Genomic epidemiology of the SARS-CoV-2 epidemic in Zimbabwe: Role of international travel and regional migration in spread

Authors: Tapfumanei Mashe^{1*}, Faustinos Tatenda Takawira¹, Leonardo de Oliveira Martins², Muchaneta Gudza-Mugabe³, Joconiah Chirenda⁴, Manes Munyanyi⁶, Blessmore V Chaibva⁵, Andrew Tarupiwa¹, Hlanai Gumbo¹, Agnes Juru¹, Charles Nyagupe¹, Isaac Phiri⁶, Portia Manangazira⁶, Alexander Goredema⁶, Sydney Danda⁶, Israel Chabata⁶, Janet Jonga⁶, Rutendo Munharira⁶, Kudzai Masunda⁷, Innocent Mukeredzi⁷, Douglas Mangwanya⁶, The COVID-19 Genomics UK (COG-UK) Consortium^{9,#}, SARS-CoV-2 Research Group^{10,#}, Alex Trotter², Thanh Le Viet², Gemma Kay², David Baker², Gaetan Thilliez², Ana-Victoria Gutierrez², Justin O'Grady², Maxwell Hove⁵, Sekesai Mutapuri-Zinyowera¹, Andrew J. Page², Robert A. Kingsley^{2,8}, Gibson Mhlanga⁶

Affiliations:

1. National Microbiology Reference Laboratory, Ministry of Health and Child Care, Zimbabwe, P.O. Box ST 749, Southerton, Harare, Zimbabwe.
2. Quadram Institute Bioscience, Norwich, United Kingdom
3. World Health Organization, Zimbabwe, 82-84 Enterprise Road, Highlands P.O. Box CY 348 Causeway, Harare, Zimbabwe
4. Department of Surgical Sciences, Faculty of Medicine and Health Sciences, University of Zimbabwe
5. Directorate of Pharmacy Services, Ministry of Health and Child Care, Zimbabwe
6. Ministry of Health and Child Care, Zimbabwe
7. Harare City Health Department, Beatrice Road Infectious Diseases Hospital, Harare, Zimbabwe
8. School of Biological Sciences, University of East Anglia, Norwich, United Kingdom
9. <https://www.cogconsortium.uk>
Full list of COG-UK consortium and names and SARS-CoV-2 Research Group affiliations are in the appendix
10. SARS-CoV-2 Research Group

***Corresponding author.** E-mail: mashet2006@yahoo.co.uk

Keywords: sequencing, virus, transmission, epidemiology, COVID-19

Abstract

Zimbabwe reported its first case of SARS-CoV-2 infection in March 2020, and case numbers increased to more than 8,099 to 16th October 2020. An understanding of the SARS-CoV-2 outbreak in Zimbabwe will assist in the implementation of effective public health interventions to control transmission. Nasopharyngeal samples from 92,299 suspected and confirmed COVID-19 cases reported in Zimbabwe between 20 March and 16 October 2020 were obtained. Available demographic data associated with those cases identified as positive (8,099) were analysed to describe the national breakdown of positive cases over time in more detail (geographical location, sex, age and travel history). The whole genome sequence (WGS) of one hundred SARS-CoV-2-positive samples from the first 120 days of the epidemic in Zimbabwe was determined to identify their relationship to one another and WGS from global samples. Overall, a greater proportion of infections were in males (55.5%) than females (44.85%), although in older age groups more females were affected than males. Most COVID-19 cases (57 %) were in the 20-40 age group. Eight lineages, from at least 25 separate introductions into the region were found using comparative genomics. Of these, 95% had the D614G mutation on the spike protein which was associated with higher transmissibility than the ancestral strain. Early introductions and spread of SARS-CoV-2 were predominantly associated with genomes common in Europe and the United States of America (USA), and few common in Asia at this time. As the pandemic evolved, travel-associated cases from South Africa and other neighbouring countries were also recorded. Transmission within quarantine centres occurred when travelling nationals returning to Zimbabwe. International and regional migration followed by local transmission were identified as accounting for the development of the SARS-CoV-2 epidemic in Zimbabwe. Based on this, rapid implementation of public health interventions are critical to reduce local transmission of SARS-CoV-2. Impact of the predominant G614 strain on severity of symptoms in COVID-19 cases needs further investigation.

Introduction

'Severe Acute Respiratory Syndrome Coronavirus 2' (SARS-CoV-2) was confirmed as the causative agent of 'Coronavirus Disease 2019' (COVID-19) (Huang *et al.*, 2020). The virus spread to more than 200 countries, led to more than 47 million confirmed infections and more than 1.2 million deaths (as of 4th of November 2020). The COVID-19 pandemic constitutes the largest global public health crisis in a century, with daunting health and socioeconomic challenges (Khandelwal, 2020). Countries in the African continent have particular challenges with respect to controlling COVID-19 infections and reducing mortality rates; coupled with weak public health delivery systems they have large populations of vulnerable people e.g. those immunocompromised from HIV/AIDS, with non-communicable diseases such as diabetes, hypertension and anaemia, or malnutrition (Kaseje, 2020).

The first COVID-19 case in Africa was recorded in Egypt on the 14th of February, 2020 (Lone and Ahmad, 2020), and on 27th February, the sub-Saharan African region recorded its first case in Nigeria. Further cases in sub-Saharan Africa were recorded in South Africa and Zimbabwe, on March 5th and 20th, respectively (Msomi *et al.*, 2020). The initial cases identified in Africa were mostly introduced from Europe, the Middle East and the United States of America (Hâncean *et al.*, 2020; Lone and Ahmad, 2020; Mehtar *et al.*, 2020).

On 30th March, after confirmation of the first death (24th March 2020) Zimbabwe introduced a 21-day self-quarantine for citizens arriving from affected territories. On the 23rd of March 2020, the country banned: non-essential in-bound travel and transportation except for returning residents; entertainment and recreational activities; public gatherings of > 50 people; and limited hospital visits to one relative per patient per day. Other measures implemented included compulsory wearing of face masks in all settings, social distancing and hand sanitization. A national COVID-19 coordinating committee chaired by a Cabinet Minister, with a full time senior public health specialist as its coordinator, was set up. At all levels from province to district, coordinating committees were established. Rapid Response Teams at all levels were set up to investigate presumed COVID-19 cases, follow-up on all contacts, collect samples from contacts and supervise the burial of confirmed COVID-19 deaths. Treatment centres were set up, mainly Harare and Bulawayo city. Due to the lockdowns imposed in other countries, there was an influx of Zimbabwean residents returning home. This led to establishment of Government run quarantine centres. Initially, all cases housed at quarantine centres had follow up laboratory testing on day 1, 7 and 21 to assess infection and recovery rate. After having completed the 21-day quarantine period with negative results, the returnees/travellers were discharged. The intervention meant that even those who were not showing signs and symptoms of COVID-19 were tested. On 30th

March, total shutdown of all services, except essential services like health care centres, grocery stores and government activities was implemented.

The COVID-19 pandemic, and the actions taken in response to it, disrupted healthcare services and, as a result, overwhelmed the major healthcare programmes in Africa focused on controlling HIV, tuberculosis and malaria. Facilities and staff are taken up with attending to the overwhelmingly high demand for care required by patients with severe COVID-19, leaving other patients without the resources they need. This will have far-reaching consequences for other diseases, poverty, food security, and economic growth in Africa (Walker *et al.*, 2020).

There is an urgent need to understand the epidemiology of SARS-CoV-2 in Africa, particularly the dynamics of its virulence, transmission, mutation capacity and symptom variation amongst infected people (Kalk and Schultz, 2020). The SARS-CoV-2 mutation rate is estimated to be approximately 2.5 single nucleotide polymorphisms (SNPs) per month (Duchene *et al.*, 2020; Meredith *et al.*, 2020). Genome sequencing of SARS-CoV-2 is important to monitor these changes because they are relevant to vaccine efficacy, can be used to infer transmission dynamics, and inform future public health measures targeting its spread. Genomic surveillance backed up with detailed epidemiological data on SARS-CoV-2 can identify which lineages of the virus are circulating in the human population. Information on how these change over time can serve as an indicator of the success of control measures, how often new sources of virus are introduced from other geographical areas, and how the virus has evolved in response to interventions. Despite having inadequate health systems, the number of reported cases in many African countries, specifically Zimbabwe, has been rising more slowly than predicted (Nkengasong and Mankoula, 2020).

Zimbabwe took the initiative to sequence SARS-CoV-2 samples from 100 patients collected over the period March to June 2020. The objectives were to use the sequence results to: understand initial transmission events; understand early domestic transmission of the virus in Zimbabwe; add context to the regional and global data; and to evaluate the role of rapid whole genome sequencing for outbreak analysis in this setting.

