

## **Viral genetic sequencing identifies staff transmission of COVID-19 is important in a community hospital outbreak**

Jane A.H. Masoli\* MBChB<sup>1,2</sup>, Aaron Jeffries PhD<sup>\*4</sup>, Ben Temperton PhD<sup>\*3</sup>, Cressida Auckland MBBS<sup>2</sup>, Michelle Michelsen BSc<sup>3</sup>, Joanna Warwick-Dugdale PhD<sup>3</sup>, Robyn Manley PhD<sup>3</sup>, Audrey Farbos MSc<sup>3</sup>, Sian Ellard PhD<sup>4</sup>, Beatrice Knight PhD<sup>2,4</sup>, Claire Bewshea MSc<sup>4</sup>, Christine Sambles PhD<sup>3</sup>, James Harrison PhD<sup>3</sup>, Ben Bunce MSc<sup>2</sup>, Alexis Carr MBBS<sup>2</sup>, Andrew T. Hattersley DM<sup>4</sup>, Stephen Ll. Michell PhD<sup>3</sup>, David J. Studholme PhD<sup>3</sup>

1. Epidemiology and Public Health Group, College of Medicine and Health, University of Exeter, Exeter, UK
2. Royal Devon and Exeter Hospital NHS Trust, Barrack Road, Exeter, UK.
3. Biosciences. College of Life and Environmental Sciences, University of Exeter, Exeter, UK.
4. Institute of Biomedical and Clinical Science, College of Medicine and Health, University of Exeter, Exeter, UK

\* Jane A.H. Masoli, Aaron Jeffries and Ben Temperton contributed equally to this work.

Corresponding author: David Studholme: [d.j.studholme@exeter.ac.uk](mailto:d.j.studholme@exeter.ac.uk)

### **Key points:**

- Detailed analysis of a large community hospital outbreak in older adults and staff with concurrent clinical and genomic data, including working patterns.
- Staff transmission was important in this community hospital outbreak.
- We found plausible associations between staff and patient cases.

## **Abstract**

### **Background**

We have successfully used whole-genome sequencing to provide additional information for transmission pathways in infectious spread. We report and interpret genomic sequencing results in clinical context from a large outbreak of COVID-19 with 46 cases across staff and patients in a community hospital in the UK.

### **Methods**

Following multiple symptomatic cases within a two-week period, all staff and patients were screened by RT-PCR and staff subsequently had serology tests.

### **Results**

Thirty staff (25%) and 16 patients (62%) tested positive for COVID-19. Genomic sequencing data showed significant overlap of viral haplotypes in staff who had overlapping shift patterns. Patient haplotypes were more distinct from each other but had overlap with staff haplotypes.

### **Conclusions**

This study includes clinical and genomic epidemiological detail that demonstrates the value of a combined approach. Viral genetic sequencing has identified that staff transmission of COVID-19 was important in this community hospital outbreak.

## Introduction

Whole-genome sequencing and phylogenetic analysis of pathogen isolates can be combined with traditional epidemiology to define transmission pathways and track outbreaks.<sup>1</sup> Applied to bacterial pathogens such as methicillin-resistant *Staphylococcus aureus* (MRSA),<sup>2</sup> sequencing has distinguished between closely related genotypes that would be otherwise indistinguishable, thereby improving the resolution of analyses of pathogen transmission. Viruses have high rates of mutation and replication, resulting in rapid genetic variation.<sup>3</sup> It has been proposed that viral genome sequencing and phylogenetic analyses could be key in viral epidemics by adding molecular precision<sup>4</sup> and this approach was pioneered in near real time during the West African ebola epidemic of 2014 to 2016<sup>5</sup>. The SARS-CoV-2 pandemic has precipitated unprecedented levels of application of viral genome sequencing, which has been used to inform clinical decision-making.<sup>5</sup> Clinically, spread of COVID-19 can occur in the community or within hospitals and can be acquired from multiple sources. Outbreak analysis best practice combines detailed clinical epidemiology with supporting microbiological testing.<sup>6</sup> However, it is extremely difficult to ascertain the source of a positive case and standard testing by reverse transcriptase–polymerase chain reaction (RT-PCR)<sup>7</sup> does not aid understanding of the transmission pathway. Phylogenetic analysis of genome sequences from outbreaks can add essential detail. This is exemplified in New Zealand where genomic surveillance has been successfully deployed to understand the origins of SARS-CoV-2 in circulation and patterns of their transmission.<sup>8</sup>

The COVID-19 Genomics UK (COG-UK) Consortium<sup>9</sup> was established to support rapid whole-genome viral sequencing for confirmed COVID-19 cases<sup>10</sup>. As of June 2020, the epidemic in the UK comprised at least 1356 independently introduced lineages. Comparison of the genome sequences of viral isolates from a cluster of cases can be used to test

hypothetical epidemiological links between those cases and provide supporting or contradictory evidence.

