medRxiv preprint doi: https://doi.org/10.1101/2022.02.17.22269742; this version posted February 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Genomically informed clinical comparison of three epidemic waves of COVID-19 in Malawi

Authors: Catherine Anscombe*[1, 2], Samantha Lissauer*[&1,4], Herbert Thole*[1], Jamie Rylance[1,2], Dingase Dula[1], Mavis Menyere[1], Belson Kutambe[1], Charlotte van der Veer[1,11], Tamara Phiri[9], Ndaziona P. Banda[8], Kwazizira S. Mndolo[9], Kelvin Mponda[9], Chimota Phiri[9], Jane Mallewa[8], Mulinda Nyirenda[8,9], Grace Katha[9], Henry Mwandumba[1,2,8], Stephen B. Gordon[1,2], Kondwani C. Jambo[1,2,8], Jennifer Cornick[1,4], Nicholas Feasey[1,2], Blantyre COVID-19 Consortium, Kayla G. Barnes[+1,4,5,6,7], Ben Morton[+1,2,10], Philip M. Ashton[+&1,4]

1. Malawi-Liverpool-Wellcome Clinical Research Programme, Kamuzu University of Health Sciences, Blantyre, Malawi
2. Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, United Kingdom
3. Liverpool University Hospitals NHS Foundation Trust, Liverpool, United Kingdom
4. Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, United Kingdom
5. Harvard School of Public Health, Boston, USA
6. Broad Institute of MIT and Harvard, Cambridge, USA
7. University of Glasgow MRC Centre for Virus Research, Glasgow, UK
8. Kamuzu University of Health Sciences (formerly University of Malawi-College of Medicine) Blantyre, Malawi
9. Department of Medicine, Queen Elizabeth Central Hospital, Blantyre, Malawi
10. Liverpool University Hospitals Foundation Trust, Liverpool, UK.
11. Institute of Life Course and Medical Sciences, University of Liverpool, Liverpool, United Kingdom

* Contributed equally, + Contributed equally, & Corresponding authors
Corresponding authors: Philip Ashton (philip.ashton@liverpool.ac.uk) & Samantha Lissauer (slissauer@mlw.mw)

medRxiv preprint doi: https://doi.org/10.1101/2022.02.17.22269742; this version posted February 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Blantyre COVID-19 Consortium

Clinical
Wezzie Kalua, Peter Mandala, Barbara Katutula, Rosaleen Ng'oma, Steven Lanken, Jacob Phulusa, Mercy Mkandawire, Sylvester Kaimba, Sharon Nthala, Edna Nsomba, Lucy Keyala, Beatrice Chinoko, Markus Gmeiner, Vella Kaudzu, Bridget Freyne, Peter MacPherson, Todd D. Swarthout and Pui-Ying Iroh Tam.

Laboratory
Simon Sichone, Ajisa Ahmadu, Grace Stima, Mazuba Masina, Oscar Kanjewa, Vita Nyasulu, End Chinyama, Allan Zuza, Brigitte Denis, Evance Storey, Nedson Bondera, Danford Matchado, Adams Chande, Arthur Chingota, Chimenya Ntwea, Langford Mkandawire, Chimwemwe Mhango, Agness Lakudzala, Mphatso Chaponda, Percy Mwenechanya, Leonard Mvaya and Dumizulu Tembo

Data and statistics
Marc Y. R. Henrion, James Chirombo, Paul Kambiya, Clemens Masesa & Joel Gondwe

Keywords

COVID; SARS-CoV-2; ISARIC; Delta; mortality; LMIC; Malawi; Africa

medRxiv preprint doi: https://doi.org/10.1101/2022.02.17.22269742; this version posted February 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Abstract

Background

Compared to the abundance of clinical, molecular, and genomic information available on patients hospitalised with COVID-19 disease from high-income countries, there is a paucity of data from low-income countries.

Methods

We enrolled 245 hospitalised patients with PCR confirmed COVID-19 disease at Queen Elizabeth Central Hospital, the main hospital for southern Malawi, between July 2020 and September 2021. The recruitment period covered three waves of SARS-CoV-2 infections in Malawi. Clinical and diagnostic data were collected using the ISARIC clinical characterization protocol for COVID-19. The viral material from PCR-positive swabs was amplified with a tiling PCR scheme and sequenced using the MinION sequencer in Malawi. Consensus genomes were generated using the ARTIC pipeline and lineage assignment was performed using Pangolin.

