

Retrospective screening of routine respiratory samples revealed undetected community transmission and missed intervention opportunities for SARS-CoV-2 in the United Kingdom

Joseph G. Chappell^{1,2,3}, Theocharis Tsoleridis^{1,2,3}, Gemma Clark⁴, Louise Berry⁵, Nadine Holmes¹, Christopher Moore¹, Matthew Carlile¹, Fei Sang¹, Bisrat J. Debebe¹, Victoria Wright¹, William L. Irving^{1,2,3}, Brian J. Thomson^{2,5}, Timothy C. J. Boswell⁴, Iona Willingham⁴, Amelia Joseph⁴, Wendy Smith⁴, Manjinder Khakh⁴, Vicki M. Fleming⁴, Michelle M. Lister⁴, Hannah C. Howson-Wells⁴, Edward C. Holmes⁶, Matthew W. Loose¹, Jonathan K. Ball^{1,2,3,*}, C. Patrick McClure^{1,2,3} and on behalf of the COG-UK consortium

Abstract

In the early phases of the SARS coronavirus type 2 (SARS-CoV-2) pandemic, testing focused on individuals fitting a strict case definition involving a limited set of symptoms together with an identified epidemiological risk, such as contact with an infected individual or travel to a high-risk area. To assess whether this impaired our ability to detect and control early introductions of the virus into the UK, we PCR-tested archival specimens collected on admission to a large UK teaching hospital who retrospectively were identified as having a clinical presentation compatible with COVID-19. In addition, we screened available archival specimens submitted for respiratory virus diagnosis, and dating back to early January 2020, for the presence of SARS-CoV-2 RNA. Our data provides evidence for widespread community circulation of SARS-CoV-2 in early February 2020 and into March that was undetected at the time due to restrictive case definitions informing testing policy. Genome sequence data showed that many of these early cases were infected with a distinct lineage of the virus. Sequences obtained from the first officially recorded case in Nottinghamshire – a traveller returning from Daegu, South Korea – also clustered with these early UK sequences suggesting acquisition of the virus occurred in the UK and not Daegu. Analysis of a larger sample of sequences obtained in the Nottinghamshire area revealed multiple viral introductions, mainly in late February and through March. These data highlight the importance of timely and extensive community testing to prevent future widespread transmission of the virus.

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a novel zoonotic virus, first identified in the city of Wuhan in the Chinese province of Hubei, following a cluster of patients presenting with severe pneumonia [1]. Since this first detection in December 2019, SARS-CoV-2 has rapidly spread across the globe and, as of 30 March 2021, there have been a total of 127349248 confirmed cases globally, resulting in 2787593 deaths [2]. Within the UK, there have been 4337700 lab-confirmed cases and 126615 deaths, as of 30 March 2021 [2].

Infection with SARS-CoV-2 can lead to the development of coronavirus disease 2019 (COVID-19), characterised by fever, persistent cough, fatigue and shortness of breath [3, 4]. In severe cases, this can progress into acute respiratory distress syndrome (ARDS), which often requires artificial ventilation, and even multi-organ failure and death [5]. Despite these serious potential sequelae, many cases present asymptotically or with only mild disease. Asymptomatic and pre-symptomatic carriage of SARS-CoV-2 is now well documented [6, 7] and transmission has been reported

Received 09 December 2020; Accepted 09 April 2021; Published 16 June 2021

Author affiliations: ¹School of Life Sciences, University of Nottingham, Nottingham, UK; ²NIHR Nottingham Biomedical Research Centre, Nottingham, UK; ³Wolfson Centre for Emerging Virus Research, Nottingham, UK; ⁴Clinical Microbiology, Nottingham University Hospitals, Nottingham, UK; ⁵Nottingham University Hospitals, Nottingham, UK; ⁶Marie Bashir Institute for Infectious Diseases and Biosecurity, University of Sydney, Sydney, Australia.

*Correspondence: Jonathan K. Ball, mrzjb@exmail.nottingham.ac.uk

Keywords: SARS-CoV-2; COVID-19; Molecular Epidemiology; Whole-Genome Sequencing; Pooled Screening; Community Transmission.

Abbreviations: COG-UK, Covid-19 genomics UK consortium; COVID-19, coronavirus disease 2019; NHS, National Health Service; PHE, Public Health England; RT-PCR, reverse transcription polymerase chain reaction; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; WGS, whole genome sequencing; WHO, World Health Organization.

Five supplementary figures are available with the online version of this article.

001595 © 2021 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License. This article was made open access via a Publish and Read agreement between the Microbiology Society and the corresponding author's institution.

[8–10], which is thought to be related to high levels of viral shedding in the upper respiratory tract during the early stages of infection [11]. Due to the difficulty in identifying infected asymptomatic or pre-symptomatic individuals, SARS-CoV-2 has been able to rapidly spread, particularly in healthcare and age care environments [12].

Initial SARS-CoV-2 RT-PCR testing in the UK was offered via referral to Public Health England (PHE) national and regional diagnostic laboratories, and required strict epidemiological and clinical criteria to be met, specifically a recent travel history to Hubei province or contact with a known case and one or more of fever, shortness of breath or new and persistent dry cough. These case definitions were revised on several occasions to include travel to mainland China and several other Asian countries initially (7 February), then expanded further to include Iran and northern Italy (25 February) before finally being removed as essential criteria for diagnostic testing on 12 March [13]. Importantly, a case definition that relied heavily on travel history or exposure to a known infected individual likely resulted in undetected cases and transmission in both healthcare and community settings. A broadening of epidemiological criteria in COVID-19 case definitions was associated with an increased proportion of COVID-19 cases being identified [14]. Furthermore, as SARS-CoV-2 testing was initially only available via PHE laboratories, and testing within NHS laboratories was not rolled out until March 2020, this further restricted the capacity to detect early cases and transmission events.

The first confirmed SARS-CoV-2 case detected in the United Kingdom travelled from Hubei province on 23 January 2020 and became symptomatic on 26 January. This patient then transmitted the virus to a household contact, who also became symptomatic 2 days later [15]. The third diagnosed case was in a traveller returning from Singapore on 6 February; this patient had stopped in France where they infected seven others, before travelling to and seeding several infections in the UK [16]. Retrospective phylogenetic modelling based on genome and associated metadata has conservatively estimated a minimum of 1356 independent introductions of the virus into the UK, primarily from travellers originating in Spain, France or Italy during mid-March [17], although definitive proof of early introduction of SARS-CoV-2 into the UK before widespread testing has been lacking. There have been several reports of early circulation of SARS-CoV-2 in these mainland European countries [18–21], however, a World Health Organization (WHO) report concluded that these reports remain unconfirmed due in part to a lack of sequence data [22]. Similar cases of SARS-CoV-2 have been reported in the UK [23], but the authenticity of these reports have again not been supported with any sequence data.