We report the analysis of a clinical outbreak of COVID-19 at a community hospital in Devon, United Kingdom, and show whole-genome viral sequencing combined with clinical context to hypothesise likely transmission routes in this outbreak.

## **Methods**

### **Study design and participants**

We investigated a COVID-19 outbreak that occurred in a 32-bed community hospital with 118 staff. The community hospital focuses on rehabilitation, with the majority of patients transferred following inpatient stays at the acute hospital.

### **COVID-19 testing**

The full inpatient cohort of 26 and 118 staff were tested for COVID-19 by reverse transcriptase– polymerase chain reaction (RT-PCR) on nasopharyngeal and throat swabs within the outbreak period.

RT-PCR testing used the platforms SmartCycler® (Cepheid, Sunnyvale, CA, USA) and ELITE InGenius® (ELITech Group, Torino, Italy).

Serological testing of COVID-19 via antibody detection was introduced in the UK in May 2020. Antibody testing on serology was performed in 115 staff when first available in the UK NHS, seven weeks after the initial RT-PCR screening. Total antibody count has reported to yield up to 100% sensitivity from two weeks after symptom onset<sup>7,11</sup>. Serology was performed to ascertain whether initial screening at outbreak accounted for the majority of cases in the staff.

## **Clinical analysis**

Detailed information was obtained for clinical epidemiological evaluation from outbreak surveillance and supported by additional detail for participants recruited to the Exeter COVID-19 Sequencing project (ExCoSe). The sources used for clinical analysis consisted of inpatient medical records, staff symptom questionnaires, and staff rotas. We ascertained symptom timelines, exposures, co-working, COVID-19 severity and outcomes.

## **Viral genome sequencing**

44 of 46 positive cases had retained swab material from the RT-PCR on which whole-genome viral sequencing was attempted. RNA was prepared and sequenced using protocols published by the ARTIC network (<https://artic.network/ncov-2019>), based on 400 bp PCR tiling amplicons to sequence the whole 30 kb genome. Briefly, RNA extracted from nasopharyngeal swabs was reverse transcribed and used as a template in for two separate multiplexed PCRs for 35 cycles. The two multiplexed PCRs were then pooled for each sample, purified with 1x AMPure XP bead purification (Beckman Coulter, USA), quantified by fluorimetry and normalized in concentration. Sequencing libraries were prepared (Oxford Nanopore Technologies, UK) by adding barcode and adaptor to the amplicon DNA; the full protocol is available at <https://dx.doi.org/10.17504/protocols.io.bdp7i5rn>. Sequencing was performed on an Oxford Nanopore MinION. Each sequencing run comprised 23 samples plus one negative control (carried through from RT-PCR step). Duration of sequencing runs sequenced were 16 to 24 hours, with high-accuracy real-time base calling and demultiplexing (Guppy 3.5.2) and monitoring with MinKNOW v3.6.5 (Oxford Nanopore Technologies) and RAMPART<sup>12</sup>. The

ARTIC pipeline (<https://github.com/artic-network/artic-ncov2019.git>) was then used for subsequent filtering, assembly and variant calling.

## **Sequence Analysis**

We aligned genome sequences against the Wuhan-Hu-1 reference genome sequence (GenBank MN908947.3<sup>13</sup>) using MAFFT v7.310.<sup>14</sup> Haplotypes were inferred from the resulting alignment using a custom script ([https://github.com/davidjstudholme/get\\_viral\\_haplotypes](https://github.com/davidjstudholme/get_viral_haplotypes)) that identifies genomic sites that show variation and are unambiguously called (i.e. no Ns) across all analysed genome sequences to generate a Nexus-formatted output file.<sup>15</sup> This Nexus file served as input into Popart<sup>16</sup> to generate median-spanning networks.<sup>17</sup>

## **Ethical Approval**

The genomic sequencing was performed as part of the COG-UK study. COG-UK genomic sequencing is conducted as part of surveillance for COVID-19 infections under the auspices of Section 251 of the NHS Act 2006. The COG-UK study protocol was approved by the Public Health England Research Ethics Governance Group (reference: R&D NR0195).