Material and methods

Epidemiological analysis. Nasopharyngeal samples (92,299 in total including repeats) were collected from suspected COVID-19 cases across Zimbabwe between the 20th March (first case identified) and the 16th October 2020 were diagnosed routinely at nine local

laboratories to various regions affected using previously described PCR methods to confirm positive cases (Corman *et al.*, 2020). Additional laboratories were established during the national scale up to help management of COVID-19 pandemic. At sampling the demographic data including gender, age and history of travel were recorded. Demographic, clinical and laboratory data associated with the positive cases (8,099) was collated and analysed in order to describe the epidemiology of SARS-CoV-2 using an analytical cross-sectional study design. Specifically, we determined the number of SARS-CoV-2 infections detected on each day since the first case in Zimbabwe, the distribution of detected cases in ten geographical regions of Zimbabwe, the number of infections in males and females, and the number of infection by age stratification. Access to these data (and associated clinical samples used for genomic analysis) was through coordination with teams from the Epidemiology and Diseases Control Unit of the Ministry of Health and Child Care of Zimbabwe.

Selection of samples for genomic evaluation. Samples from the first cases (244) by date of collection (20th March to 30th June) that gave a positive PCR test were considered for genome sequencing. Samples that had a Ct value of less than 30 in the original PCR test were processed for sequencing as described below. Of these, 100 samples provided good quality sequence (<30% N's) were used analysed by genome sequence comparison and phylogenetic analysis.

RNA extraction. Samples for whole genome sequencing were processed using the QuantStudio 3 (Applied Biosystems) diagnostic platform at the National Microbiology Reference Laboratory (NMRL), Zimbabwe. RNA was extracted from nasopharyngeal swab samples using the NucliESNS easyMag (Biomerieux) instrument according to the manufacturer's instructions. Left over aliquots of the RNA used for realtime PCR detection were transported to UK for sequencing. The lower cycle threshold (Ct) or take-off value produced by the SARS-CoV-2 assays in the QuantStudio 3 assays were used to determine whether samples needed to be diluted according to the ARTIC protocol (Quick, 2020).

ARTIC SARS-CoV-2 multiplex tiling PCR. cDNA and multiplex PCR reactions were prepared following the ARTIC nCoV-2019 sequencing protocol v2 (Quick, 2020). Dilutions of RNA were prepared when required based on Ct values following the guidelines from the ARTIC protocol. V3 CoV-2 primers (https://github.com/artic-network/artic-ncov2019/tree/master/primer_schemes/nCoV-2019/V3) were used to perform the multiplex PCR for SARS-CoV-2 according to the ARTIC protocol (Quick, 2020) with minor changes. Due to variable Ct values all samples in the two ARTIC multiplex PCRs were run for 35 cycles. Odd and even PCR reactions were pooled and cleaned using a 1x SPRI bead clean

using KAPA Pure Beads (Roche Catalogue No. 07983298001) according to manufacturer instructions. PCR products were eluted in 30 µl of 10 mM Tris-HCL buffer, pH 7.5. cDNA was quantified using QuantiFluor® ONE dsDNA System (Promega, WI, USA). Libraries were then prepared for sequencing on the Illumina or Nanopore platform and sequenced as described previously (Baker *et al.*, 2020). All sequence data are freely available in the GISAID database (Supplementary Table 1 and Supplementary Table 2). Sequence data of SARS-CoV-2 from South Africa was reported previously (Msomi *et al.*, 2020).

Sequence analysis. The raw reads were demultiplexed using bcl2fastq (v2.20) (Illumina Inc.) allowing for zero mismatches in the dual barcodes to produce FASTQ files. The reads were used to generate a consensus sequence for each sample using an open source pipeline adapted from <https://github.com/connor-lab/ncov2019-artic-nf> (<https://github.com/quadram-institute-bioscience/ncov2019-artic-nf/tree/qib>). Briefly, the reads had adapters trimmed with TrimGalore (<https://github.com/FelixKrueger/TrimGalore>), were aligned to the Wuhan Hu-1 reference genome (accession MN908947.3) using BWA-MEM (v0.7.17) (Li, 2013), the ARTIC amplicons were trimmed and a consensus built using iVar (v.1.2) (Grubaugh *et al.*, 2019).

Quality control of the sequenced genomes. Samples were prepared and sequenced in 96-well plates with one cDNA negative control per plate and one RNA extraction negative control, where applicable. Contaminated samples were removed from analysis. Consensus sequence was defined as passing quality control if greater than 50% of the genome was covered by confident calls or there was at least one contiguous sequence of more than 10,000 bases and no evidence of contamination in the negative control. This is regarded as the minimum amount of data to be phylogenetically useful. A confident call is defined as having 10x depth of coverage. If the coverage falls below these thresholds, the bases are masked with Ns. Low quality variants are also masked with Ns. The QC threshold for inclusion in Global Initiative on Sharing All Influenza Data (GISAID) database is higher, requiring that greater than 90% of the genome is covered by confident calls and that there is no evidence of contamination.

Clustering and phylogenetic analysis. Lineages (Rambaut *et al.*, 2020) assigned to each consensus genome were determined using Pangolin (<https://github.com/cov-lineages/pangolin>) from the Rambaut group on MRC CLIMB (Connor *et al.*, 2016) using pangoLEARN training data set from 2020.08.29, with uncertain classifications (low support or discordance from tree) curated by hand. Only samples that passed quality control were considered for genomic analysis. Global sequences (Supplementary Table 2) that were potentially related to those from Zimbabwe were included in the phylogenetic analysis,

leading to 6467 genomes being included using the software peroba (<https://github.com/quadram-institute-bioscience/peroba>). All sequences were aligned against the reference Wuhan Hu-1 using MAFFT (Kato and Standley, 2013), and the maximum likelihood tree was estimated using fasttree and refined with iqtree2 under the HKY+G model (Price *et al.*, 2010; Minh *et al.*, 2019). Using PastML (Ishikawa *et al.*, 2019) for ancestral state reconstruction, local outbreaks were found and redundant global sequences were removed for visualisation purposes.

Results:

Epidemiology of SARS-CoV-2 infections in Zimbabwe from March to October 2020.

Epidemiological analysis was based on 8,099 positive cases identified from the 92,299 national samples made available between the 20th March (first case) and 16th October 2020. Bulawayo had the highest infection rate at 221 per 100,000 population, followed by Harare at 144 per 100,000 population and then Matabeleland South at 106 cases per 100,000 population. Manicaland, Harare and Bulawayo provinces had the highest case fatality rates (CFR): 4.4%, 3.7% and 3.0% of the population, respectively (Table 1). Matabeleland South had the lowest CFR: 0.9%. The country's overall CFR rate was 3% with almost half of the deaths (47%) happened in the community rather than in hospitals.

Table 1. SARS-CoV-2 cases and deaths per province (as of 16th October 2020)

Province	Population (est. 2019)	Total number of cases	Infection rate (per 100,000 population)	Number of deaths	Case Fatality Rate (CFR)
Bulawayo	700,466	1,545	221	47	3.0
Harare	2,276,285	3,286	144	121	3.7
Manicaland	1,879,129	499	27	22	4.4
Mashonaland Central	1,235,657	208	17	4	1.9
Mashonaland East	1,441,974	412	29	4	1.0
Mashonaland West	1,609,978	350	22	12	3.4
Midlands	1,731,435	644	37	9	1.4
Masvingo	1,592,217	236	15	2	0.8
Matabeleland North	803,048	141	18	3	2.1
Matabeleland South	733,226	778	106	7	0.9
Total	14,003,415	8,099	58	231	2.9

The majority of cases recorded prior to 20th of July were travel associated. The number of cases with no recent travel reported remained below 50 during this period of time. People travelling from South Africa accounted for the greatest proportion (83.4%) of travel-associated cases (Fig. 1). From the time the first case was recorded to about 120 days later, there were more travel-associated cases than locally-transmitted cases, and the majority of travel-related cases were asymptomatic. During July the number of cases with no recent travel reported increased rapidly, confirming the start of community transmission. An increase in overall number of daily cases was observed from day 60 after the first case and peaked during the first half of August. Thereafter the number of cases decreased to the middle of October when approximately 20 cases per day were identified.

Figure 1. Epidemic curve of SARS-CoV-2 confirmed cases as of 16 October 2020. (A) Cumulative number of positive tests for SARS-CoV-2 in individuals reporting recent travel (probable imported cases, green line) or no recent travel (domestic cases, blue line). **(B)** Number of positive tests for SARS-CoV-2 by day (red bar).

Of those infected, 4455 (55%) were men and 3644 (45%) were women. Health workers constituted about 8% of the total number of reported cases. The most affected age group was the 31-40 years group (Fig. 2). The proportion of cases recorded in all age group categories were slightly greater amongst men than women except in the 91-100 years age group (Fig. 2). The greatest skew towards men was in the age range 41-50 while in females it was amongst the 91-100 age group. Overall, most deaths occurred amongst men (59%) and the largest number of deaths were in the 41-50 years age group. The CFR increased from age-group 41-50 to >90 years. The under-1-year CFR was 5% (Fig. 2).