Results

Sequencing data showed that wave one was predominantly B.1 (8/11 samples), wave two consisted entirely of Beta variant of concern (VOC) (6/6), and wave three was predominantly Delta VOC (25/26). Patients recruited during the second and third waves had progressively fewer underlying chronic conditions, and in the third wave had a shorter time to presentation (2 days vs 5 in the original wave). Multivariable logistic regression demonstrated increased mortality in wave three, dominated by the Delta VOC, compared to previous waves (OR 6.6 [CI 1.1-38.8]).

Conclusions

Patients hospitalised with COVID-19 disease and who were recruited to the ISARIC cohort, in Blantyre during the Delta wave had more acute symptom onset; fewer underlying conditions; and were more likely to die. Whilst we demonstrate the value of linking virus sequence data with clinical outcome data in a low-income setting, this study also highlights the considerable barriers to establishing sequencing capacity in a setting heavily affected by disruptions in supply chain and inequity of resource distribution.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.17.22269742; this version posted February 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Introduction

Policy makers need robust data to inform the clinical and public health response to the COVID-19 pandemic. The International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC) has developed a variety of tools and protocols to support the collection and analysis of data during the pandemic (1–3). These simplify the establishment of observational cohorts, and enable high-quality, harmonised, clinical research in response to emerging threats.

At Queen Elizabeth Central Hospital (QECH), Blantyre, patients have been enrolled under the ISARIC Tier 1 protocol since April 2020 (4). We previously demonstrated that, in the first wave of infection, patients admitted to hospital with suspected COVID-19 who were PCR negative, but IgG positive for SARS-CoV-2 had analogous immunological profiles to those who were PCR positive. These patients were less likely to receive COVID-19 specific treatments such as dexamethasone. Previously, however, there was limited sequencing capacity at our institution and no description of viral genomes was possible.

Genome sequencing has been essential to the global response to the COVID-19 pandemic. The early release of the Wuhan-1 genome sequence (5) enabled the development of specific diagnostic tests (6) and the design of mRNA vaccines, used to such great success in high-income countries (7,8). The evolution of the virus has led to the emergence of lineages designated variants of concern (VOCs), usually detected and defined by genome sequencing, and this has been one of the defining features of the pandemic to date (9,10). These VOCs have caused further global waves of infection with specific political and public health responses required for Alpha, Beta, Delta and Omicron VOCs. Linking of genomic data to clinical and public health data is important in determining the impact of viral mutations on disease severity and outcomes, particularly in areas where resources are constrained and there are high rates co-morbidity including HIV infection and TB (11).

Here, we describe the sequencing of the SARS-CoV-2 genomes from swabs collected from adult patients admitted to the hospital with symptomatic COVID-19 during three sequential waves of the pandemic. We place clinical outcome data in pathogen genomic context, to improve our understanding of the genomic epidemiology of the SARS-CoV-2 pandemic in Blantyre, Malawi.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.17.22269742; this version posted February 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Methods

Study design and recruitment

We prospectively recruited adult patients (>18 years) using the tier one sampling strategy from the International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC) Clinical Characterisation Protocol (CCP) (3), as previously described (4). Patients were recruited at Queen Elizabeth Central Hospital (QECH), Blantyre, Malawi, which is a large referral hospital in Southern Malawi.

During the recruitment period, patients with COVID-19 were cohorted in wards capable of providing oxygen therapy, but without capacity for invasive mechanical ventilation, intensive care facilities, continuous positive airways pressure (CPAP) or high flow oxygen.

Patients with suspected or confirmed SARS-CoV-2 infection were approached for informed consent with an aim to recruit within 72 hours of hospital admission. Respiratory samples (combined nasopharyngeal and oropharyngeal swab) and peripheral blood samples were collected at the point of patient recruitment. SARS-CoV-2 PCR diagnostic testing was carried out as previously described (4). For this study, only patients with a positive SARS-CoV-2 PCR test were included. Clinical and therapeutic data was taken from the clinical records, measured observations and history. Study protocols were approved by the Malawi National Health Science Research Committee (NHSRC, 20/02/2518 and 19/08/2246) and Liverpool School of Tropical Medicine Research Ethics Committee (LSTM REC, 20/026 and 19/017).