To better understand the prevalence and emergence of SARS-CoV-2 in the UK before the broadening of case definition criteria and wider testing, retrospective PCR testing of archived diagnostic specimens submitted for respiratory virus screening, review of case histories to identify individuals with symptoms compatible with SARS-CoV-2

infection. To investigate how any undetected cases of SARS-CoV-2 infection may have influenced the genomic epidemiology within Nottingham in the months following the rollout of localised testing, we analysed SARS-CoV-2 genomes sequenced from Nottingham patients between March and June 2020. This study was conducted in a large teaching hospital located in Nottingham and representative of provincial cities throughout the UK. We describe the detection of SARS-CoV-2 from eight patients admitted to hospital with severe respiratory distress who were not tested at the time because they had no travel history or contact with someone infected and therefore did not meet the case definition applied at the time. Sequence analysis of these early cases showed that they belonged to a distinct B-lineage of SARS-CoV-2 which dominated the early phases of the local outbreak. Analysis of further sequences, collected as part of the COG-UK initiative [24] from patients who tested positive after the rollout of local testing, highlighted extensive introductions of the virus into the region.

METHODS

Sample collection

A total of 1660 respiratory specimens (throat swabs, nose swabs, nasopharyngeal aspirates, bronchoalveolar lavages and endotracheal tube secretions) from 1378 patients were collected between 2 January and 11 March 2020 for routine diagnostic investigation [25] for which surplus total nucleic acid was available. To facilitate rapid testing, we created 169 pools each containing up to 10 samples, which were then subjected to SARS-CoV-2-specific PCR.

RT-PCR screening and high-throughput sequence analysis

cDNA was synthesised from each of the nucleic acid pools using RNA to cDNA EcoDry (Random Hexamers) (Takara Bio Europe, Saint-Germain-en-Laye, France). Due to the national shortage of qPCR testing reagents and equipment at the time of this study, the cDNA was initially screened with an in-house PCR assay targeting the RNA-dependent RNA polymerase (RdRp) gene region of SARS-CoV-2, producing an 186 bp amplicon. Positive samples were then confirmed using a larger 366 bp amplicon assay, also targeting the RdRp. All positive samples were further confirmed by both Sanger and whole genome sequencing (WGS) and multiple negative controls were included in each run. For both assays, 5 µl of cDNA was added to PCR reactions containing 5 µl of 10× PCR buffer, 0.25 µl of HotStarTaq DNA Polymerase (QIAGEN, Hilden, Germany), 2 µl each of both forward and reverse primers (10 pmol µl⁻¹), 2 µl dNTPs (10 mM) (Sigma-Aldrich, St Louis, USA) and 33.75 µl of DEPC-treated water. Cycling conditions were 95 °C for 15 min, 55 cycles of 95 °C, 58 and 72 °C for 20 s each, followed by a final extension of 72 °C for 30 s. The primer sequences for the initial 186 bp assay was qCOV19f: 5'-CAATAGCCGCCACTAGAGGA and qCOV19r: 5'-GAGCAAGAACAAGTGAGGCC. The sequences for the larger assays were COV19_Cf:

5'-CGCCACTAGAGGAGCTACTG and COV19_Cr:
5'-GCCGTGACAGCTTGACAAAT.

Positive pools were de-multiplexed and the individual nucleic acid extracts were subject to the same methodology as the pooled samples. WGS was achieved using the ARTIC amplicon sequencing protocol [26], the complete methodology is available in the supplementary information.

Following the implementation of localised NUH SARS-CoV-2 testing using the RealStar SARS-CoV-2 RT-PCR Kit (Altona, Hamburg, Germany), positive samples were routinely subject to WGS. Selection was based on the daily laboratory capacity; when sequencing capacity exceeded sample number, sequencing was attempted from all positive samples with a diagnostic Ct value of 30 or under for one or both gene targets. Where the number of samples exceeded the daily capacity, samples were selected in numerical order to fill capacity, but duplicated ward sources were omitted to increase breadth of sampling.

Phylogenetic analysis

In total, 28124 SARS-CoV-2 whole genomes with >95% coverage were downloaded from GISAID on 10 June 2020. The processing of the genomes was performed with the Geneious Prime 2019.0.4 software. The genomes were sub-divided based on their lineage as indicated by the Pangolin tool [27]. For each lineage, the sequences were aligned and the 5' and 3' ends were trimmed. Maximum likelihood phylogenetic trees were generated for multiple lineages (B.1, B.1.1, B.1.1.1, B.1.p11, B.1.5, B.1.34, B.1.36, B.2, B.2.1, B.2.2, B.2.5, B.2.6, B.3, B.6, B.9, B.10, B.15) using IQ-TREE2 [28] employing the GTR+F+R3 model of nucleotide substitution as suggested by the software's model finder, with 1000 SH-like approximate likelihood ratio test (SH-aLRT) [29]. Each tree was rooted using the early Wuhan virus hCoV-19/Wuhan/WH04/2020|EPI_ISL_406801|2020-01-05 that was sampled in January 2020.

RESULTS

Clinical diagnosis of SARS-CoV-2

Prior to the rollout of localised SARS-CoV-2 testing on 12 March 2020, nine cases (denoted Patients 9–17) matching the clinical and epidemiological contemporary case definitions were identified and referred to PHE for testing (Table 1). Complete sequence data was available for one patient (Patient 9), a traveller returning from South Korea, tested on 28 February 2020.

A review of case histories revealed an additional five individuals whose symptoms were compatible with COVID-19 but who were not tested as they did not meet the epidemiological criteria of the contemporary case definition (Patients 4–8). As these were suspected to have SARS-CoV-2 infection, their respiratory samples were included in the verification of a local SARS-CoV-2 PCR test within the Nottingham diagnostic virology laboratory, and subsequently referred to PHE for confirmatory testing. These patients were all males over the age of 50 whose samples were collected between 2 and 12 March. Complete virus genome

sequences were obtained from all five patients: three were classified as lineage B (Patients 4–6), one as lineage B.2.2 (Patient 7) and one as lineage B.2.5 (Patient 8).