Figure 2. Relationship between age, gender and infection with SARS-Cov-2. (A) number of positive cases in each age group. (B) number of deaths in each age-group. (C) overall case fatality rate (CFR) in each age group. Cases for female (Blue bars) and male (green bar) are indicated. The small infographic in **A** and **B** represents the overall percentage of each gender infected (blue = female, green = male).

At least 25 independent introductions of SARS-CoV-2 into Zimbabwe was associated with eight global lineages. The whole genome sequence of 100 SARS-CoV-2 samples from the first 3 months of the epidemic in Zimbabwe were analysed to investigate their phylogenetic relationship and infer spread. To place the genomes from Zimbabwe SARS-CoV-2 samples into the context of closely-related genomes from global samples a maximum likelihood phylogenetic tree of 6467 genomes and subsamples was constructed (Figure 3). The 100 genomes from Zimbabwe could be placed within eight previously defined high-order global sub-lineages (Table 2). Four samples (4%), were from high-order lineage A, with the remaining genomes (n=96, 96%) from lineage B. The most common sub-lineage was B.1.1 (51% of samples), followed by lineage B.1 (25% of samples) (Table 2).

Table 2: Summary of global lineages circulating in Zimbabwe, South Africa and Globally (as of 2020-11-08).

Lineage	Zimbabwe (n=100)		South Africa (n=420)		Global (n=62138)	
	no. samples	Percent	no. samples	Percent	no. samples	Percent
A	4	4	1	0.08	711	0.5
B.1	25	25	61	4.98	21383	15
B.1.1	51	51	274	22.37	31648	22.1
B.1.5	9	9	45	3.67	6253	4.4
B.1.23	6	6	0	0	12	0.01
B.1.1.57	3	3	20	1.63	22	0.02
B.1.1.62	1	1	18	1.47	24	0.02
B.2	1	1	1	0.08	2083	1.5

A maximum likelihood tree based on variation in the nucleotide sequence revealed a population structure for Zimbabwean cases of COVID-19 that grouped into 12 clusters each comprised of multiple genomes that emerged from nodes distributed throughout the tree; 15 genomes were present as a separate lineage (Figure 3). If each of these clusters

represented separate introductions then we can infer that 27 was the lower limit for the number of independent introductions evident and for which there was good supporting evidence. However, in the case of four clusters or single genome lineages, the tree topology suggested that the genomes arose by divergence from just two introductions into Zimbabwe. In each case, two lineages shared a common ancestor with no intermediary nodes formed with genomes from other countries (blue and green lineages in Figure 3). By this more conservative estimation, the minimum number of introductions into Zimbabwe that our data support was 25. The 12 clusters were each comprised between 2 and 30 genomes with 0-3 SNPs between members of the cluster. Seven clusters contained more than one genome that had an identical genome sequence to at least one genome in the same cluster.

Two large clusters within lineage B.1.1 were present on extended branches of the phylogenetic tree composed of genomes that were distinct from all other to any genome present in the GISAID database by 4-5 SNPs. These introductions may have been from geographical locations where sequence of SARS-CoV-2 was not reported or the result of sequence divergence during transmission within Zimbabwe from an earlier case within Zimbabwe not included in our analysis.

Figure 3. Phylogenetic relationship of SARS-CoV-2 genomes from Zimbabwe in context with closely related genomes from global cases. Maximum Likelihood phylogenetic tree of SARS-CoV-2 genomes from Zimbabwe in the context of 465 of the most closely-related genomes from global samples. Lineages connecting Zimbabwe clusters (red) and global lineages (black) are indicated. Country (inner circle) and continent (middle circle) of sampling, and global lineage designation (outer circle) are colour coded as indicated in the key (insert).

The D614G genotype was the most frequent variant in early spread events in Zimbabwe. There has been considerable interest in the emergence of a variant lineage of SARS-CoV-2 with a non-synonymous mutation resulting in substitution of an aspartate with a glycine residue at position 614 (D614G) of the spike (S) protein; this variant is linked to increased transmissibility in the upper respiratory tract and increased fitness (Plante *et al.*, 2020; Volz *et al.*, 2020). We therefore determined the distribution and frequency of the D614G variant in SARS-Cov-2 genomes from Zimbabwe. A total of 95 samples out of 100 contained the D614G mutation. Four of the five samples with the ancestral 614D form were

from a single cluster of identical genomes from lineage A which was associated with travel from Dubai; the other was from lineage B.2 and associated with travel from USA. All five were from cases presenting in April or May 2020.

Genomic epidemiology identifies intercontinental transmission and local transmission in the first 120 days of the epidemic.

National surveillance in Zimbabwe identified the first three cases of COVID-19 in the second half of March 2020 in three individuals with a recent history of travel from outside of Zimbabwe. These patients had arrived from the United Kingdom (UK), United States (US) and Dubai and corresponded to samples ZW25, ZW29 and ZW70, respectively (Figure 4). Samples ZW25 and ZW29 were indistinguishable from one another suggesting a recent common source or direct transmission. However, this genome sequence was extremely common in the GISAID database (Shu and McCauley, 2017) where identical genomes have been reported from multiple countries, although predominantly Europe, consistent with rapid transmission of this clone internationally (Figure 4).

Genome ZW70 was from a traveller arriving from Dubai and was identical to three other genomes: ZW168, ZW169 and ZW214. Local epidemiological investigation indicated that these were infections resulting from cohabitation and may represent the earliest cases of local transmission. This interpretation was supported by the genome sequence data. Dubai airport is an important hub for flights from many locations in the world including Asia. Genomes ZW168, ZW169 and ZW214 all belonged to lineage A, which is most widely distributed in China.

An additional seven and three genomes were associated with arrivals from the UK and the US, respectively, in either late March or April. International travel was suspended following the arrival of these cases. Nine of these genomes belonged to the B1 lineage, while the last from a case associated with travel to the US belonged to the B2 lineage. Although in the same B1 lineage, eight of these travel-associated cases were on different sublineage and, with one exception, had no identical genomes from non-travel associated COVID-19 cases in our dataset. The one exception was sample ZW25BY that was from a case recorded in Bulawayo in April, which was identical to samples ZW-391 and ZW-EC, also from Bulawayo, with no travel-history recorded which, therefore, represents potential domestic transmission.

An increase in cases associated with Zimbabwean residents returning from neighbouring countries, especially South Africa, during May 2020 was reported by local surveillance. At this time SARS-CoV-2 was reported to be spreading rapidly in South Africa and therefore a two-week quarantine in designated accommodation was introduced. Although the majority of returnees were asymptomatic, they were tested and on multiple occasions were positive for

SARS-CoV-2. A total of 26 samples from quarantined individuals housed in six of these accommodations were sequenced. In several quarantine accommodation settings, genomes from multiple samples from different individuals resident in the same quarantine location were identical or differed by only a single nucleotide in the genome. For example in quarantine location A, two clusters were identified that consisted of 10 and 14 samples (May 2020), respectively all of which were in the B.1.1 lineage with genomes that were indistinguishable from one another; this was consistent with direct transmission. Since returnees from multiple locations were quarantined together, the most likely explanation is that transmission occurred in the quarantine accommodation rather than from a common outside source. Later spread events were primarily linked to returnees from South Africa and potential transmission in quarantine facilities.

Figure 4. Phylogenetic relationship amongst 100 SARS-Cov-2 samples from Zimbabwe. A maximum likelihood tree based on SNPs in the genome of 100 SARS-Cov-2 genomes from Zimbabwe from between March and June 2020. The tree is rooted with the original SARS-Cov-2 genome from Wuhan. Sample Identifier, district and province location from which the sample was collected, country of travel if applicable or domestic case, quarantine facility identity (A-E) or no quarantine (white), month of isolation, SARS-Cov-2 global lineage and residue 614 of the spike 'S' protein are indicated in columns.

Discussion:

Epidemiological analysis identified two phases of the SARS-CoV-2 epidemic in Zimbabwe. The vast majority of cases in the first 3 months were associated with international or regional travel. This was confirmed by the epidemiological data linking most confirmed cases with a history of having travelled outside of the country exacerbated by rapid influx of Zimbabwean citizens returning from other countries already experiencing the disease. A rapid increase in the number of cases was associated with probable domestic transmission after July 2020.

Zimbabwe is a landlocked country in Southern Africa between the Zambezi and Limpopo rivers and bordering South Africa, Botswana, Zambia and Mozambique. Travel history is essential for understanding the introduction of SARS-Cov-2 into a nation, and this was confirmed by the data from Zimbabwe. Initial cases recorded within the first 120 days of the outbreak were associated with travel from international locations already experiencing the disease. The majority of the initial COVID-19 cases in Zimbabwe (20th March – 30th June 2020) were Zimbabwean residents returning from neighbouring South Africa, Botswana and Mozambique. Most were associated with travel from South Africa related to informal trade between these two countries by cross-border traders, the majority of which are aged 25-45 years. The combination of informal cross-border traders and formally returning residents may have contributed to the subsequent increased local transmission in Zimbabwe. Other cases, however, were from Europe, Asia and USA, the epicentres of the SARS-Cov-2 epidemic. However, the rise in cases was slower than predicted and mirrored what has been seen in many African countries (Nkengasong and Mankoula, 2020). This slow rise in cases could be attributed to the measures the country instituted which included wearing of face masks, hand sanitization and lockdown. These measures have been shown to minimise and contain the outbreak in other countries, including France (Cauchemez *et al.*, 2020).