Statistical analysis

Clinical data were analysed using Stata V15.1 (StataCorp, Stata Statistical Software: Release 15, College Station, Texas, USA). Categorical variables were compared using Fisher's exact test. Continuous variables were tested for normality and appropriate statistical tests were applied; non-normally distributed measurements are expressed as the median [IQR] and were analysed by the Kruskal-Wallis test to compare clinical parameters across the three waves. The primary outcome variable was survival to hospital discharge. We selected the following covariates *a priori* to determine potential predictors of mortality: pandemic infection wave (W1: 04/2020 – 10/2020, W2: 11/2020 – 03/2021 and W3: 04/2021 – 08/2021); vaccine status; age; sex; HIV infection status; prior diagnosis cardiac disease; prior diagnosis diabetes mellitus; time from symptoms to hospital admission; respiratory rate; and SpO_2. All the above variables are available at, or shortly after, hospital admission. Univariable and multivariable logistic regression analyses were fitted using the STATA "logistic" command to generate odds ratios and confidence intervals (data and code available in supplementary materials). The overall statistical significance of the difference in mortality between waves was assessed using a likelihood ratio test, comparing the univariable model against a null, intercept-only model and the full multivariable model against a null model with all covariates except for the categorical variable encoding the epidemic wave. Exact binomial confidence intervals for the proportion of each genotype during each wave were calculated in R v4.1.0 (12) using the binom.test function.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.17.22269742; this version posted February 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

SARS-CoV-2 molecular biology and genome sequencing

Samples were extracted using the Qiasymphony-DSP mini kit 200 (Qiagen, UK) with offboard lysis. Samples were then tested using the CDC N1 assay to confirm the Ct values before sequencing. ARTIC protocol V2 sequencing protocol was used until June 2021, after which we switched to the V3 protocol. ARTIC version 3 primers were used for the tiling PCR until we switched to the University of Zambia (UNZA) primer set that provided good results for Delta VOC in August 2021 (13). Initially two primer pools were used, however a third pool was made for primer pairs that commonly had lower depth compared to the average (details Supplementary Table 1). PCR cycling conditions were adapted to the new sequencing primers, with annealing temperature changed to 60°C. Sequencing was carried out with the Oxford Nanopore Technologies MinION sequencer. Samples that had poor coverage (<70%) with the ARTIC primer set were repeated with the UNZA primer set.

Analysis of SARS-CoV-2 sequencing data

Raw FAST5 data produced by the MinION were processed with Guppy v5.0.7. FAST5s were basecalled with guppy_basecaller, basecalled FASTQs were assigned to barcodes using guppy_barcoder, including the `--require_barcodes_both_ends` flag. The per-sample FASTQ files were processed with the artic pipeline using the `medaka` option (14). The lineage of each consensus genome was identified using pangolin with the following versions; pangolin v3.1.17, pangolearn 2021-12-06, constellations v0.1.1, scorpio v0.3.16, pango-designation used by pangoLEARN/Usher v1.2.105, pango-designation aliases v1.2.122 (15). Samples were re-analysed when the Pangolin database was updated. The run was repeated if there was contamination in the negative control.

To set reasonable Ct thresholds for selecting samples to sequence in future work, we plotted the true positive rate versus the false positive rate (i.e. ROC curves) for a range of Ct thresholds from 15 to 40, where the true positive rate was defined as the proportion of samples with a genome coverage >=70% that had a Ct below the threshold. The false-positive rate was defined as the proportion of samples with a genome coverage <70% that had a Ct below the threshold. Code to calculate the values for the ROC curves is available here - https://gist.github.com/flashton2003/bb690261106dc98bb1ae5de8a0e61199.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.17.22269742; this version posted February 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Results

Clinical Characteristics

Between July 2020 and September 2021, we recruited 245 adults with COVID-19, using the ISARIC Clinical Characterisation Protocol. Participant characteristics are given in Table 1. Recruitment spanned three distinct waves of COVID-19 in Malawi; 1st wave n=48 (July-November 2020), 2nd wave n=94 (December 2020-March 2021), 3rd wave n=103 (June 2021-October 2021). More participants were recruited in waves 2 and 3, reflecting the epidemiology of COVID-19 in Malawi (Supplementary Figure 1). All participants had SARS-CoV-2 positivity confirmed by nucleic acid amplification tests (NAAT).