Retrospective detection of SARS-CoV-2 in residual diagnostic material

Residual diagnostic material collected prior to 11 March, and not subjected to the SARS-CoV-2 testing described above, was retrospectively screened for SARS-CoV-2. The majority of these were throat swabs or nasopharyngeal aspirates, with only a small proportion of nasal swabs, sputum or other samples (Fig. S1a, available in the online version of this article). The patient demographics indicated a slightly higher inclusion of males than females (Fig. S1b). The proportion of children under the age of ten were also higher than other ages (Fig. S1c). All samples had originally been sent to the diagnostic virology laboratory for respiratory virus testing. Overall, 41% of the samples contained at least one respiratory virus, with rhinovirus being the most frequently detected, followed by respiratory syncytial virus and human metapneumovirus (Fig. S1d). Non-SARS-CoV-2 coronaviruses were detected in 4.16% of samples. An average of 170 samples per week were collected throughout January, February and the first week of March, peaking at 234 samples collected between 2 and 8 March 2020 (Fig. S1e).

Of the 1660 samples tested, three SARS-CoV-2 positive samples were identified through retrospective PCR screening, each from a separate pool. These positive samples were collected from patients on 21 February (Patient 1), 2 March (Patient 2) and 8 March (Patient 3). All three samples were throat swabs and there were no detectable co-infections with other respiratory viruses. Patient 1 is the earliest known detection of SARS-CoV-2 in Nottingham, collected 1 week before the previous earliest known positive meeting the contemporary case criteria (Table 1). This patient was a 75-year-old female, admitted to hospital following a fall and suffering from respiratory failure. Her condition worsened, requiring ventilation and the patient ultimately developed multi-organ failure and died on 3 March. No recent travel history or contact with a recently returned traveller was identified during care such that the PHE-defined case definition for SARS-CoV-2 testing at the time was not met. Patient 2 was a 64-year-old male being treated for a suspected liver abscess, presenting with repeated fevers, crackles were noted on lung auscultation but there were no abnormalities on chest X-ray. Patient 3 was a 66-year-old male admitted to hospital with a sore throat and symptoms of Guillain-Barre syndrome, with mild abnormalities also noted on a chest X-ray. Both Patients 2 and 3 recovered from their infections.

The complete SARS-CoV-2 genome was sequenced from all three samples; all were classified as lineage B.

Prevalence of SARS-CoV-2 lineages in Nottingham UK following the rollout of localised testing

A total of 1405 patients tested positive for SARS-CoV-2 between 12 March and 2 June. The total number of positives detected per day peaked on 8 April before gradually declining from 22 April onwards (Fig. 1).

Table 1. Clinical details for patients in whom SARS-CoV-2 was detected prior to 12 March 2020. A and E: Accident and emergency department, ICU: Intensive care unit, SOB: Shortness of breath.

Detection method	Patient ID	Sequence no.	Sample type	Patient contact	Collection date	Age	Sex	Travel history	Hospital admission	Tested by NUH	Clinical features	Lineage
Retrospective PCR on residual diagnostic material	1	NOTT-1118C7	Throat Swab	No contact	21/02/20	75	F	None reported	Yes	Yes	A and E admission following a fall, severe respiratory failure. Recent history of rigours and chills in the past week plus some diarrhoea. Admitted to ICU, ventilated, developed fatal multiple organ failure.	B
	2	NOTT-10DF8D	Throat Swab	No contact	02/03/20	64	M	None reported	Yes	No	Possible liver Abscess Repeated fevers Clear Chest X-Ray.	B
	3	NOTT-10F167	Throat Swab	No contact	08/03/20	66	M	None reported	Yes	No	Sore throat Guillain-Barre syndrome Mild Chest X-ray changes.	B
	4	NOTT-10DEBD	BAL	No contact	07/03/20	69	M	None reported	Yes	Yes	3 day fever, headache and confusion. Spouse admitted 2 weeks earlier with same symptoms Worsening O ₂ , ICU admission.	B
	5	NOTT-10DEAE	BAL	No contact	07/03/20	63	M	None reported	Yes	Yes	Fever and shortness of breath for 3 days, fatigue for 7 days. No coryza, cough or myalgia Admitted to ICU and ventilated. Developed fatal multi-organ failure.	B
	6	NOTT-10DECC	Throat Swab	No contact	11/03/20	65	M	None reported	No	Yes	2 day history of fever, sore throat, dry cough, fatigue.	B
	7	NOTT-10DEF9 NOTT-10DEDB	Throat Swab	No contact	11/03/20	54	M	None reported	Yes	Yes	Admitted via A and E. SOB, productive cough, sore throat, mild headache and chest pain. Cough for 2 days.	B.2.2
	8	NOTT-10E48D	Throat Swab	No contact	12/03	70	M	None reported	Yes	Yes	Presented to A and E with SOB, started on the day of admission.	B.2.5

Continued

Table 1. Continued

Detection method	Patient ID	Sequence no.	Sample type	Patient contact	Collection date	Age	Sex	Travel history	Hospital admission	Tested by NUH	Clinical features	Lineage
Referred for PHE Testing	9	hCoV-19/England/2010401504/2020	Throat Swab	10	28/02	38	M	South Korea	No	No	Fever	B
	10	No sequence data	Throat Swab	9	05/03	34	F	Contact	No	No	Community test, no admission	Unknown
	11	No sequence data	Throat Swab	12	03/03	57	M	Northern Italy	No	No	Dry cough and fever	Unknown
	12	No sequence data	Throat Swab	11	06/03	55	F	Northern Italy	No	No	Cough	Unknown
	13	No sequence data	Throat Swab	None	03/03	61	F	Northern Italy	No	No	Cough	Unknown
	14	No sequence data	Throat Swab	None	06/03	72	F	Northern Italy	No	No	Cough	Unknown
	15	No sequence data	Throat Swab	16	10/03	41	F	Switzerland	No	No	Community test, no admission	Unknown
	16	No sequence data	Throat Swab	15	10/03	44	F	Switzerland	No	No	Community test, no admission	Unknown
	17	No sequence data	Throat Swab	None	10/03	34	M	None	No	No	Cough and fever	Unknown