The SARS-Cov-2 pandemic is an emerging infectious disease with no known public health interventions except the measures implemented in China, Europe and elsewhere. Zimbabwe's response was shaped by experiences from other countries, where the epidemic

was described first (La Regina *et al.*, 2020; Shekhar *et al.*, 2020; Xu *et al.*, 2020). For the first 120 days following the first case of Covid-19 in Zimbabwe the total number of detected cases remained below 50, suggesting that public health interventions were effective in limiting local transmission to a minimum. The observed delayed local transmission may be as a result of the surveillance systems employed. Limiting movement by traders and cargo to a small manageable number, implementation of preventative equipment (e.g. hand sanitization, face mask wearing) and response measures within the first 30 days after the first case could explain this phenomenon. Though measures to minimise transmission were in place, a lot of irregularities emerged such as residents crossing borders illegally and/or non-compliance of quarantine requirements and relaxation of lockdown measures which could have contributed to an increase in local transmission. This was also evidenced by having the greatest number of cases centred around Bulawayo, Gweru and Harare cities with close proximity to land borders or international airports and densely populations. These factors may explain the propagated picture of the COVID-19 epidemic curve in Zimbabwe. Nonetheless, current evidence suggests that quarantine of travel-associated cases and early suspension of tourism-associated travel may have been effective based on our genomic analysis of cases in the first 120 days following the first case.

The factors associated with contraction of COVID-19 in Zimbabwe were consistent with factors associated with contraction in high income countries. The overall CFR of SARS Cov-2 was 2.9%, which was comparable with CFRs reported elsewhere in Asia, Europe and USA (Niforatos *et al.*, 2020; Pachetti *et al.*, 2020; Samaddar *et al.*, 2020). Compared with other countries, younger males aged 31- 40 years, were infected with SARS-CoV-2 in Zimbabwe during the first 120 days of the epidemic. However, the older population above 60 years was more likely to die from COVID-19 disease compared with the younger age groups, this agrees with what have been observed in other regions (Brandén *et al.*, 2020; Dhama *et al.*, 2020). Outbreak descriptions from Europe, Asia and the United States of America, indicated that older members of the population with pre-existing medical conditions were at increased risk of both becoming infected with SARS-Cov-2 and of dying from infection (Sanyaolu *et al.*, 2020; Schultze *et al.*, 2020). In the United Kingdom (UK) a considerably larger proportion of cases from ages 10-50 were in females, with cases in males being greatest in 0-10 years and 60-80 age categories (Page *et al.*, 2020). These observations may have been due to two demographic characteristics prevalent in Zimbabwe. First, the younger age-group, 31-50 years is the most mobile and some have migrated out of the country in search of employment (Potts, 2010; Crush *et al.*, 2015). Secondly, the same age group are sexually active and contribute the highest prevalence's of HIV/ AIDS infection in the Southern Africa region (Gouws *et al.*, 2008).

Fourteen out of the 100 SARS-CoV-2 cases sequenced reported that they had travelled from outside the Southern Africa region. Of these cases two resulted from local transmission with a positive epidemiological link. Amongst the remaining twelve cases there was no evidence of onward transmission based on the SARS-CoV-2 genomes sampled and sequenced in the subsequent two months. The sequences of SARS-CoV-2 genomes associated with travellers returning from neighbouring Southern African countries were distinct from those associated with travellers from the US, UK and Dubai recorded earlier in the epidemic; most were from lineage B.1.1, which was also the most common lineage amongst 2016 sequences from 18 African countries reported recently (www.afro.who.int). Interestingly, lineage B.1.1.1 which accounted for 16% of the samples reported in South Africa was not found in our analysis. This may be due to the fact that South Africa is a major airline hub in Southern Africa, connecting most of the countries to Europe, Asia and USA.

The D614G genotype was the most frequent variant in early spread events in Zimbabwe, consistent with the reported trend for increased frequency of this genotype in many locations globally, even when the ancestral genotype (614D) was previously established (Korber *et al.*, 2020). There has been considerable interest in the emergence of a variant lineage of SARS-CoV-2 with a non-synonymous mutation resulting in substitution of 614G in which a glycine residue was replaced by an aspartate residue at position 614 (D614G) of the spike (S) protein. The S-protein is thought to be important for entry of the virus into host cells and D614G has been implicated in increased cell entry (Korber *et al.*, 2020); the mechanism for this was via increasing the likelihood of binding the ACE2 receptor on the cell surface (Yurkovetskiy *et al.*, 2020). Consistent with this idea, population genetic modelling studies indicated that the 614G variant increased in frequency in the UK relative to the 614D ancestral variant, suggesting a selective advantage for 614G (Volz *et al.*, 2020). This protein is also the target in anti-viral (remdesivir and favipiravir) treatment trials for COVID-19 (Saxena, 2020). The mutation of SARS-CoV-2 is important because vaccine-escape strains may emerge quickly and there is also the potential for detection-failure if the target used in the RT-PCR tests for SARS-CoV-2 may accumulate mutations affecting the test. A total of 95 of the 100 SARS-CoV-2 genomes used in our analysis contained the D614G mutation. All five of the genomes with the ancestral 614D genotype were from samples from March or April 2020, with none observed in sequences from samples taken in May or June. The dominance of the 614G genotype in Zimbabwe likely reflects the relatively high frequency of the 614G genotype in countries from which the virus was introduced. However, since it has been suggested that the 614G genotype is present in greater numbers in the upper respiratory tract in infected people it may also have been more likely to spread once introduced into the Zimbabwean population.

A combination of local surveillance and whole genome sequencing revealed the early intercontinental and subsequent local international spread of the SARS-CoV-2 into Zimbabwe. The analysis provides insights into the impact of interventions on transmission and the basis for improved future responses to this and other pandemics. Our whole genome sequence analysis provides a baseline view of the SARS-CoV-2 introduced into Zimbabwe that will place later domestic spread of the virus within a Zimbabwean context. As the national response to the SARS-CoV-2 pandemic continues the capability to test, isolate and track the evolution of SARS-CoV-2, is crucial to assess ongoing mutation of the virus.

Conclusion:

Phylogenomic analysis highlights the dominance of regional migration over intercontinental travel in the spread of SARS-CoV-2 in Zimbabwe. The majority of SARS-CoV-2 genomes from lineages B.1 and B.1.1 were recovered from residents returning from South Africa. These lineages appear to be predominant in Africa and continuous efforts for widespread sequencing is important to provide insights about their origin, and the evolution of local/imported viral dissemination in Zimbabwe. As the epidemic unfolded most of the more than 8,000 cases were due to local transmission. Risk factors for local transmission have not been described. Continuous surveillance using whole-genome sequencing is required to estimate the presence of mutations that may be increasing transmissibility. In addition to describing the immunological response to SARS-CoV-2 infection, data from SARS-CoV-2 genomic surveillance may contribute to the development of vaccines against this novel virus. Zimbabwe will need to strengthen health systems so that public health interventions that are effective against SARS-CoV-2 transmission elsewhere in the world, are also effective in Zimbabwe.

Limitations

Samples used for whole genome sequencing were sampled from the first 120 days of the epidemic. Potential subsequent circulation of descendants of these early intercontinental introductions after around 120 days once transmission was increasing exponentially, remains to be determined.

Ethical approval:

Ethical approval for the study was exempted by the Medical Research Council of Zimbabwe as this study was in response to a public health emergency.

Acknowledgements:

The authors thank all local clinical and laboratory staff for their contribution and dedication to the SARS-CoV-2 Surveillance Network. Special thanks go to Elizabeth Tabitha Abbew, for all her expert advice and proofreading of the article. The authors acknowledge the Epidemiology and Diseases Control Department of the Ministry of Health and Child Care who provided full access to all the COVID-19 datasets exploited in this study. We thank members of the COVID-19 Genomics Consortium UK for their contributions to generating some of the data used in this study. We thank the laboratories who submitted their data to GISAID, full details are listed in the Supplementary Material.

Funding statements:

This work was supported by a combination of routine work of scientists at the National Microbiology Reference Laboratory, Biomedical Research and Training Institute, Beatrice Road Infectious Diseases Hospital, National Virology Laboratory and National TB Reference Laboratory. The authors gratefully acknowledge the support of the Biotechnology and Biological Sciences Research Council (BBSRC); this research was funded in part by the BBSRC Institute Strategic Programme Microbes in the Food Chain BB/R012504/1 and its constituent projects BBS/E/F/000PR10348, BBS/E/F/000PR10349, BBS/E/F/000PR10351, and BBS/E/F/000PR10352 and by the Quadram Institute Bioscience BBSRC funded Core Capability Grant (project number BB/CCG1860/1). The COVID-19 Genomics UK (COG-UK) Consortium 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.