There were no significant differences in sex or median age between the waves (Table 1), however, there was a significant reduction (p=0.001) in time from symptom onset to presentation in wave three (median two days [IQR 1-5]) compared to wave one (median five days [IQR 2-8]) or two (median four days [IQR 2-9]). There was a decrease in the proportion of patients with cardiac disease (30% and 23.4% vs 3.9%, P <0.001) and diabetes (40% vs 19.2% vs 19.4% p=0.012) across waves. There was no difference in overall cohort survival on direct comparison (91.7% vs 90.4% vs 84.5%, P-value = 0.305), although there was a trend towards reduced survival in wave 3. Length of hospital stay was similar for all three waves (median eight days). There was a trend toward increased use of oxygen and significantly higher administration of oral and IV steroids during wave 3 (60.4% vs 59.6% vs 86.4% p=<0.001). Low numbers of patients were vaccinated; within this cohort 19/103 (18.4%) wave 3 participants had received the first dose (vaccine was unavailable in previous waves). Of these 16/19 (84.2%) survived to hospital discharge compared to 71/84 (84.5%) who had not been vaccinated (p=0.97).

Univariable logistic regression analysis demonstrated that age ≥70 (OR 13.64 CI: 1.62 – 114.52), respiratory rate (OR 9.35 CI: 1.88 – 46.53) and $SpO_2 \leq 87\%$ (OR 14.56 CI: 4.94 – 42.33) were associated with increased mortality (Table 2). After adjustment of all *a priori* specified variables within a multivariable model age ≥70 (OR 20.20 CI: 1.59 – 256.26), $SpO_2 \leq 87\%$ (OR 20.15 CI: 3.54 – 114.68) and admission during wave 3 (OR 6.59 CI: 1.11 – 38.85) were independently associated with increased mortality for our patient cohort. There was no contribution to outcome from vaccine status, sex, HIV infection, presence of co-morbidities days from symptoms to admission or respiratory rate within the multivariable model (Table 2). The multivariable likelihood ratio test for presence or absence of admission wave within the model demonstrated a significant effect ($Chi^2 = 6.31$, p = 0.043).

medRxiv preprint doi: https://doi.org/10.1101/2022.02.17.22269742; this version posted February 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table 1: Comparison of the demographic and clinical characteristics of COVID patients enrolled in ISARIC during three waves. UVA: Universal Vital Assessment score (16). TB positivity was defined according to presence of positive urinary LAM, GeneXpert or sputum test during hospital admission. Diabetes and Cardiac disease status ascertained from patient history and medical notes. # Proportion (%) positivity calculated using the denominator for individual variables (unknown status classified as missing data) and compared using the Chi^2 test. §: Median and IQR were compared using the Kruskal-Wallis test

Score	**Wave 1 (n=48)**	**Wave 2 (n=94)**	**W3 (n=103)**	**P-value**
Female	31.3% (15)	41.5% (39)	29.1% (30)	0.167 #
Age	52 (43 – 64)	46 (37 – 58)	51 (36 – 64)	0.258 §
Symptoms to admission (days)	5 (2 – 8)	4 (2 – 9)	2 (1 – 5)	**0.001** §
Admission to sample (days)	4 (2 – 5)	3 (2 – 7)	4 (2 – 5)	0.912 §
HIV seropositive	22.9% (11)	29.8% (28)	20.4% (21)	0.297#
TB positive	2.1% (1)	1.1% (1)	1.9% (2)	0.856#
Malaria positive	4.2% (2)	2.1% (2)	1.0% (1)	0.432#
Cardiac disease	30.0% (13)	23.4% (22)	3.9% (4)	**<0.001**#
Diabetes	40.0% (18)	19.2% (18)	19.4% (20)	**0.012**#
Oxygen at enrolment	50.0% (23)	58.5% (55)	65.1% (67)	0.215#
UVA	2 (0 – 3.5)	2 (0 – 3)	2 (0 – 4)	0.665 §
Beta-lactam antibiotic	81.3% (39)	68.1% (64)	90.3% (93)	**<0.001**#
Steroids administered (IV or PO)	60.4% (29)	59.6% (56)	86.4% (89)	**<0.001**#
Hospital survival	44 (91.7%)	85 (90.4%)	87 (84.5%)	0.305#
Survivor length of stay (days)	8 (6 – 18)	8 (4 – 16)	8 (6 – 11)	0.659 §