SARS-CoV-2 positive samples were randomly selected for whole-genome sequencing; of these, 679 were successfully sequenced and classified into lineages. A total of 23 lineages were detected, with lineages B, B.1, B.1.1, B.1.1.1 and B.1.p11 the most frequent (Fig. 2a). Lineages B.1, B.1.1 and B were also the three most prevalent lineages globally (Fig. S2). Between 2 March and 3 June 2020, lineages B, B.1.1.1, B.1.1.9, B.1.8, B.1.p11, B.15, B.1.34, B.2.2, B.2.6 and B.6 were proportionally more prevalent in Nottingham than the rest of the UK. Additionally, over 20% of lineage B.1.1.9, B.1.8, B.1.34 and B.6 UK sequences were generated from specimens submitted to Nottingham diagnostic laboratories for testing (Fig. 2b). The diversity of lineages detected each week varied over time. Diversity was greatest between week 12 and 15 (16 March to 12 April) and peaked in week 13, during which 19 lineages were detected. The frequency of lineages B.1.1 and B.1.1.1 increased over time, while the other common lineages (B.1 and B.1.p11) decreased in frequency. The three lineages detected prior to the rollout of localised screening - B, B.2.2 and B.2.5 - all decreased in frequency; B.2.2 was last detected in week 17, B.2.5 in week 13 and B in week 21 (Fig. 3). Fourteen lineages had not been detected in the month prior to 2 June, six of which had not been detected elsewhere in the UK for over a month. Only five lineages were detected in the final week of this study (Fig. S3).

Genomic epidemiology of SARS-CoV-2 in Nottingham

The majority of sequences obtained from patients during the retrospective 'look-back' analysis of archival respiratory samples were assigned to lineage B using the Pangolin tool. These sequences, together with other lineage B sequences sampled in Nottingham (following the introduction of localised screening), the UK and worldwide were then combined and subjected to phylogenetic analysis.

The resulting phylogeny shows that most lineage B Nottingham-derived sequences form a distinct clade within the B-lineage with strong statistical support (Figs 4 and S3). Sequences collected from other regions in the UK (Sheffield, Cambridge and Exeter) and internationally (Iceland, Portugal, USA) were also interspersed throughout this clade, suggesting an epidemiological connection to those infections in Nottingham. The sequence of Patient 1, which was derived from a specimen collected on 21 February, indicates that this individual was part of a transmission chain following the early introduction of this virus strain into the region. Importantly, this sequence clustered with, and predates what was previously thought to be the first case of SARS-CoV-2 infection, which was identified in a traveller returning from South Korea (Patient 9, Table 1).

The retrospectively detected B.2.2 sequence (Patient 7) forms a distinct clade with three other Nottingham-derived sequences detected in later weeks (Fig. S4g), suggesting an epidemiological link between them.

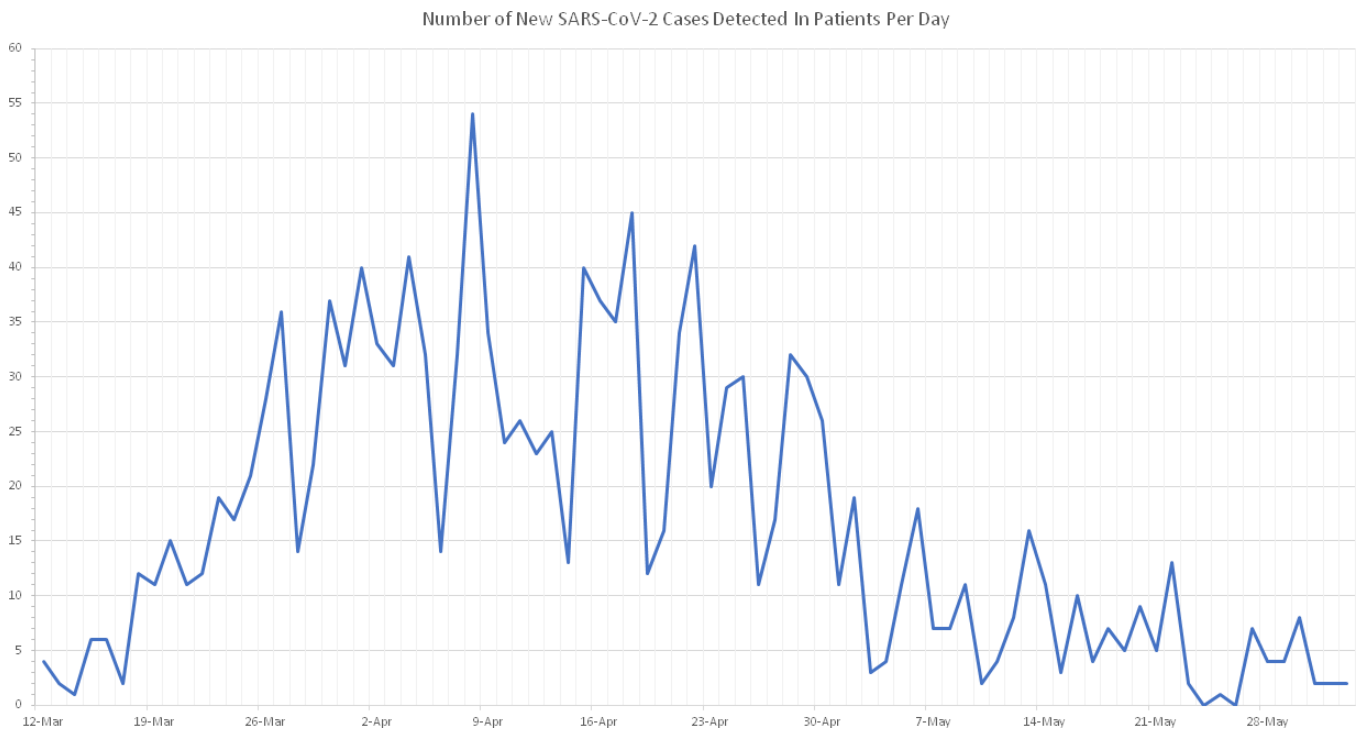


Fig. 1. Number of new SARS-CoV-2 cases detected in Nottingham University Hospitals patients per day, from the rollout of localised testing on 12 March until 2 June 2020.

Phylogenetic analysis of the remaining Nottingham-derived sequences was performed to determine the number of potential introductions of virus into the Nottingham area in the initial stages of the outbreak.

Lineages B.1.5, B.1.36, B.2.6, B.6, B.9 and B.15 were represented by a single sequence, each indicating a unique regional introduction of virus (Fig. S4a–f). Several Nottingham-derived sequences clustered together with sequences representing lineages B.1.34 and B.2.2, again most likely following a single introduction (Fig. S4b, g). Analysis of lineages B.1.1.1, B.1.P11, B.2.5 and B.3 shows that each lineage was independently introduced into Nottingham on at least two occasions (Fig. S4h–k), whilst lineages B.1.1, B.2 and B.2.1 appear to have been introduced at least three times (Fig 4l–n) and lineage B.1 appears to have been introduced on several occasions (Fig. S4o). Poor branch support in several of these phylogenies, lineage B.1 in particular, precludes the accurate determination of number of introductions, and possibly represents an underestimate of the number of unique introductions. On many occasions, Nottingham sequences clustered with sequences from other UK cities and/or other countries, indicating transmission of the virus via travelling, as well as the potential origins and/or destination of the transmission chains. Interestingly, in lineages B, B.1.1, B.1.1.1, B.1.p11, B.1.34, B.2 and B.2.2, small clusters of Nottingham sequences with high SH-aLRT support were observed (data not shown), suggestive of local community transmission chains of the virus.