Author contributions:

Leadership and supervision: AJP, RAK, JOG, SMZ

Funding acquisition: AJP, RAK, JOG

Metadata curation: TM, GT, SMZ, GM, MGM

Project administration: TM, RAK, AJP, GM

Samples and logistics: TM, GT, AVG, TT, MGM, TA, HG, SARS-CoV-2 Research Group, Zimbabwe

Sequencing and bioinformatics analysis: DJB, AJP, LOM

Data analysis: TM, TT, LM, AP, RAK, IP, PM, AG, SD, IC, JJ, RM, KM, IM, DM, MH

Paper writing: TM, RAK, LM, JC, BVC, FTK, FTT, JC, MGM, GM, SMZ

Appendix 1 (The COVID-19 Genomics UK (COG-UK) Consortium)

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 Loman¹⁵.

Leadership, supervision, sequencing, analysis, funding acquisition, metadata curation, project administration, samples, logistics, and visualisation:

Samuel C Robson⁶⁸.

Leadership, supervision, project administration, visualisation, samples, logistics, metadata curation and software and analysis tools:

Tanya Golubchik²⁷.

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 Bonsall²⁷.

Leadership and supervision, sequencing, analysis, funding acquisition, visualisation and software and analysis tools:

Ali R Awan⁷⁴.

Leadership and supervision, funding acquisition, sequencing, analysis, metadata curation, samples and logistics:

Sally Corden³³.

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 Bashton⁶⁰, and Nicole Pacchiarini³³.

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 Silva³⁸, and Dennis Wang³⁸.

Project administration, samples, logistics, leadership and supervision:

Monique Andersson²⁸, Anoop J Chauhan⁷⁰, Mariateresa de Cesare²⁶, Catherine Ludden^{1,3}, and Tabitha W Mahungu⁹¹.

Sequencing, analysis, project administration and metadata curation:

Rebecca Dewar²⁰, and Martin P McHugh²⁰.

Samples, logistics, metadata curation and project administration:

Natasha G Jesudason²¹, Kathy K Li MBCh²¹, Rajiv N Shah²¹, and Yusri Taha⁶⁶.

Leadership, supervision, funding acquisition and metadata curation:

Kate E Templeton²⁰.

Leadership, supervision, funding acquisition, sequencing and analysis:

Simon Cottrell³³, Justin O'Grady⁵¹, Andrew Rambaut¹⁹, and Colin P Smith⁹³.

Leadership, supervision, metadata curation, sequencing and analysis:

Matthew T.G. Holden⁸⁷, and Emma C Thomson²¹.

Leadership, supervision, samples, logistics and metadata curation:

Samuel Moses^{81, 82}.

Sequencing, analysis, leadership, supervision, samples and logistics:

Meera Chand⁷, Chrystala Constantinidou⁷¹, Alistair C Darby⁴⁶, Julian A Hiscox⁴⁶, Steve Paterson⁴⁶, and Meera Unnikrishnan⁷¹.

Sequencing, analysis, leadership and supervision and software and analysis tools:

Andrew J Page⁵¹, and Erik M Volz⁹⁶.

Samples, logistics, sequencing, analysis and metadata curation:

Charlotte J Houldcroft⁸, Aminu S Jahun¹¹, James P McKenna⁸⁸, Luke W Meredith¹¹, Andrew Nelson⁶¹, Sarojini Pandey⁷², and Gregory R Young⁶⁰.

Sequencing, analysis, metadata curation, and software and analysis tools:

Anna Price³⁴, Sara Rey³³, Sunando Roy⁴¹, Ben Temperton⁴⁹, and Matthew Wyles³⁸.

Sequencing, analysis, metadata curation and visualisation:

Stefan Rooke¹⁹, and Sharif Shaaban⁸⁷.

Visualisation, sequencing, analysis and software and analysis tools:

Helen Adams³⁵, Yann Bourgeois⁶⁹, Katie F Loveson⁶⁸, Áine O'Toole¹⁹, and Richard Stark⁷¹.

Project administration, leadership and supervision:

Ewan M Harrison^{1,3}, David Heyburn³³, and Sharon J Peacock^{2,3}

Project administration and funding acquisition:

David Buck²⁶, and Michaela John³⁶

Sequencing, analysis and project administration:

Dorota Jamrozy¹, and Joshua Quick¹⁵

Samples, logistics, and project administration:

Rahul Batra⁷⁸, Katherine L Bellis^{1,3}, Beth Blane³, Sophia T Girgis³, Angie Green²⁶, Anita Justice²⁸, Mark Kristiansen⁴¹, and Rachel J Williams⁴¹.

Project administration, software and analysis tools:

Radoslaw Poplawski¹⁵.

Project administration and visualisation:

Garry P Scarlett⁶⁹.

Leadership, supervision, and funding acquisition:

John A Todd²⁶, Christophe Fraser²⁷, Judith Breuer^{40,41}, Sergi Castellano⁴¹, Stephen L Michell⁴⁹, Dimitris Gramatopoulos⁷³, and Jonathan Edgeworth⁷⁸.

Leadership, supervision and metadata curation:

Gemma L Kay⁵¹.

Leadership, supervision, sequencing and analysis:

Ana da Silva Filipe²¹, Aaron R Jeffries⁴⁹, Sascha Ott⁷¹, Oliver Pybus²⁴, David L Robertson²¹, David A Simpson⁶, and Chris Williams³³.

Samples, logistics, leadership and supervision:

Cressida Auckland⁵⁰, John Boyes⁸³, Samir Dervisevic⁵², Sian Ellard^{49, 50}, Sonia Goncalves¹, Emma J Meader⁵¹, Peter Muir², Husam Osman⁹⁵, Reenesh Prakash⁵², Venkat Sivaprakasam¹⁸, and Ian B Vipond².

Leadership, supervision and visualisation

Jane AH Masoli^{49, 50}.

Sequencing, analysis and metadata curation

Nabil-Fareed Alikhan⁵¹, Matthew Carlile⁵⁴, Noel Craine³³, Sam T Haldenby⁴⁶, Nadine Holmes⁵⁴, Ronan A Lyons³⁷, Christopher Moore⁵⁴, Malorie Perry³³, Ben Warne⁸⁰, and Thomas Williams¹⁹.

Samples, logistics and metadata curation:

Lisa Berry⁷², Andrew Bosworth⁹⁵, Julianne Rose Brown⁴⁰, Sharon Campbell⁶⁷, Anna Casey¹⁷, Gemma Clark⁵⁶, Jennifer Collins⁶⁶, Alison Cox^{43, 44}, Thomas Davis⁸⁴, Gary Eltringham⁶⁶, Cariad Evans^{38, 39}, Clive Graham⁶⁴, Fenella Halstead¹⁸, Kathryn Ann Harris⁴⁰, Christopher Holmes⁵⁸, Stephanie Hutchings², Miren Iturriza-Gomara⁴⁶, Kate Johnson^{38, 39}, Katie Jones⁷², Alexander J Keeley³⁸, Bridget A Knight^{49, 50}, Cherian Koshy⁹⁰, Steven Liggett⁶³, Hannah Lowe⁸¹, Anita O Lucaci⁴⁶, Jessica Lynch^{25, 29}, Patrick C McClure⁵⁵, Nathan Moore³¹, Matilde Mori^{25, 29, 32}, David G Partridge^{38, 39}, Pinglawathee Madona^{43, 44}, Hannah M Pymont², Paul Anthony Randell^{43, 44}, Mohammad Raza^{38, 39}, Felicity Ryan⁸¹, Robert Shaw²⁸, Tim J Sloan⁵⁷, and Emma Swindells⁶⁵.

Sequencing, analysis, Samples and logistics:

Alexander Adams³³, Hibo Asad³³, Alec Birchley³³, Tony Thomas Brooks⁴¹, Giselda Bucca⁹³, Ethan Butcher⁷⁰, Sarah L Caddy¹³, Laura G Caller^{2, 3, 12}, Yasmin Chaudhry¹¹, Jason Coombes³³, Michelle Cronin³³, Patricia L Dyal⁴¹, Johnathan M Evans³³, Laia Fina³³, Bree Gatica-Wilcox³³, Iliana Georgana¹¹, Lauren Gilbert³³, Lee Graham³³, Danielle C Groves³⁸, Grant Hall¹¹, Ember Hilvers³³, Myra Hosmillo¹¹, Hannah Jones³³, Sophie Jones³³, Fahad A Khokhar¹³, Sara Kumziene-Summerhayes³³, George MacIntyre-Cockett²⁶, Rocio T Martinez Nunez⁹⁴, Caoimhe McKerr³³, Claire McMurray¹⁵, Richard Myers⁷, Yasmin Nicole Panchbhaya⁴¹, Malte L Pinckert¹¹, Amy Plimmer³³, Joanne Stockton¹⁵, Sarah Taylor³³, Alicia Thornton⁷, Amy Trebes²⁶, Alexander J Trotter⁵¹, Helena Jane Tutill⁴¹, Charlotte A Williams⁴¹, Anna Yakovleva¹¹ and Wen C Yew⁶².