medRxiv preprint doi: https://doi.org/10.1101/2022.02.17.22269742; this version posted February 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table 2: Clinical factors associated with mortality for SARS-CoV-2 PCR confirmed patients admitted to hospital with severe acute respiratory infection. Univariable and multivariable logistic regression analysis with all pre-specified parameters included within the final multivariable model. Final multivariable model: n=226, $chi^2 = 62.80$, Pseudo $R^2 = 0.363$.

Variable	**Univariable**			**Multivariable**		
	Odds ratio	P value	Confidence Interval	Odds ratio	P value	Confidence Interval
Wave						
2	1.16	0.808	0.34 – 4.00	1.64	0.552	0.32 – 8.43
3	2.02	0.231	0.64 – 6.41	6.59	**0.037**	1.11 – 38.85
Vaccinated	1.44	0.581	0.39 – 5.29	0.59	0.569	0.09 – 3.66
Age						
30-39	1.56	0.715	0.14 – 18.23	0.44	0.580	0.03 – 7.86
40-49	5.87	0.101	0.71 – 48.73	2.04	0.540	0.21 – 20.16
50-59	2.09	0.531	0.21 – 21.10	0.26	0.395	0.01 – 5.68
60-69	3.33	0.291	0.35 – 31.44	0.75	0.834	0.05 – 11.14
≥70	13.64	0.016	1.62 – 114.52	20.20	**0.020**	1.59 – 256.26
Male	0.60	0.206	0.27 – 1.32	0.49	0.239	0.15 – 1.60
HIV infected	1.43	0.453	0.56 – 3.66	2.36	0.224	0.59 – 9.42
HIV unknown	1.73	0.236	0.67 – 4.46	1.55	0.537	0.39 – 6.25
Cardiac disease	1.08	0.877	0.39 – 3.04	1.00	0.998	0.19 – 5.18
Diabetes	0.85	0.739	0.33 – 2.20	0.82	0.772	0.21 – 3.19
Symptoms to admission (days)						
4-6	2.20	0.122	0.81- 6.01	2.59	0.210	0.58 – 11.43
7-9	1.89	0.280	0.60 – 5.99	6.22	0.056	0.95 – 40.66
≥10	1.60	0.412	0.53 – 4.61	1.75	0.483	0.37 – 8.31
Respiratory rate						
20-24	1.40	0.674	0.29 – 6.81	0.56	0.542	0.09 – 3.56
25-29	2.29	0.319	0.45 – 11.68	0.26	0.199	0.03 – 2.05
≥30	9.35	0.006	1.88 – 46.53	1.62	0.652	0.20 – 13.21
SpO2						
93-95	1.20	0.781	0.33 – 4.27	0.82	0.801	0.17 – 3.88
88-92	1.90	0.296	0.57 – 6.32	0.04	0.954	0.26 – 4.26
≤87	14.56	<0.001	4.94 – 42.33	20.15	**0.001**	3.54 – 114.68

medRxiv preprint doi: https://doi.org/10.1101/2022.02.17.22269742; this version posted February 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Molecular testing

Confirmatory reverse transcription-quantitative PCR (RT-qPCR) at enrolment demonstrated that 102/245 participants remained positive. Ct values were available for 95/102 confirmatory RT-qPCR positive cases, and there was no significant difference in median Ct between waves (Supplementary Figure 2).

Sequencing results

We sequenced 102 samples from 102 patients and obtained 43 genomes with more than 70% coverage at 20x depth (Supplementary Table 2). Low coverage of the genome (<70%) was related to low viral load. This was true for both ARTIC v3 and UNZA tiling PCR primer sets separately (Figure 1). Overall, the median Ct value of samples with <70% coverage was 30.7, compared with 24.5 for those above this threshold (Supplementary Table 2). ARTIC v3 produced significantly lower median genome coverage than UNZA for samples with Ct values less than 30 (68% vs 76%, Kolmogorov-Smirnov P-value = 0.0003).