DISCUSSION

The complete genome sequencing of SARS-CoV-2 combined with retrospective clinical and molecular evaluation has facilitated analysis of the introduction and subsequent prevalence of multiple lineages of the virus into Nottinghamshire in the East Midlands region of the UK. Our study reveals multiple community-acquired cases of SARS-CoV-2 lineage B viruses presenting at a regional healthcare centre, but failing to meet contemporary case criteria for testing in a diagnostic system with restricted capability and up to 1 month before the government imposed countrywide lockdown measures.

Patient 1 in this study is, to the best of our knowledge, the earliest described community-acquired case of SARS-CoV-2 in the UK to be confirmed with viral sequence data, admitted to hospital care on 21 February 2020, and was also the first UK COVID-19 death, preceding the earliest known death by 2 days [30]. A median incubation period estimate of COVID-19 of 5.1 days (4.5–5.8) from infection to the presentation of initial symptoms [31], in addition to the week prior to hospitalisation when the patient was also symptomatic, suggests that infection could have taken place as early as 9 February. This patient had no history of either travel or contact with travellers and so infection must therefore have occurred locally, suggesting an active wider network of community transmission than previously suspected. The first PHE-confirmed case of SARS-CoV-2 in Nottinghamshire meeting contemporary clinical testing criteria, Patient 9, was

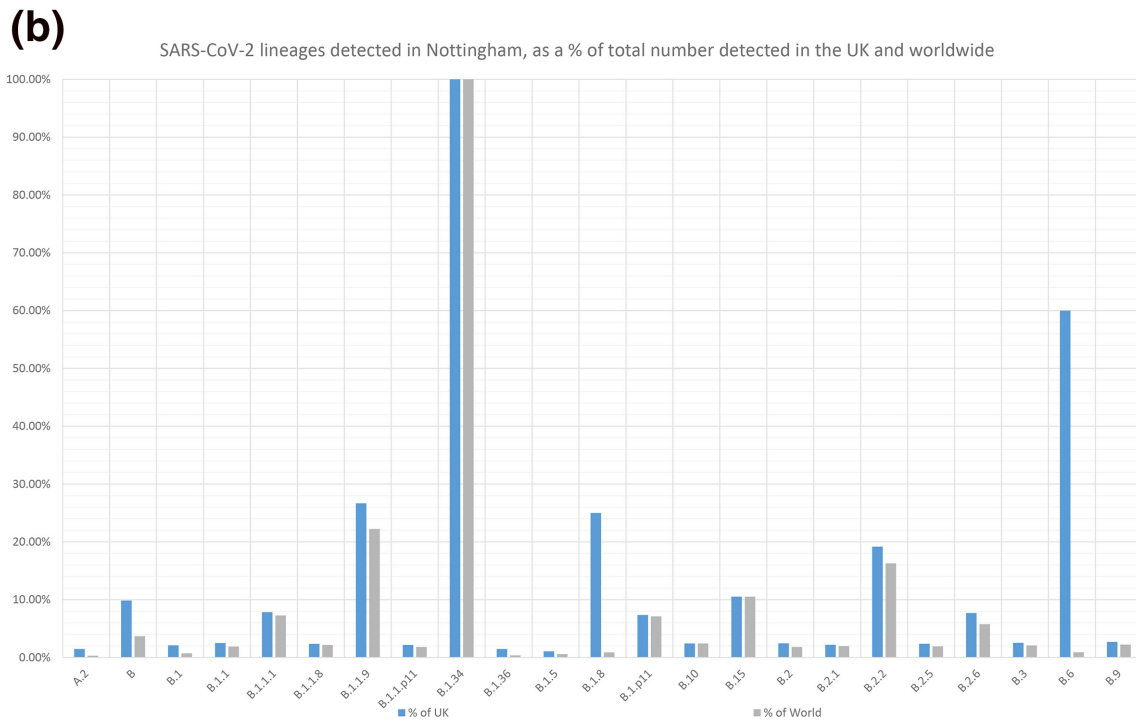
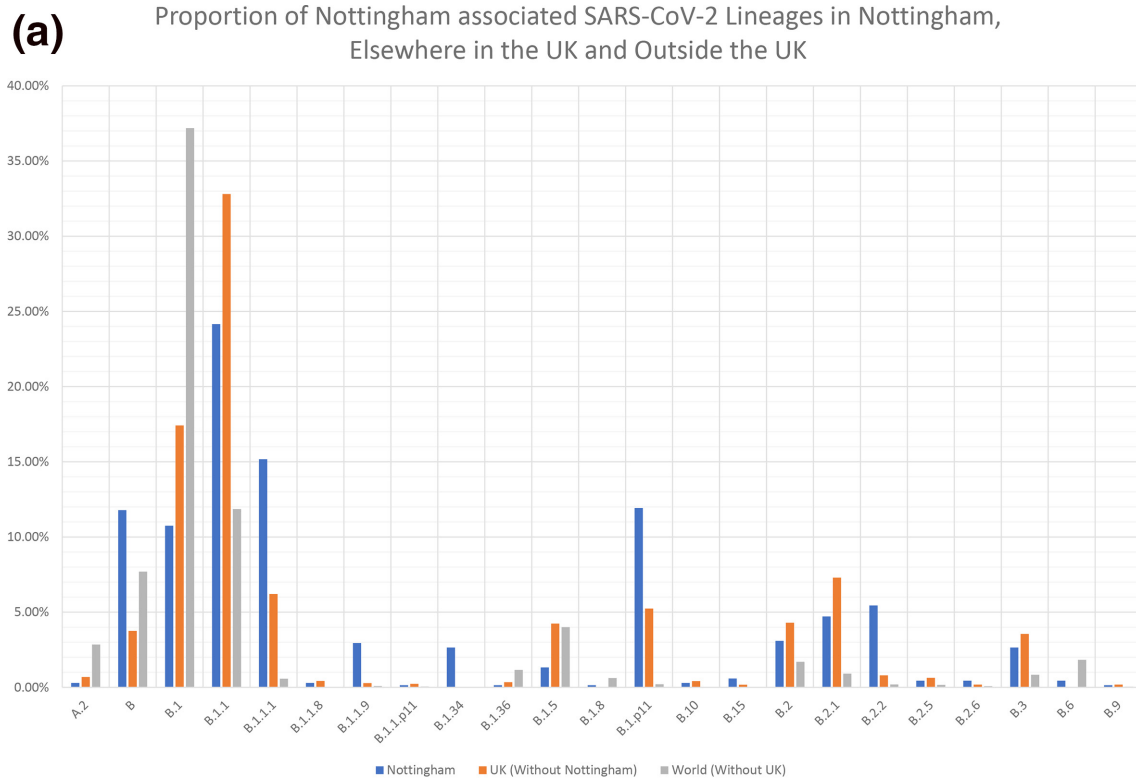