Participants were identified during routine clinical practice and recruited via the Royal Devon and Exeter Tissue Bank (RDETB – ethical approval REC no 11/SW/0018) set up to facilitate research through the collection of "spare" clinical samples available at the time of routine procedures.

Approval for the ExCoSe study (STB63/CTB58) was obtained from the RDETB Steering Committee, who are delegated to approve study specific collections. The RDETB is managed within the NIHR Exeter Clinical Research Facility.

## **Results**

### **Description of outbreak**

Multiple cases were detected, following the onset of symptoms in both staff and patient groups, therefore screening was conducted on all inpatients (n=26) and staff (n=118).

### **PCR Testing**

RT-PCR tested positive in 30 of 118 staff (25%) and 16 of 26 (62%) patients.

### **Staff cases**

Six of the staff cases were asymptomatic at the point of testing, with two remaining asymptomatic. The majority developed mild to moderate symptoms at worst but one case had severe COVID-19 requiring hospitalisation. The first cases of symptomatic staff members preceded any positive results in inpatients. All symptomatic staff remained away from work appropriately. This outbreak occurred before routine testing was available for healthcare workers. Two positive staff members had been symptomatic, followed Public Health England guidance on absence and return to work and had returned to work following their period of absence prior to their positive COVID-19 swab.

### **Patients**

The mean age of the 16 COVID-19 positive patients was 82 years (SD 8.7) and 63% were female. All of the COVID-19 positive patients had recently been inpatients at the acute hospital trust prior to transfer to community hospital, with a mean time from transfer to confirmed COVID-19 of 18 days (SD 9.5). Any patients with a clinical suspicion of COVID-19 were tested and confirmed COVID-19 negative prior to transfer to the community

hospital, as per the policy at that point; guidance was subsequently updated to test all prior to transfer. The patients had significant co-morbidity (mean seven co-morbidities) and polypharmacy (mean eight co-prescribed regular medications). Seven of the patients were asymptomatic for COVID-19 at the point of positive swab. Thirteen of the patients were transferred back to the acute trust, although this did not correlate with severity of symptoms as it was early in the pandemic and these patients were anticipated to potentially require further medical input. Four patients died within 30 days of COVID-19 positive test.

### **Serology**

Subsequent serological testing was performed in 115 of the 118 staff who were tested by RT-PCR. Twenty-nine staff were PCR positive and serology positive. Ten additional staff members had positive serology. There was one case who was PCR positive and had successful viral sequencing for whom the serology was negative on two repeated tests.

### **Viral genome sequence**

Viral RNA was available for 28/30 staff and all 16 patients. Near-complete viral genome sequences were successfully generated for 18 staff and 13 patients, plus 122 additional viral genomes from cases in the Exeter area at that point. Median depth of sequencing across each genome ranged between 1,657 x and 27,377 x. Each genome had at least 20 x coverage across 90.16 – 99.87 % of its length.

Two hundred and thirty-one sites in the genome showed single nucleotide polymorphism (SNP). The 31 community hospital samples fell into seven distinct haplotypes (A-H), defined by 29 SNPs that were polymorphic among these; the other 202 SNPs were invariant among



the outbreak samples. The relationships between the viral haplotypes are illustrated in Figure 1A, placing the local cases into regional context of viral genetic diversity, and Figure 2. Civet and Pangolin (<https://articnetwork.github.io/civet/>) assigned all sequenced genomes to global lineage B1.1 and UK lineage UK2266 except for one sample (our haplotype H), which was assigned to global lineage B1 and UK lineage UK5741.

Overall, the viral sequences from the 13 patients represented seven haplotypes as illustrated in Figure 2. Two were shared with staff: A (n=6 patients) and B (n=2 patients) and five were unique (C, D, E, F and G). The viral sequences from most of the 18 staff fall into two haplotypes A (n=10 staff) and B (n=7 staff). Haplotype H was identified in a new staff member who had not previously had overlapping shifts with other COVID-19 positive staff members. Given the estimated mutation rate of approximately two changes per genome per month<sup>4</sup>, while we cannot confidently propose links between cases whose haplotypes are differentiated by one or two SNPs, it is plausible that they are connected.