Characteristics of the sub-group of patients whose SARS-CoV-2 consensus genome had >=70% coverage are available in Supplementary Table 3. Successful sequencing was more likely in females who formed only 34% of participants but gave rise to 63% of high coverage sequences.

We produced ROC curves showing the True Positive Rate and False Positive Rate at a range of Ct thresholds (Supplementary Figure 2). Based on visual inspection of these ROC curves, we chose Ct value thresholds of 28 for ARTIC v3 and 27 for UNZA as they provided a balance between reducing wasted sequencing runs, and generating as many sequences as possible for our purposes.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.17.22269742; this version posted February 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

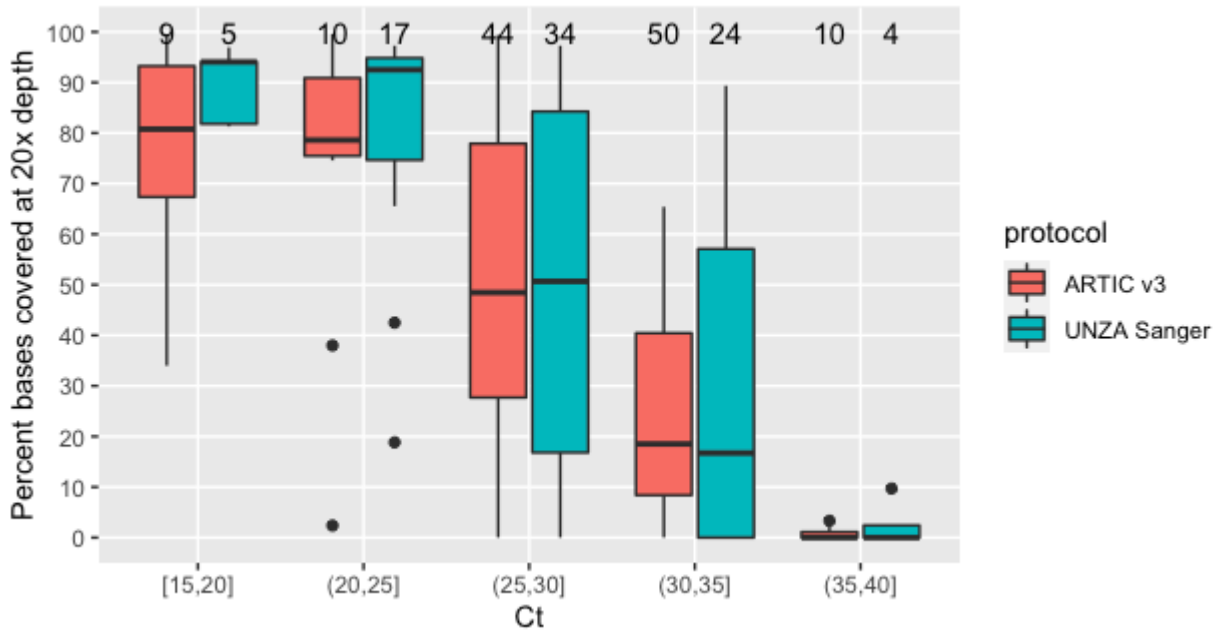

Figure 1: Relationship between PCR Ct value and the percentage of the SARS-CoV-2 reference genome covered to at least 20x depth. The number at the top of each column is the number of samples for the two protocols in each bin of the box plot.

Identification of SARS-CoV-2 lineages

We observed three pangolin lineages among the 11 SARS-CoV-2 samples from wave 1 (Figure 2, Supplementary Table 2), with the most frequently identified pangolin lineage being B.1 (n=8), followed by B.1.1 (n=2) and B.1.1.448 (n=1). One hundred percent (6/6) of samples from wave 2 were VOC Beta (exact binomial 95% CI of the estimate in the untested population = 54-100%) and 96% (25/26) of samples from wave 3 were VOC Delta (95% CI 80-100%) (Figure 2). One sample received at the beginning of June 2021 was VOC Beta. We observed seven pangolin lineages among the 25 VOC Delta samples sequenced during wave 3; 11 AY.75.1, 8 B.1.617.2, 2 AY.75 and 1 each of AY.50, AY.59, AY.122 and AY.72 (Supplementary Figure 3). Due to low numbers of successfully sequenced isolates during the second wave, we also investigated the genotype of samples from Malawi submitted to GISAID during this time; Beta VOC accounted for 324 of the 349 (93%, 90-95%) SARS-CoV-2 genomes from Malawi in GISAID which were sampled.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.17.22269742; this version posted February 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