Fig. 2. Proportion of SARS-CoV-2 lineages detected between 21 February and 2 June 2020 in Nottingham, compared with the prevalence of the same lineages in the rest of the UK and elsewhere in the world during the same time period (a). The total count of these lineages is also shown as a percentage of the total count of the same lineages detected in the rest of the UK and world (b).

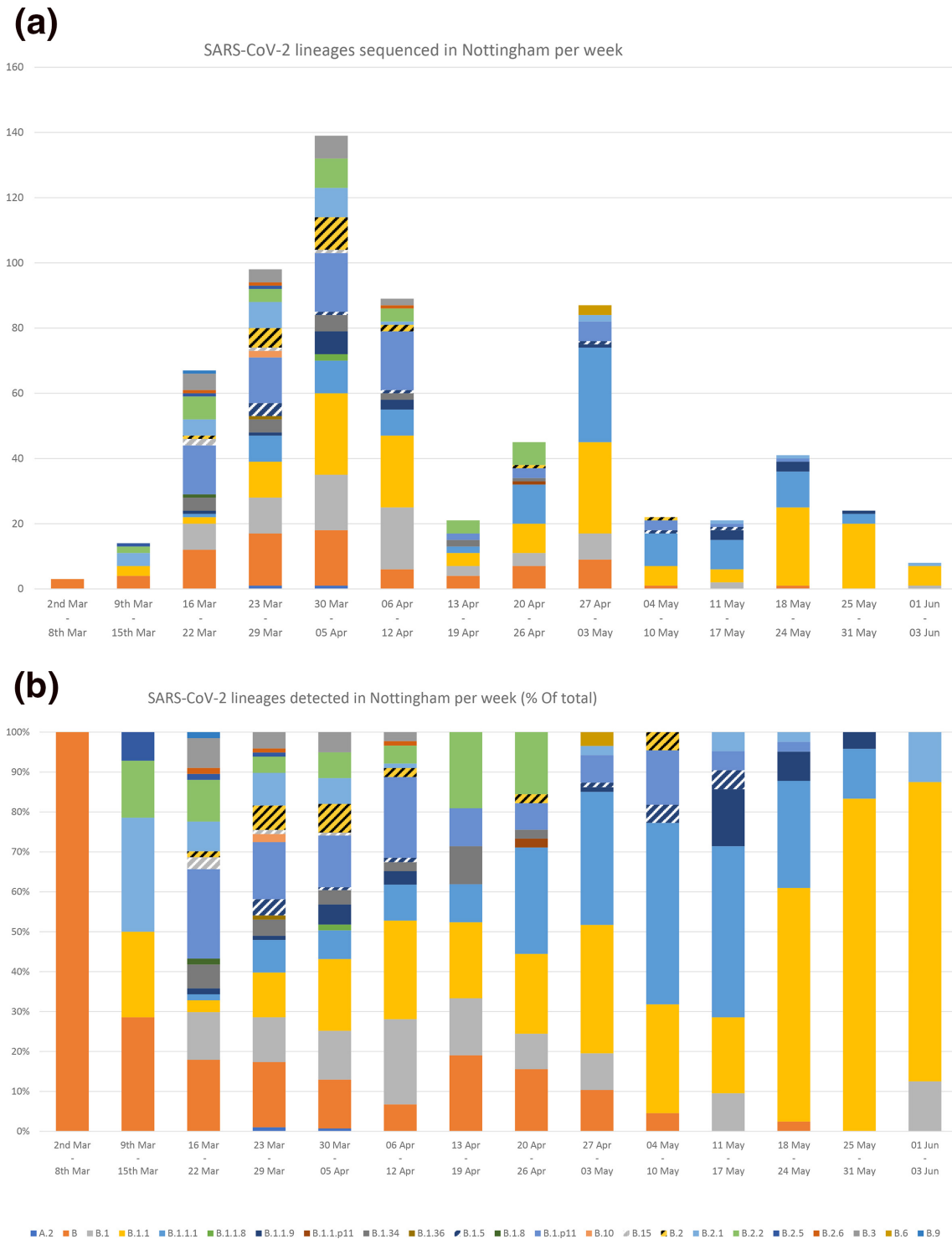


Fig. 3. Frequency of SARS-CoV-2 lineages detected in Nottingham between 21 February and 2 June 2020, per week, as total count (a) and as a proportion of the total number (b).

sampled 1 week after Patient 1. This patient had recently travelled to South Korea, but their associated viral sequence belonged to the same clade of lineage B as Patient 1, which has been infrequently observed in global data sets. The patient

developed a fever 5 days after returning to Nottingham, which is consistent with both acquisition in South Korea and local acquisition in the UK before or after international travel [31]. Locally acquired infection in Nottingham is probably the



Fig. 4. Phylogenetic relationships of SARS-CoV-2 sequences from lineage B based on their entire genome (29412nt). This is an extraction of a sub-tree that contains most of the lineage B sequences detected in Nottingham, including those identified during retrospective PCR testing. The earliest known sequence is shown in green, other sequences identified in patients during the retrospective analysis are shown in purple, sequences obtained from the first confirmed local case (a traveller returning from South Korea) are shown in blue and sequences obtained from other local patients identified after introduction of wide-screen testing are shown in red. For clarity some branches have been collapsed. Where a branch has been collapsed the geographical locations of where the sequences were obtained are shown at the tip, and the number of individual sequences represented is shown in parentheses. The tree was rooted on a Wuhan sequence sampled on 2020-01-05. Reference sequences are indicated by their GISAID accession numbers. Branch lengths are drawn to a scale of nucleotide substitutions per site. Numbers above individual branches indicate SH-aLRT bootstrap support. The complete, unedited lineage B phylogenetic tree is presented in Fig. S4.

most likely scenario, and certainly the most parsimonious explanation, given our evidence that this viral lineage was already circulating in the Nottingham area. This highlights the utility of next-generation sequencing and phylogenetic analysis in delineating the epidemiology of virus outbreaks.