The sequencing-based networks in Figure 1 reveal that the most-frequent haplotype (A) in the outbreak cases is frequent across samples sequenced regionally (n= 37 (haplotype A) out of 141 sequenced at that point). This suggests that this haplotype is common and therefore its co-incidence among 16 of 31 cases could be explained by either spread from a single initial case or multiple independent acquisitions from the community, or a combination of both of these. However, the second-most frequent haplotype (B) among outbreak cases is not seen in other samples external to the outbreak (Figure 1), nor in other samples sequenced to date nationally, strongly suggesting that those cases are epidemiologically linked. Two further cases were separated by a single mutation from this haplotype (haplotypes E and F) suggesting they could have arisen from haplotype B, and so were included in analysis of this cluster. We therefore undertook a more detailed analysis on the cluster of cases with

haplotype B plus haplotypes E and F to try to identify the likely spread of infection to guide policy to prevent future outbreaks (Figure 3).

### **Cluster analysis**

The patients with haplotypes E and F and three of the four patients with haplotype B were asymptomatic and screened for COVID-19 by PCR test as part of the outbreak analysis, with one complaining of a mild cough following positive result. Patient 4 was negative at initial screening and subsequently positive. Detailed case-note review revealed the timeline of exposure and symptom onset (Figure 3). Three staff members within the cluster were symptomatic and off work before any patient cases were confirmed positive and prior to the transfer of patients 1 and 2 from the acute hospital to the community hospital. It is noticeable that the cluster of staff who share haplotype B sequencing had overlapping shift patterns over the two-week period preceding symptom onset, with regular points of contact at shift handover. Additionally, the staff cases within this cluster all developed symptoms within a one-week period.

### **Discussion**

Genomic sequencing has shown that the majority of staff cases in this outbreak had a limited lineage, while patient cases were more likely to be distinct. The most common haplotype in this outbreak maps to the lineage sequenced most frequently both regionally and nationally. Distinct mutations have arisen in this viral lineage to result in the cluster of cases described with haplotypes B, E and F, which have only been identified in this outbreak.

The cluster of cases within this distinct haplotype suggests a likely transmission pathway within the community hospital. COVID-19 positive patients more frequently had a unique haplotype than did staff, suggesting that patient-to-staff transmission was rare. The staff's

viral genomes predominantly fell within two haplotypes, one of which was unique to this outbreak. At the time of this reported outbreak, the South West of England had the lowest COVID-19 incidence in the country.<sup>18</sup> Such a large cluster of cases, within two (closely related) haplotypes in staff, indicates it is highly likely that there was significant staff-to-staff transmission in this outbreak. Patients had a more diverse range of haplotypes, but it is also probable that there was some staff-to-patient transmission, given overlapping haplotypes between staff and inpatients who had been in the community hospital for an average of 18 days prior to screening.

Whilst the route of transmission between staff cannot be definitely assigned, it is noted that the staff with shared haplotypes had overlapping shift patterns. Healthcare workers have higher COVID-19 exposure than the general population due to the nature of their work<sup>19</sup> and have been commonly affected worldwide.<sup>20</sup> There have been prior reported outbreaks in which healthcare workers have represented the majority of cases affected,<sup>21</sup> and nosocomial infection rates likely transmitted by healthcare workers have risen, leading to regular healthcare worker testing.<sup>22</sup> In the UK COVID-19 was found to be RT-PCR positive in 14% of healthcare workers reporting symptoms.<sup>23</sup> Rivett et al. described large-scale screening of 1,032 asymptomatic healthcare workers, with 3% positive.<sup>24</sup> Those with symptoms can follow public health measures on isolation, but asymptomatic carriage is problematic.

In this outbreak viral sequencing provided key information on which cases are likely to be linked versus those from different sources. This illustrates the utility of sequencing and supports the use of rapid viral sequencing in the evaluation of COVID-19 outbreaks. We have shown that in future COVID-19 surge episodes or outbreaks whole-genome sequencing could be used effectively to support clinical evaluation. This method could be employed alongside contact tracing to track COVID-19 spread in real-time.

A limitation of the study is that viral sequence was not generated in all cases, due to unavailable samples or insufficient sequence, probably reflecting low viral load at testing. This outbreak was prior to widespread testing among healthcare workers or asymptomatic inpatients outside outbreak scenarios and if events repeated now it is likely that testing would have occurred earlier in individual symptomatic cases.

