

56. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014). doi:10.1093/bioinformatics/btu170 Medline

57. C. Wymant, F. Blanquart, T. Golubchik, A. Gall, M. Bakker, D. Bezemer, N. J. Croucher, M. Hall, M. Hillebregt, S. H. Ong, O. Ratmann, J. Albert, N. Bannert, J. Fellay, K. Fransen, A. Gourlay, M. K. Grabowski, B. Gunsenheimer-Bartmeyer, H. F. Günthard, P. Kivelä, R. Kouyos, O. Laeyendecker, K. Liitsola, L. Meyer, K. Porter, M. Ristola, A. van Sighem, B. Berkhout, M. Cornelissen, P. Kellam, P. Reiss, C. Fraser; BEEHIVE Collaboration, Easy and accurate reconstruction of whole HIV genomes from short-read sequence data with shiver. *Virus Evol.* **4**, vey007 (2018). doi:10.1093/ve/vey007 Medline

58. Sanger Institute, “Tools directory” (Sanger Institute, 2021); https://www.sanger.ac.uk/science/tools.

59. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). doi:10.1038/nmeth.1923 Medline

60. A. O’Toole, V. Hill, J. T. McCrone, E. Scher, A. Rambaut, “Pangolin COVID-19 lineage assigner v2.1.7, lineages version 2021-01-20” (Centre for Genomic Pathogen Surveillance, 2021); https://pangolin.cog-uk.io/.

61. C. Mavian, S. Marini, M. Prosperi, M. Salemi, A snapshot of SARS-CoV-2 genome availability up to April 2020 and its implications. *JMIR Public Health Surveill.* **6**, e19170 (2020). doi:10.2196/19170 Medline

62. Y. Shu, J. McCauley, GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill.* **22**, 30494 (2017). doi:10.2807/1560-7917.ES.2017.22.13.30494 Medline

63. L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012). doi:10.1093/bioinformatics/bts565 Medline

64. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002). doi:10.1093/nar/gkf436 Medline

65. M. Nei, T. Gojobori, Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**, 418–426 (1986). Medline

66. A. M. Kozlov, D. Darriba, T. Flouri, B. Morel, A. Stamatakis, RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference.

Bioinformatics **35**, 4453–4455 (2019). doi:10.1093/bioinformatics/btz305 Medline

67. X. Didelot, D. J. Wilson, ClonalFrameML: Efficient inference of recombination in whole bacterial genomes. *PLOS Comput. Biol.* **11**, e1004041 (2015). doi:10.1371/journal.pcbi.1004041 Medline

68. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018). doi:10.1093/molbev/msx281 Medline

69. COVID-19 Genomics UK (COG-UK) consortiumcontact@cogconsortium.uk, An integrated national scale SARS-CoV-2 genomic surveillance network. *Lancet Microbe* **1**, e99–e100 (2020). doi:10.1016/S2666-5247(20)30054-9 Medline

70. Data for: K. A. Lythgoe, M. Hall, L. Ferretti, M. de Cesare, G. MacIntyre-Cockett, A. Trebes, M. Andersson, N. Otecko, E. L. Wise, N. Moore, J. Lynch, S. Kidd, N. Cortes, M. Mori, R. Williams, G. Vernet, A. Justice, A. Green, S. M. Nicholls, M. A. Ansari, L. Abeler-Dörner, C. E. Moore, T. E. A. Peto, D. W. Eyre, R. Shaw, P. Simmonds, D. Buck, J. A. Todd on behalf of the Oxford Virus Sequencing Analysis Group (OVSG), T. R. Connor, S. Ashraf, A. da Silva Filipe, J. Shepherd, E. C. Thomson, The COVID-19 Genomics UK (COG-UK) Consortium, D. Bonsall, C. Fraser, T. Golubchik, SARS-CoV-2 within-host diversity and transmission. Zenodo (2021); https://doi.org/10.5281/zenodo.4570598.