Whilst initially common, the proportion of lineage B in Nottinghamshire fell sharply as the pandemic continued. At the time of writing, lineage B was last detected on 18 May in Nottingham and using the same nomenclature proposal in which lineages were defined, lineages are classed as ‘unobserved’ after a month of not being detected [32].

The frequency of SARS-CoV-2 detection and lineage diversity in Nottingham began to increase during mid-March and peaked during late March/early April. This is consistent

with the findings of Pybus *et al.* [17], where the frequency of TMRCAs (Time of the most recent common ancestor), and therefore transmissions, in the UK peaks during late March. Some virus lineages appeared to have been introduced on multiple occasions and were transmitted widely, whereas others had fewer introductions and/or were associated with limited onward spread. Despite the unprecedented national and global effort, only a fraction of SARS-CoV-2 infected individuals were diagnosed and sequenced, with further loss of sequence information due to a high-quality threshold, such that many novel introductions are likely to be undetected. The complete genome sequence data sets from which the phylogenetic analyses were derived are heavily biased towards UK samples, due to unevenness

in global sequencing activity: by late April 2020, 53% of available genomes were sampled in the UK [33], despite accounting for only 3.4% of reported cases [34]. This imbalance will hinder inferences about the definitive sources of introduction to the UK, but conversely has allowed greater insight into subsequent national SARS-CoV-2 transmission within the UK and onwards to other destinations.

It is likely that the true prevalence of the virus within the local study setting and indeed the wider UK during the time-frame of this study was much higher than we have reported here, particularly from early-February onwards, as our study was limited to symptomatic individuals requiring secondary medical care. Asymptomatically infected individuals, along with those presenting with mild or paucisymptomatic infection, are likely to comprise a significant proportion of the total number of infections. Prevalence of asymptomatic infections in representative American and Icelandic populations is high: between 40 and 45% of total infections [35], and as high as 87.5% in younger cohorts [36]. Early epidemiological modelling predicted that for every hospitalised COVID-19 case in the UK, there may have been a further 120–124 infected individuals [37], whilst 89 and 86% of infections US and China, respectively, were undetected during the early months of the pandemic [38], supporting the hypothesis of a potentially significant incidence of undetected infection within the wider community.

Although a large cohort of surplus diagnostic material dating back to January 2020 was screened, it is possible that further SARS-CoV-2 positives within remain undetected. The pooling of respiratory samples, for example, may have contributed to a decrease in the detection rate. However, we consider this unlikely as similar pooling of samples has been shown to result in only a minor increase in reported Ct-values [39]. Also, the retrospective screening of surplus diagnostic material only captures respiratory samples taken from a specific subgroup of patients: those admitted to hospital predominantly with a suspected respiratory infection. Further false negatives within our cohort may have also arisen as our retrospective assay was not fully evaluated for clinical use and the broad clinical manifestation of COVID-19 was still emerging at the time of investigation.

The salient finding from our study is that simple opportunities to identify early cases of SARS-CoV-2 infection were missed due to overly stringent case-criteria. Had the diagnostic criteria been widened earlier to include patients with compatible symptoms but no travel history, it is likely that earlier imported infections would have been detected, enabling rapid deployment of infection control measures that may have prevented onwards transmission. However, the diagnostic capacity available nationally was not sufficient at the time to process the volume of testing required with a broader case definition. This would have been ameliorated by increased local testing within PHE and NHS diagnostic laboratories earlier in the epidemic. Many of these laboratories possessed the knowledge and experience necessary to initiate local testing but were unable to do so due to a delayed mandate and the availability of commercial testing platforms, on which many diagnostic services rely. For future pandemic

preparedness, the UK urgently needs to invest in and expand diagnostic capacity within NHS and PHE diagnostic laboratory services, ensuring the rapid dissemination of diagnostic protocols, a stable supply chain of reagents and instruments required to sustain a rapid increase in capacity and provide a fully integrated end-to-end result system to ensure appropriate and timely infection prevention and control measures can be taken within secondary care and community settings. Any lasting investment in the human resources and associated infrastructure to achieve a more agile epidemic response both nationally and globally will undoubtedly save lives and drastically reduce the adverse impact of such outbreaks on the economy.

Funding information

This work was supported by grants from the Medical Research Council UK (MR/R010307/1 and MR/S009434/1), which are both part of the EDCTP2 programme supported by the European Union, The University of Nottingham Campaign and Alumni Relations Office research donations award and the EU H2020 programme (Project 727393 – PaleBlu). Whole genome sequencing of SARS-CoV-2 was funded by COG-UK; COG-UK is supported by funding from the Medical Research Council (MRC) part of UK Research and Innovation (UKRI), the National Institute of Health Research (NIHR) and Genome Research Limited, operating as the Wellcome Sanger Institute.

Acknowledgements

The names and affiliations of COG-UK members can be found in the supplementary material available with the online version of this article. We would also like to thank Professor Oliver Pybus, University of Oxford for helpful discussion.

Author contributions

The study was designed by C. P. M., J. G. C and J. K. B. Clinical specimens and details were provided by G. C., M. K., V. M. F., M. M. L., H. H-W., W. S., T. C. J. B., L. B., I. W., A. J. and W. L. I., retrospective RT-PCR screening and preparation of samples for sequencing was completed by J. G. C. and C. P. M., complete genome sequencing was performed by M. L., N. H., C. M., M. C., F. S., B. J. D., and V. M. Data was analysed by T. T., E. H., C. P. M., J. G. C., M. L. and J. K. B. The manuscript was written by J. G. C., and revised by C. P. M., J. K. B, T. T., W. L. I., L. B., G. C., E. H. and M. L. All authors read the final manuscript.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

1. Lu H, Stratton CW, Tang YW. Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. *J Med Virol* 2020;92:401–402.
2. World Health Organization. Who coronavirus disease (COVID-19) Dashboard. Who.int. <https://covid19.who.int/>
3. Chen N, Zhou M, Dong X, Qu J, Gong F et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet* 2020;395:507–513.
4. Huang C, Wang Y, Li X, Ren L, Zhao J et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* 2020;395:497–506.
5. Matthay MA, Aldrich JM, Gotts JE. Treatment for severe acute respiratory distress syndrome from COVID-19. *Lancet Respir Med* 2020;8:433–434.
6. Treibel TA, Manisty C, Burton M, McKnight Áine, Lambourne J et al. COVID-19: PCR screening of asymptomatic health-care workers at London Hospital. *The Lancet* 2020;395:1608–1610.
7. Rivett L, Sridhar S, Sparkes D, Routledge M, Jones NK, et al. Screening of healthcare workers for SARS-CoV-2 highlights the role of asymptomatic carriage in COVID-19 transmission. *eLife* 2020;9.

8. Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G et al. Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *N Engl J Med* 2020;382:970–971.
9. Wei WE, Li Z, Chiew CJ, Yong SE, Toh MP. Presymptomatic Transmission of SARS-CoV-2 - Singapore, January 23-March 16, 2020. *MMWR Morb Mortal Wkly Rep* 2020;69:411–415.
10. Bai Y, Yao L, Wei T, Tian F, Jin D-Y et al. Presumed asymptomatic carrier transmission of COVID-19. *JAMA* 2020;323:1406–1407.
11. Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S et al. Virological assessment of hospitalized patients with COVID-2019. *Nature* 2020;581:465–.
12. Arons MM, Hatfield KM, Reddy SC, Kimball A, James A et al. Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. *N Engl J Med* 2020;382:2081–2090.
13. Public Health England. COVID-19: Investigation and Initial Clinical Management of Possible Cases. Gov.Uk, 2020, pp. 1–5.
14. Tsang TK, Wu P, Lin Y, Lau EHY, Leung GM et al. Effect of changing case definitions for COVID-19 on the epidemic curve and transmission parameters in mainland China: a modelling study. *Lancet Public Health* 2020;5:e296:e289–e296.
15. Lillie PJ, Samson A, Li A, Adams K, Capstick R et al. Novel coronavirus disease (Covid-19): the first two patients in the UK with person to person transmission. *Journal of Infection* 2020;80:578–606.
16. Spiteri G, Fielding J, Diercke M, Campese C, Enouf V, et al. First cases of coronavirus disease 2019 (covid-19) in the WHO European region, 24 January to 21 February 2020. *Euro Surveill* 2020;25
17. Pybus O, Rambaut A, Zarebski AE, Moritz U, Jackson B. Preliminary analysis of SARS-CoV-2 importation & establishment of UK transmission lineages 8; 2020.
18. Deslandes A, Berti V, Tandjaoui-Lambotte Y, Alloui C, Carbonnelle E. SARS-CoV-2 was already spreading in France in late December 2019. *Int J Antimicrob Agents* 2020;55:106006.
19. Carrat F, Figoni J, Henny J, Desenclos JC, Kab S. Evidence of early circulation of SARS-CoV-2 in France: findings from the population-based “CONSTANCES” cohort. *Eur J Epidemiol* 2021;36:219–222.
20. La Rosa G, Mancini P, Bonanno Ferraro G, Veneri C, Iaconelli M et al. SARS-CoV-2 has been circulating in northern Italy since December 2019: evidence from environmental monitoring. *Sci Total Environ* 2021;750:141711.
21. Chavarria-Miró G, Anfruns-Estrada E, Guix S, Paraira M, Galofré B. Sentinel surveillance of SARS-CoV-2 in wastewater anticipates the occurrence of COVID-19 cases. *medRxiv* 2020;2020.06.13.20129627.
22. World Health Organization. Who-convened global study of origins of SARS-CoV-2: China part. <https://www.who.int/publications/i/item/who-convened-global-study-of-origins-of-sars-cov-2-china-part> [accessed 2021].
23. Skopeliti C. Coronavirus: man, 84, 'died from covid-19 in january' as daughter attacks china for 'cover-up.' *The Independent* 2020. <https://www.independent.co.uk/news/uk/home-news/coronavirus-uk-china-death-peter-attwood-b421030.html>
24. The COVID-19 Genomics UK (COG-UK) consortium. An integrated National Scale SARS-COV-2 Genomic surveillance Network. *The Lancet Microbe* 2020.
25. Bagasi AA, Howson-Wells HC, Clark G, Tarr AW, Soo S. Human bocavirus infection and respiratory tract disease identified in a UK patient cohort. *J Clin Virol* 2020;104453.
26. GitHub - artic-network/artic-ncov2019: ARTIC nanopore protocol for nCoV2019 novel coronavirus. 2019. <https://github.com/artic-network/artic-ncov2019>
27. O'Toole A, McCrone J. Hcov-2019/pangolin: Software package for assigning SARS-COV-2 genome sequences to global lineages. 2019. <https://github.com/hCoV-2019/pangolin>
28. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* 2020;37:1530–1534.
29. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 2010;59:307–321.
30. CMO for England announces first death of patient with COVID-19-GOV.UK. 2020. <https://www.gov.uk/government/news/cmo-for-england-announces-first-death-of-patient-with-covid-19>
31. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q. The incubation period of coronavirus disease 2019 (CoVID-19) from publicly reported confirmed cases: Estimation and application. *Ann Intern Med* 2020;172:577–582.
32. Rambaut A, Holmes EC, Hill V, OToole A, McCrone J. A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology. *bioRxiv* 2020;2020.04.17.046086.
33. Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J. Micro-react: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb Genom* 2016;2:e000093e000093.
34. World Health Organization. Who Coronavirus (covid-19) dashboard. <https://covid19.who.int/>
35. Oran DP, Topol EJ. Prevalence of Asymptomatic SARS-CoV-2 Infection : A Narrative Review. *Ann Intern Med* 2020;173:362–367.
36. Lytras T, Dellis G, Flountzi A, Hatzianastasiou S, Nikolopoulou G, et al. High prevalence of sars-cov-2 infection in repatriation flights to Greece from three european countries. *J Travel Med* 2020;27
37. Jit M, Jombart T, Nightingale ES, Endo A, Abbott S et al. Estimating number of cases and spread of coronavirus disease (COVID-19) using critical care admissions, United Kingdom, February to March 2020. *Eurosurveillance* 2020;25:4–8.
38. Wu SL, Mertens AN, Crider YS, Nguyen A, Pokpongkiat NN. Substantial underestimation of SARS-CoV-2 infection in the United States. *Nature Communications* 2020;11:1–10.
39. Lohse S, Pfuhl T, Berkó-Göttel B, Rissland J, Geißler T. Pooling of samples for testing for SARS-CoV-2 in asymptomatic people. *The Lancet Infectious Diseases* 2020;3099:2019–2020.

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.