A key strength of this study is that it combined viral genomic sequencing with epidemiological clinical analysis, RT-PCR and subsequent serological testing to facilitate full outbreak analysis. The screening of all staff members first by RT-PCR and then by serology testing has given a complete picture in this outbreak. The subsequent antibody testing shows that the majority of cases occurred in the staff during this outbreak . The COG-UK genetic sequencing provides excellent coverage for genomic sequencing in the UK and better clinical corroboration could aid understanding in outbreaks and more generally, such as whether new COVID-19 cases over time in the same setting (e.g. a specific care home) are the same haplotype or if distinct viral haplotypes are being introduced by staff or visitors. This could significantly affect policy, but requires centres having access to both clinical detail and viral genomic sequencing, a particular strength of this study, in which the NHS trust and University teams work in close collaboration. It is particularly challenging to combine sequencing and clinical information from non-hospital settings such as care homes due to centralisation of testing practices.

## **Conclusion**

This case demonstrates the added value of genomic epidemiology in outbreak analysis. The likely staff-to-staff transmission and potential staff-to-patient transmission should be heeded. It reinforces the need for implementing infection control strategies and social distancing within health and social care environments between healthcare workers<sup>19,25</sup>, even in the

absence of symptoms. Ward handovers and other meeting events, plus staff rest spaces and office areas may be particular risks and need to be urgently re-evaluated and adapted during the pandemic of this highly infectious virus. The lessons learnt from this are transferrable to hospitals and care home settings, in which staff groups provide care and assistance to older people. Viral genetic sequencing has identified that staff transmission of COVID-19 was important in this community hospital outbreak.

### **Funding**

COG-UK is supported by funding from the Medical Research Council (MRC) part of UK Research & Innovation (UKRI), the National Institute of Health Research (NIHR) and Genome Research Limited, operating as the “Wellcome Sanger Institute”.

JM is funded by the National Institute for Health Research (NIHR) (Doctoral Research Fellowship; DRF-2014-07-177). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

### **Author contributions**

Case identification, patient recruitment, data acquisition and clinical interpretation: JM, CA, BK, AC, AH, CB. Process of genetic sequencing and data production: SM, AJ, BT, J W-D, RM, AF, CS, BB, JH, SE, MM. JM, AJ, BT, AH & DS drafted the report. JM and DS are guarantors for the data. All authors approved the final version of the report

### **Acknowledgements**

We would like to acknowledge the COVID-19 Genomics UK (COG-UK) consortium, which provides the infrastructure for COVID-19 sequencing across the UK.

We would like to acknowledge the Royal Devon and Exeter Tissue Bank team, which supported recruitment and data collection via the ExCoSe study.

## References

1. Ladhani SN, Chow JY, Janarthanan R, et al. Investigation of SARS-CoV-2 outbreaks in six care homes in London, April 2020. *EClinicalMedicine*. 2020;26. doi:10.1016/j.eclinm.2020.100533
2. Köser CU, Holden MTG, Ellington MJ, et al. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N Engl J Med*. 2012;366(24):2267-2275. doi:10.1056/NEJMoa1109910
3. Duffy S, Shackelton LA, Holmes EC. Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet*. 2008;9(4):267-276. doi:10.1038/nrg2323
4. Grubaugh ND, Ladner JT, Lemey P, et al. Tracking virus outbreaks in the twenty-first century. *Nat Microbiol*. 2019;4(1):10-19. doi:10.1038/s41564-018-0296-2
5. Gire SK, Goba A, Andersen KG, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science (80- )*. 2014;345(6202):1369-1372. doi:10.1126/science.1259657
6. Public Health England. Communicable Disease Outbreak Management. Operational guidance. *Public Heal Engl*. 2014:1-66. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/343723/12\\_8\\_2014\\_CD\\_Outbreak\\_Guidance\\_REandCT\\_2\\_\\_2\\_.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/343723/12_8_2014_CD_Outbreak_Guidance_REandCT_2__2_.pdf).
7. Sethuraman N, Jeremiah SS, Ryo A. Interpreting Diagnostic Tests for SARS-CoV-2. *JAMA*. 2020;323(22):2249. doi:10.1001/jama.2020.8259
8. Steffel J, Giugliano RP, Braunwald E, et al. Edoxaban Versus Warfarin in Atrial Fibrillation Patients at Risk of Falling: ENGAGE AF-TIMI 48 Analysis. *J Am Coll Cardiol*. 2016;68(11):1169-1178. doi:10.1016/j.jacc.2016.06.034
9. COG-UK Consortium – COVID-19 Genomics UK Consortium. <https://www.cogconsortium.uk/>. Accessed May 19, 2020.
10. An integrated national scale SARS-CoV-2 genomic surveillance network. *The Lancet Microbe*. 2020;1(3):e99-e100. doi:10.1016/s2666-5247(20)30054-9
11. Lou B, Li T-D, Zheng S-F, et al. Serology characteristics of SARS-CoV-2 infection since exposure and post symptom onset. *Eur Respir J*. May 2020. doi:10.1183/13993003.00763-2020
12. Oude Munnink BB, Nieuwenhuijse DF, Stein M, et al. Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat Med*. 2020;26(9):1405-1410. doi:10.1038/s41591-020-0997-y
13. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579(7798):265-269. doi:10.1038/s41586-020-2008-3
14. Katoh K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002;30(14):3059-3066. doi:10.1093/nar/gk436
15. Maddison DR, Swofford DL, Maddison WP. NEXUS: An Extensible File Format for