Sequencing, analysis and software and analysis tools:

Mohammad T Alam⁷¹, Laura Baxter⁷¹, Olivia Boyd⁹⁶, Fabricia F. Nascimento⁹⁶, Timothy M Freeman³⁸, Lily Geidelberg⁹⁶, Joseph Hughes²¹, David Jorgensen⁹⁶, Benjamin B Lindsey³⁸, Richard J Orton²¹, Manon Ragonnet-Cronin⁹⁶, Joel Southgate^{33, 34}, and Sreenu Vattipally²¹.

Samples, logistics and software and analysis tools:

Igor Starinskij²³.

Visualisation and software and analysis tools:

Joshua B Singer²¹, Khalil Abudahab^{1, 30}, Leonardo de Oliveira Martins⁵¹, Thanh Le-Viet⁵¹, Mirko Menegazzo³⁰, Ben EW Taylor^{1, 30}, and Corin A Yeats³⁰.

Project Administration:

Sophie Palmer³, Carol M Churcher³, Alisha Davies³³, Elen De Lacy³³, Fatima Downing³³, Sue Edwards³³, Nikki Smith³⁸, Francesc Coll⁹⁷, Nazreen F Hadjirin³ and Frances Bolt^{44, 45}.

Leadership and supervision:

Alex Alderton¹, Matt Berriman¹, Ian G Charles⁵¹, Nicholas Cortes³¹, Tanya Curran⁸⁸, John Danesh¹, Sahar Eldirdiri⁸⁴, Ngozi Elumogo⁵², Andrew Hattersley^{49, 50}, Alison Holmes^{44, 45}, Robin Howe³³, Rachel Jones³³, Anita Kenyon⁸⁴, Robert A Kingsley⁵¹, Dominic Kwiatkowski^{1, 9}, Cordelia Langford¹, Jenifer Mason⁴⁸, Alison E Mather⁵¹, Lizzie Meadows⁵¹, Sian Morgan³⁶, James Price^{44, 45}, Trevor I Robinson⁴⁸, Giri Shankar³³, John Wain⁵¹, and Mark A Webber⁵¹.

Metadata curation:

Declan T Bradley^{5, 6}, Michael R Chapman^{1, 3, 4}, Derrick Crooke²⁸, David Eyre²⁸, Martyn Guest³⁴, Huw Gulliver³⁴, Sarah Hoosdally²⁸, Christine Kitchen³⁴, Ian Merrick³⁴, Siddharth Mookerjee^{44, 45}, Robert Munn³⁴, Timothy Peto²⁸, Will Potter⁵², Dheeraj K Sethi⁵², Wendy Smith⁵⁶, Luke B Snell^{75, 94}, Rachael Stanley⁵², Claire Stuart⁵² and Elizabeth Wastenge²⁰.

Sequencing and analysis:

Erwan Acheson⁶, Safiah Afifi³⁶, Elias Allara^{2, 3}, Roberto Amato¹, Adrienn Angyal³⁸, Elihu Aranday-Cortes²¹, Cristina Ariani¹, Jordan Ashworth¹⁹, Stephen Attwood²⁴, Alp Aydin⁵¹, David J Baker⁵¹, Carlos E Balcazar¹⁹, Angela Beckett⁶⁸, Robert Beer³⁶, Gilberto

Betancor⁷⁶, Emma Betteridge¹, David Bibby⁷, Daniel Bradshaw⁷, Catherine Bresner³⁴, Hannah E Bridgewater⁷¹, Alice Broos²¹, Rebecca Brown³⁸, Paul E Brown⁷¹, Kirstyn Brunker²², Stephen N Carmichael²¹, Jeffrey K. J. Cheng⁷¹, Dr Rachel Colquhoun¹⁹, Gavin Dabrera⁷, Johnny Debebe⁵⁴, Eleanor Drury¹, Louis du Plessis²⁴, Richard Eccles⁴⁶, Nicholas Ellaby⁷, Audrey Farbos⁴⁹, Ben Farr¹, Jacqueline Findlay⁴¹, Chloe L Fisher⁷⁴, Leysa Marie Forrest⁴¹, Sarah Francois²⁴, Lucy R. Frost⁷¹, William Fuller³⁴, Eileen Gallagher⁷, Michael D Gallagher¹⁹, Matthew Gemmell⁴⁶, Rachel AJ Gilroy⁵¹, Scott Goodwin¹, Luke R Green³⁸, Richard Gregory⁴⁶, Natalie Groves⁷, James W Harrison⁴⁹, Hassan Hartman⁷, Andrew R Hesketh⁹³, Verity Hill¹⁹, Jonathan Hubb⁷, Margaret Hughes⁴⁶, David K Jackson¹, Ben Jackson¹⁹, Keith James¹, Natasha Johnson²¹, Ian Johnston¹, Jon-Paul Keatley¹, Moritz Kraemer²⁴, Angie Lackenby⁷, Mara Lawniczak¹, David Lee⁷, Rich Livett¹, Stephanie Lo¹, Daniel Mair²¹, Joshua Maksimovic³⁶, Nikos Manesis⁷, Robin Manley⁴⁹, Carmen Manso⁷, Angela Marchbank³⁴, Inigo Martincorena¹, Tamyo Mbisa⁷, Kathryn McCluggage³⁶, JT McCrone¹⁹, Shahjahan Miah⁷, Michelle L Michelsen⁴⁹, Mari Morgan³³, Gaia Nebbia⁷⁸, Charlotte Nelson⁴⁶, Jenna Nichols²¹, Paola Niola⁴¹, Kyriaki Nomikou²¹, Steve Palmer¹, Naomi Park¹, Yasmin A Parr¹, Paul J Parsons³⁸, Vineet Patel⁷, Minal Patel¹, Clare Pearson^{2, 1}, Steven Platt⁷, Christoph Puethe¹, Mike Quail¹, Jayna Raghwan²⁴, Lucille Rainbow⁴⁶, Shavanthi Rajatileka¹, Mary Ramsay⁷, Paola C Resende Silva^{41, 42}, Steven Rudder⁵¹, Chris Ruis³, Christine M Sambles⁴⁹, Fei Sang⁵⁴, Ulf Schaefer⁷, Emily Scher¹⁹, Carol Scott¹, Lesley Shirley¹, Adrian W Signell⁷⁶, John Sillitoe¹, Christen Smith¹, Dr Katherine L Smollett²¹, Karla Spellman³⁶, Thomas D Stanton¹⁹, David J Studholme⁴⁹, Grace Taylor-Joyce⁷¹, Ana P Tedim⁵¹, Thomas Thompson⁶, Nicholas M Thomson⁵¹, Scott Thurston¹, Lily Tong²¹, Gerry Tonkin-Hill¹, Rachel M Tucker³⁸, Edith E Vamos⁴, Tetyana Vasylyeva²⁴, Joanna Warwick-Dugdale⁴⁹, Danni Weldon¹, Mark Whitehead⁴⁶, David Williams⁷, Kathleen A Williamson¹⁹, Harry D Wilson⁷⁶, Trudy Workman³⁴, Muhammad Yasir⁵¹, Xiaoyu Yu¹⁹, and Alex Zarebski²⁴.

Samples and logistics:

Evelien M Adriaenssens⁵¹, Shazaad S Y Ahmad^{2, 47}, Adela Alcolea-Medina^{59, 77}, John Allan⁶⁰, Patawee Asamaphan²¹, Laura Atkinson⁴⁰, Paul Baker⁶³, Jonathan Ball⁵⁵, Edward Barton⁶⁴, Mathew A Beale¹, Charlotte Beaver¹, Andrew Beggs¹⁶, Andrew Bell⁵¹, Duncan J Berger¹, Louise Berry⁵⁶, Claire M Bewshea⁴⁹, Kelly Bicknell⁷⁰, Paul Bird⁵⁸, Chloe Bishop⁷, Tim Boswell⁵⁶, Cassie Breen⁴⁸, Sarah K Buddenborg¹, Shirelle Burton-Fanning⁶⁶, Vicki Chalker⁷, Joseph G Chappell⁵⁵, Themoula Charalampous^{78, 94}, Claire Cormie³, Nick Cortes^{29, 25}, Lindsay J Coupland⁵², Angela Cowell⁴⁸, Rose K Davidson⁵³, Joana Dias³, Maria Diaz⁵¹, Thomas Dibling¹, Matthew J Dorman¹, Nichola Duckworth⁵⁷, Scott Elliott⁷⁰, Sarah Essex⁶³, Karlie Fallon⁵⁸, Theresa Feltwell⁸, Vicki M Fleming⁵⁶, Sally Forrest³, Luke