- Systematic Information. *Syst Biol*. 1997;46(4):590. doi:10.2307/2413497
16. Leigh JW, Bryant D. POPART: Full-feature software for haplotype network construction. *Methods Ecol Evol*. 2015;6(9):1110-1116. doi:10.1111/2041-210X.12410
  17. Bandelt H-J, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 1999;16(1):37-48.
  18. Public Health England. Regional Coronavirus NHS cases in the UK. <https://coronavirus.data.gov.uk/#category=regions&map=case&area=e12000009>. Published 2020. Accessed May 14, 2020.
  19. Wang J, Zhou M, Liu F. Reasons for healthcare workers becoming infected with novel coronavirus disease 2019 (COVID-19) in China. *J Hosp Infect*. 2020;105(1):100-101. doi:10.1016/j.jhin.2020.03.002
  20. Islam MS, Rahman KM, Sun Y, et al. Current knowledge of COVID-19 and infection prevention and control strategies in healthcare settings: A global analysis. *Infect Control Hosp Epidemiol*. May 2020:1-11. doi:10.1017/ice.2020.237
  21. Wang D, Hu B, Hu C, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA*. 2020;323(11):1061. doi:10.1001/jama.2020.1585
  22. Black JRM, Bailey C, Przewrocka J, Dijkstra KK, Swanton C. COVID-19: the case for health-care worker screening to prevent hospital transmission. *Lancet*. 2020;395(10234):1418-1420. doi:10.1016/S0140-6736(20)30917-X
  23. Hunter E, Price DA, Murphy E, et al. First experience of COVID-19 screening of health-care workers in England. *Lancet*. 2020;395(10234):e77-e78. doi:10.1016/S0140-6736(20)30970-3
  24. Rivett L, Sridhar S, Sparkes D, et al. Screening of healthcare workers for SARS-CoV-2 highlights the role of asymptomatic carriage in COVID-19 transmission. *Elife*. 2020;9. doi:10.7554/eLife.58728
  25. Koh D. Occupational risks for COVID-19 infection. *Occup Med (Chic Ill)*. 2020;70(1):3-5. doi:10.1093/occmed/kqaa036

**Figure 1. Median-joining networks of the viral haplotypes.** Panel A includes only cases from the community hospital outbreak while panel B includes all 141 genomes sequenced at Exeter at that time. Also included in both networks is the Wuhan-Hu-1 reference genome (GenBank: MN908947.3). The genome sequences fell into 61 distinct haplotypes overall (defined by 231 SNPs) and eight distinct haplotypes in the outbreak (defined by 29 SNPs, listed in Supplementary Table 1). Each node represents a single haplotype and its area is proportional to the numbers of samples within that haplotype; the proportions of the circle shaded in red and yellow indicate the proportions of samples in that haplotype from outbreak

and non-outbreak samples respectively. The mutations separating the nodes are indicated by the cross hatches on the connecting edges.

**Figure 2: Haplotypes within outbreak with number of cases in patients and staff.** The

Full details of the haplotypes A – G are listed in Supplementary Table 1.

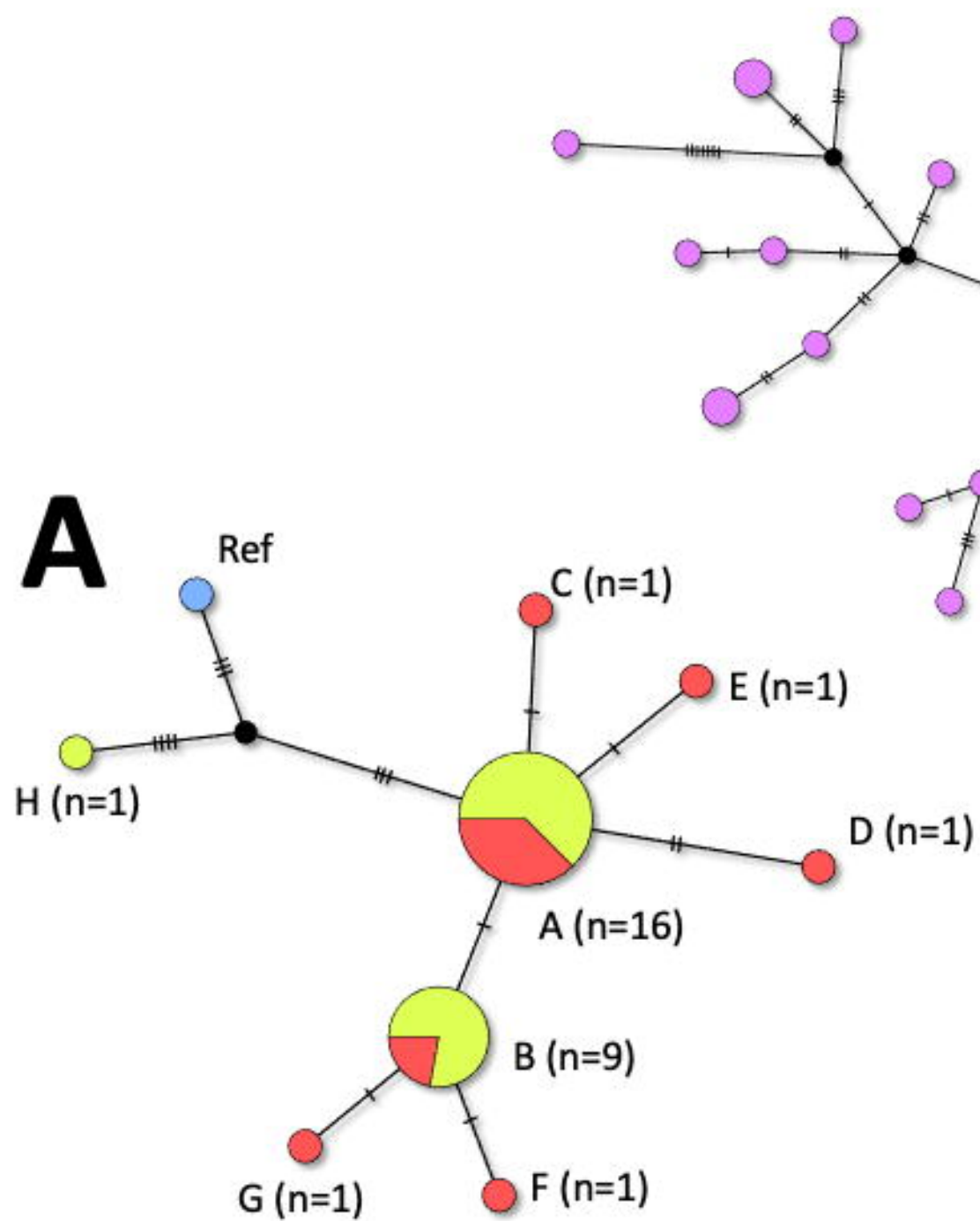
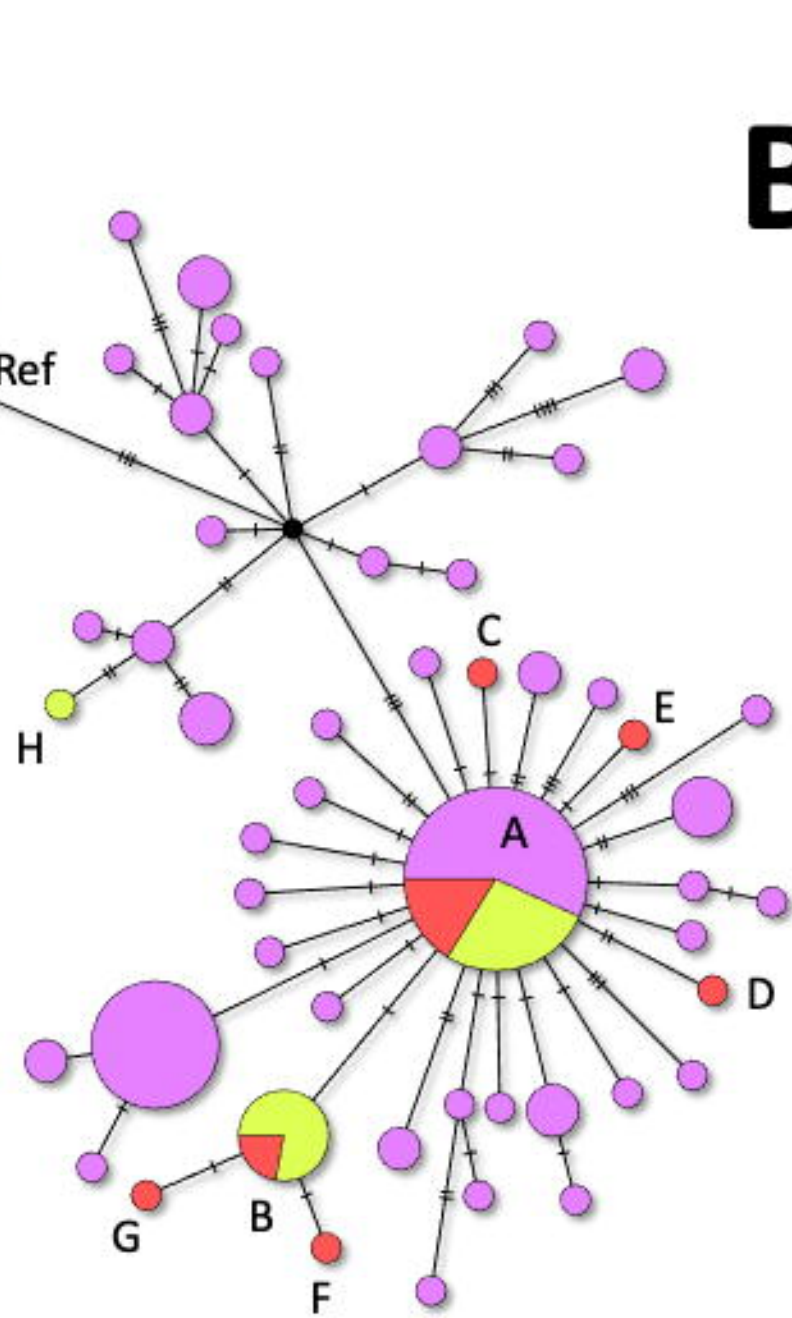
**Figure 3: Timeline for haplotype B cases including patient days in community hospital**