Foulser¹, Maria V Garcia-Casado¹, Artemis Gavriil⁴¹, Ryan P George⁴⁷, Laura Gifford³³, Harmeet K Gill³, Jane Greenaway⁶⁵, Luke Griffith⁵³, Ana Victoria Gutierrez⁵¹, Antony D Hale⁸⁵, Tanzina Haque⁹¹, Katherine L Harper⁸⁵, Ian Harrison⁷, Judith Heaney⁸⁹, Thomas Helmer⁵⁸, Ellen E Higginson³, Richard Hopes², Hannah C Howson-Wells⁵⁶, Adam D Hunter¹, Robert Impey⁷⁰, Dianne Irish-Tavares⁹¹, David A Jackson¹, Kathryn A Jackson⁴⁶, Amelia Joseph⁵⁶, Leanne Kane¹, Sally Kay¹, Leanne M Kermack³, Manjinder Khakh⁵⁶, Stephen P Kidd^{29, 25, 31}, Anastasia Kolyva⁵¹, Jack CD Lee⁴⁰, Laura Letchford¹, Nick Levene⁷⁹, Lisa J Levett⁸⁹, Michelle M Lister⁵⁶, Allyson Lloyd⁷⁰, Joshua Loh⁶⁰, Louissa R Macfarlane-Smith⁸⁵, Nicholas W Machin^{2, 47}, Mailis Maes³, Samantha McGuigan¹, Liz McMinn¹, Lamia Mestek-Boukhibar⁴¹, Zoltan Molnar⁶, Lynn Monaghan⁷⁹, Catrin Moore²⁷, Plamena Naydenova³, Alexandra S Neaverson¹, Rachel Nelson¹, Marc O Niebel²¹, Elaine O'Toole⁴⁸, Debra Padgett⁶⁴, Gaurang Patel¹, Brendan Al Payne⁶⁶, Liam Prestwood¹, Veena Raviprakash⁶⁷, Nicola Reynolds⁸⁶, Alex Richter¹⁶, Esther Robinson⁹⁵, Hazel A Rogers¹, Aileen Rowan⁹⁶, Garren Scott⁶⁴, Divya Shah⁴⁰, Nicola Sheriff⁶⁷, Graciela Sluga, Emily Souster¹, Michael Spencer-Chapman¹, Sushmita Sridhar^{1, 3}, Tracey Swingler⁵³, Julian Tang⁵⁸, Graham P Taylor⁹⁶, Theocharis Tsoleridis⁵⁵, Lance Turtle⁴⁶, Sarah Walsh⁵⁷, Michelle Wantoch⁸⁶, Joanne Watts⁴⁸, Sheila Waugh⁶⁶, Sam Weeks⁴¹, Rebecca Williams³¹, Iona Willingham⁵⁶, Emma L Wise^{25, 29, 31}, Victoria Wright⁵⁴, Sarah Wyllie⁷⁰, and Jamie Young³.

Software and analysis tools

Amy Gaskin³³, Will Rowe¹⁵, and Igor Siveroni⁹⁶.

Visualisation:

Robert Johnson⁹⁶.

Affiliations:

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.

Appendix 2 (SARS-CoV-2 Research Group, Zimbabwe)

Samples and logistics:

Kenneth Maeka¹, Peter Gumbo¹, Amini David¹, Tariro Madamombe¹, Lucy Sisya¹, Shukusho D Fungai¹, Mercy Marumure¹, Admire Bwagirire¹, Florence Malunga¹, Norah Sukutayi-Vere¹, Takudzwa Kawome¹, Stanford Mupandasekwa¹, Tanaka Sakubani¹, Exavier Mazarura¹, Babra Murwira², Monalisa Mutimutema², Joshua Mbanga³, Zephania Dhlamini³, Aleck Maunganidze³, Anita Dube³, Beauty Makamure⁴, Justin Mayini⁴, Tendai Washaya⁴, Janice Martins⁴, Isabel Mashita⁴, Happiness Mahove⁴, Norah Karonga⁴, Forget Makoga⁴, Learnmore Chawira⁵, Immaculate Govore⁵, Chipso Berejena⁶, Arnold Mukaratirwa⁶, Ruhanya Vurayayi⁶, John Macharaga⁷, Samson Chomunacho⁷, Simbarashe Mapurisa⁷, Albert Mudungwe⁷, Davison Makwarimba⁷, Raymond Ncube⁷, Beverley Chinembiri⁷, Godfrey Mhondiwa⁷, Nelia Kativu⁷, Rutendo Muchemwa⁷ and Everlyn Choto⁷.

Paper writing:

Ruhanya Vurayayi⁶.

Affiliations:

1 National Microbiology Reference Laboratory, Ministry of Health and Child Care, Zimbabwe , P.O. Box ST 749, Southerton, Harare, Zimbabwe, **2** National TB Reference Laboratory, Bulawayo, **3** National University of Science and Technology, Bulawayo, **4** Biomedical

Research and Training Institute, **5** Harare City Health Department, Beatrice Road Infectious Diseases Hospital, Harare, Zimbabwe, **6** National Virology Laboratory, Harare, Zimbabwe, **7** Sally Mugabe Central Hospital, Harare, Zimbabwe

References:

- Baker, D. J., Kay, G. L., Aydin, A., Le-Viet, T., Rudder, S., Tedim, A. P., Kolyva, A., Diaz, M., De Oliveira Martins, L., Alikhan, N.-F., Meadows, L., Bell, A., Gutierrez, A. V., Trotter, A. J., Thomson, N. M., Gilroy, R., Griffith, L., Adriaenssens, E. M., Stanley, R., Charles, I. G., Elumogo, N., Wain, J., Prakash, R., Meader, E., Mather, A. E., Webber, M. A., Dervisevic, S., Page, A. J. & O'grady, J. 2020. CoronaHiT: large scale multiplexing of SARS-CoV-2 genomes using Nanopore sequencing. *BioRxiv*, 10.1101/2020.06.24.162156: 2020.2006.2024.162156.
- Brandén, M., Aradhya, S., Kolk, M., Härkönen, J., Drefahl, S., Malmberg, B., Rostila, M., Cederström, A., Andersson, G. & Mussino, E. 2020. Residential context and COVID-19 mortality among adults aged 70 years and older in Stockholm: a population-based, observational study using individual-level data. *The Lancet Healthy Longevity*, 1: e80-e88.
- Cauchemez, S., Kiem, C. T., Paireau, J., Rolland, P. & Fontanet, A. 2020. Lockdown impact on COVID-19 epidemics in regions across metropolitan France. *The Lancet*, 396: 1068-1069.
- Connor, T. R., Loman, N. J., Thompson, S., Smith, A., Southgate, J., Poplawski, R., Bull, M. J., Richardson, E., Ismail, M. & Elwood-Thompson, S. 2016. CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community. *Microb Genom*, 2: e000086.
- Corman, V. M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D. K., Bleicker, T., Brünink, S., Schneider, J. & Schmidt, M. L. 2020. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance*, 25: 2000045.
- Crush, J., Chikanda, A. & Tawodzera, G. 2015. The third wave: mixed migration from Zimbabwe to South Africa. *Canadian Journal of African Studies/Revue canadienne des études africaines*, 49: 363-382.

- Dhama, K., Patel, S. K., Kumar, R., Rana, J., Yatoo, M. I., Kumar, A., Tiwari, R., Dhama, J., Natesan, S., Singh, R. & Harapan, H. 2020. Geriatric Population During the COVID-19 Pandemic: Problems, Considerations, Exigencies, and Beyond. *Frontiers in public health*, 8: 574198.
- Duchene, S., Featherstone, L., Haritopoulou-Sinanidou, M., Rambaut, A., Lemey, P. & Baele, G. 2020. Temporal signal and the phylodynamic threshold of SARS-CoV-2. *BioRxiv*, 10.1101/2020.05.04.077735: 2020.2005.2004.077735.
- Gouws, E., Stanecki, K. A., Lyerla, R. & Ghys, P. D. 2008. The epidemiology of HIV infection among young people aged 15–24 years in southern Africa. *Aids*, 22: S5-S16.
- Grubaugh, N. D., Gangavarapu, K., Quick, J., Matteson, N. L., De Jesus, J. G., Main, B. J., Tan, A. L., Paul, L. M., Brackney, D. E. & Grewal, S. 2019. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome biology*, 20: 1-19.
- Hâncean, M.-G., Perc, M. & Lerner, J. 2020. Early spread of COVID-19 in Romania: imported cases from Italy and human-to-human transmission networks. *Royal Society open science*, 7: 200780.
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J. & Gu, X. 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, 395: 497-506.
- Ishikawa, S. A., Zhukova, A., Iwasaki, W. & Gascuel, O. 2019. A Fast Likelihood Method to Reconstruct and Visualize Ancestral Scenarios. *Molecular Biology and Evolution*, 36: 2069-2085.
- Kalk, A. & Schultz, A. 2020. SARS-CoV-2 epidemic in African countries—are we losing perspective? *The Lancet infectious diseases*, 20: 1370.
- Kaseje, N. Why Sub-Saharan Africa needs a unique response to COVID-19. Published by World Economic Forum on, 2020.

- Katoh, K. & Standley, D. M. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, 30: 772-780.
- Khandelwal, S. K. 2020. Debating the process, impact, and handling of social and health determinants of the COVID-19 pandemic. *Indian Journal of Social Psychiatry*, 36: 64.
- Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E. E., Bhattacharya, T., Foley, B., Hastie, K. M., Parker, M. D., Partridge, D. G., Evans, C. M., Freeman, T. M., De Silva, T. I., Angyal, A., Brown, R. L., Carrilero, L., Green, L. R., Groves, D. C., Johnson, K. J., Keeley, A. J., Lindsey, B. B., Parsons, P. J., Raza, M., Rowland-Jones, S., Smith, N., Tucker, R. M., Wang, D., Wyles, M. D., Mcdanal, C., Perez, L. G., Tang, H., Moon-Walker, A., Whelan, S. P., Labranche, C. C., Saphire, E. O. & Montefiori, D. C. 2020. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*, 182: 812-827.e819.
- La Regina, M., Tanzini, M., Fineschi, V., Venneri, F., Toccafondi, G., Lachman, P., Tartaglia, R. & Group, C.-I. W. 2020. Responding to COVID-19: the experience from Italy and recommendations for management and prevention. *International Journal for Quality in Health Care*, 10.1093/intqhc/mzaa057.
- Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.
- Lone, S. A. & Ahmad, A. 2020. COVID-19 pandemic—An African perspective. *Emerging microbes & infections*: 1-28.
- Mehtar, S., Preiser, W., Lakhe, N. A., Bousso, A., Tamfum, J.-J. M., Kallay, O., Seydi, M., Zumla, A. & Nachega, J. B. 2020. Limiting the spread of COVID-19 in Africa: one size mitigation strategies do not fit all countries. *The Lancet. Global health*, 8: e881-e883.
- Meredith, L. W., Hamilton, W. L., Warne, B., Houldcroft, C. J., Hosmillo, M., Jahun, A. S., Curran, M. D., Parmar, S., Caller, L. G. & Caddy, S. L. 2020. Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-

- 19: a prospective genomic surveillance study. *The Lancet infectious diseases*, 20: 1263-1272.
- Minh, B. Q., Schmidt, H., Chernomor, O., Schrempf, D., Woodhams, M., Von Haeseler, A. & Lanfear, R. 2019. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *BioRxiv*, 10.1101/849372: 849372.
- Msomi, N., Mlisana, K., De Oliveira, T., Msomi, N., Mlisana, K., Williamson, C., Bhiman, J. N., Goedhals, D., Engelbrecht, S., Van Zyl, G., Preiser, W., Hardie, D., Hsiao, M., Mulder, N., Martin, D., Christoffels, A., York, D., Giandhari, J., Wilkinson, E., Pillay, S., Tegally, H., James, S. E., Kanzi, A., Lessells, R. J. & De Oliveira, T. 2020. A genomics network established to respond rapidly to public health threats in South Africa. *The Lancet Microbe*, 1: e229-e230.
- Niforatos, J. D., Melnick, E. R. & Faust, J. S. 2020. Covid-19 fatality is likely overestimated. *Bmj*, 368: m1113.
- Nkengasong, J. N. & Mankoula, W. 2020. Looming threat of COVID-19 infection in Africa: act collectively, and fast. *The Lancet*, 395: 841-842.
- Pachetti, M., Marini, B., Giudici, F., Benedetti, F., Angeletti, S., Ciccozzi, M., Masciovecchio, C., Ippodrino, R. & Zella, D. 2020. Impact of lockdown on Covid-19 case fatality rate and viral mutations spread in 7 countries in Europe and North America. *Journal of Translational Medicine*, 18: 338.
- Plante, J. A., Liu, Y., Liu, J., Xia, H., Johnson, B. A., Lokugamage, K. G., Zhang, X., Muruato, A. E., Zou, J., Fontes-Garfias, C. R., Mirchandani, D., Scharon, D., Bilello, J. P., Ku, Z., An, Z., Kalveram, B., Freiberg, A. N., Menachery, V. D., Xie, X., Plante, K. S., Weaver, S. C. & Shi, P.-Y. 2020. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature*, 10.1038/s41586-020-2895-3.
- Potts, D. H. 2010. *Circular migration in Zimbabwe & contemporary sub-Saharan Africa*, Boydell & Brewer.
- Price, M. N., Dehal, P. S. & Arkin, A. P. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PloS one*, 5: e9490.

- Quick, J. 2020. NCoV-2019 Sequencing Protocol V2. April. <https://doi.org/10.17504/protocols.io.bdp7i5rn>.
- Rambaut, A., Holmes, E. C., Hill, V., O'toole, Á., Mccrone, J. T., Ruis, C., Du Plessis, L. & Pybus, O. G. 2020. A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology. *BioRxiv*, 10.1101/2020.04.17.046086: 2020.2004.2017.046086.
- Samaddar, A., Gadepalli, R., Nag, V. L. & Misra, S. 2020. The Enigma of Low COVID-19 Fatality Rate in India. *Frontiers in Genetics*, 11.
- Sanyaolu, A., Okorie, C., Marinkovic, A., Patidar, R., Younis, K., Desai, P., Hosein, Z., Padda, I., Mangat, J. & Altaf, M. 2020. Comorbidity and its Impact on Patients with COVID-19. *SN comprehensive clinical medicine*, 10.1007/s42399-020-00363-4: 1-8.
- Saxena, A. 2020. Drug targets for COVID-19 therapeutics: Ongoing global efforts. *Journal of biosciences*, 45: 87.
- Schultze, A., Walker, A. J., Mackenna, B., Morton, C. E., Bhaskaran, K., Brown, J. P., Rentsch, C. T., Williamson, E., Drysdale, H., Croker, R., Bacon, S., Hulme, W., Bates, C., Curtis, H. J., Mehrkar, A., Evans, D., Inglesby, P., Cockburn, J., Mcdonald, H. I., Tomlinson, L., Mathur, R., Wing, K., Wong, A. Y. S., Forbes, H., Parry, J., Hester, F., Harper, S., Evans, S. J. W., Quint, J., Smeeth, L., Douglas, I. J. & Goldacre, B. 2020. Risk of COVID-19-related death among patients with chronic obstructive pulmonary disease or asthma prescribed inhaled corticosteroids: an observational cohort study using the OpenSAFELY platform. *The Lancet Respiratory Medicine*, 8: 1106-1120.
- Shekhar, R., Sheikh, A. B., Upadhyay, S., Atencio, J. & Kapuria, D. 2020. Early experience with COVID-19 patients at academic hospital in Southwestern United States. *Infectious Diseases*, 52: 596-599.
- Shu, Y. & Mccauley, J. 2017. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance*, 22: 30494.

- Volz, E. M., Hill, V., Mccrone, J. T., Price, A., Jorgensen, D., Toole, A., Southgate, J. A., Johnson, R., Jackson, B., Nascimento, F. F., Rey, S. M., Nicholls, S. M., Colquhoun, R. M., Da Silva Filipe, A., Shepherd, J. G., Pascall, D. J., Shah, R., Jesudason, N., Li, K., Jarrett, R., Pacchiarini, N., Bull, M., Geidelberg, L., Siveroni, I., Goodfellow, I. G., Loman, N. J., Pybus, O., Robertson, D. L., Thomson, E. C., Rambaut, A. & Connor, T. R. 2020. Evaluating the effects of SARS-CoV-2 Spike mutation D614G on transmissibility and pathogenicity. *medRxiv*, 10.1101/2020.07.31.20166082: 2020.2007.2031.20166082.
- Walker, P. G. T., Whittaker, C., Watson, O. J., Baguelin, M., Winskill, P., Hamlet, A., Djafaara, B. A., Cucunubá, Z., Olivera Mesa, D., Green, W., Thompson, H., Nayagam, S., Ainslie, K. E. C., Bhatia, S., Bhatt, S., Boonyasiri, A., Boyd, O., Brazeau, N. F., Cattarino, L., Cuomo-Dannenburg, G., Dighe, A., Donnelly, C. A., Dorigatti, I., Van Elsland, S. L., Fitzjohn, R., Fu, H., Gaythorpe, K. a. M., Geidelberg, L., Grassly, N., Haw, D., Hayes, S., Hinsley, W., Imai, N., Jorgensen, D., Knock, E., Laydon, D., Mishra, S., Nedjati-Gilani, G., Okell, L. C., Unwin, H. J., Verity, R., Vollmer, M., Walters, C. E., Wang, H., Wang, Y., Xi, X., Lalloo, D. G., Ferguson, N. M. & Ghani, A. C. 2020. The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. *Science*, 369: 413.
- Xu, W., Wu, J. & Cao, L. 2020. COVID-19 pandemic in China: Context, experience and lessons. *Health Policy and Technology*, <https://doi.org/10.1016/j.hlpt.2020.08.006>.