**and staff work days.** Day 0=first confirmed positive case within cluster. Potential exposure within community hospital is demarcated by ‘W’ for work day in staff and ‘IP’ for inpatient at community hospital for patients. Yellow shading represents symptoms. Positive COVID-19 test is represented by a red cross.

### **Supplementary Table 1**

Haplotypes of the sequenced SARS-CoV-2 genomes.



**A****B**



Haplotype C



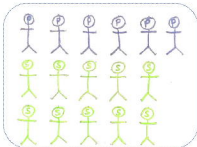
Haplotype D



Haplotype E



Haplotype F



Haplotype A



Haplotype G



Haplotype H



Haplotype B

**Key**



Staff case



Patient case

## Days

	-20	-19	-18	-17	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7			
Patient A	Transferred																IP	IP	IP	IP	X	IP	IP	IP	Discharged						
Patient B	Transferred						IP	IP	IP	IP	IP	IP	IP	IP	IP	IP	IP	IP	IP	IP	IP	IP	IP	IP	IP	IP	IP	IP	X		
Staff		W	W	W	W					W	W	W	W	X																	
staff		W	W	W			W	W			W		W		X																
staff								W			W		W	W	X																
staff		W		W					W			W		W	W	X															
staff		W	W	W	W	W			W	W	W	W	W		W	X															
staff		W			W				W			W	W	W	W		W	W	W	X											
staff			W							W							W				X										