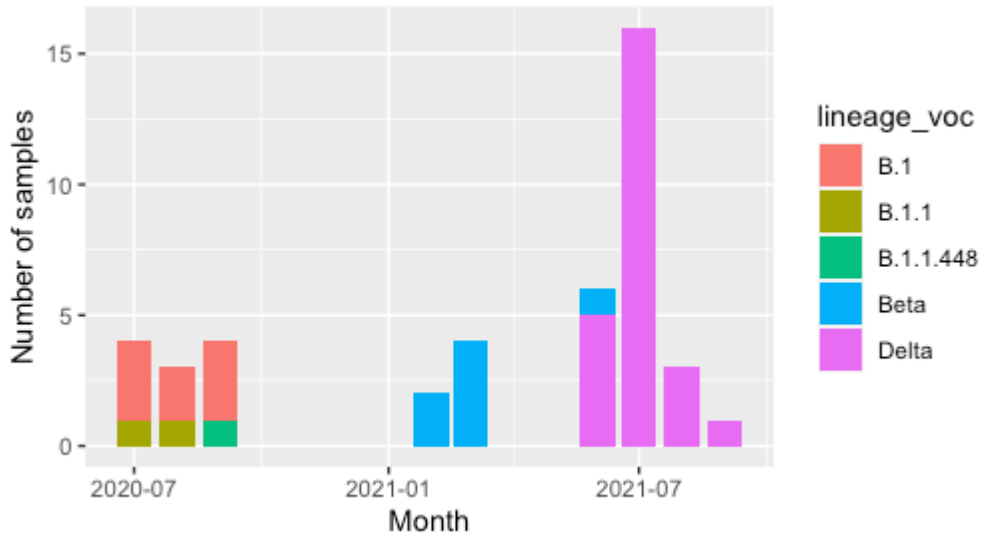

Figure 2: A histogram of the monthly number of each lineage or VOC identified in patients in our cohort.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.17.22269742; this version posted February 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Discussion

We established a platform for genome sequencing and analysis in Blantyre, Malawi and used it to sequence SARS-CoV-2 from a cohort of patients hospitalised with COVID-19 to investigate whether and how variants of concern (VOCs) influenced clinical outcomes. The first wave was predominantly B.1 and B.1.1. All successfully sequenced cases during the second wave were caused by Beta VOC. Whilst the number of successfully sequenced cases from the second wave was low, our data are consistent with data reported to GISAID from other researchers in Malawi confirming the dominance of Beta VOC in the second wave, whilst the Delta VOC dominated the third wave.

Age ≥70 and SpO_2 ≤87% at admission were independently associated with increased risk of death within both univariable and multivariable analyses. Our patient cohort presented with fewer chronic medical conditions in the second and third waves (cardiac disease and diabetes) but were more likely to be administered treatments such as steroids and antibiotics. This may represent increased adherence to local treatment guidelines and improved clinical experience in managing COVID-19 and/or that the Beta and Delta VOCs were associated with more severe illness in otherwise healthy individuals (16). Time to hospital presentation was significantly lower in the third wave, potentially suggesting that disease progression was more rapid or that patients were more aware of the need to present to hospital earlier, or that people had higher trust in the ability of the healthcare system to manage COVID-19. Multivariable analysis demonstrated that in-patient mortality amongst the recruited cohort was higher during the third/Delta VOC wave, compared to other waves (17–19). Throughout the study there was no invasive or non-invasive ventilatory support available for COVID-19 patients and no access to Interleukin 6 antagonists, which are recommended for severe disease by the WHO (since July 2021). For clinical comparisons, our recruited cohort represented a sample of those presenting to hospital, mediated by clinical decisions and guidelines which changed over time. Together with population-level changes in health-seeking behaviour, caution is warranted in the interpretation of excess mortality being due to genetic variant alone. However studies from other settings have demonstrated increased hospitalisation or death in patients infected with the Delta VOC compared to other genetic lineages (17,19). There is a paucity of linked clinical data and sequencing data from LMIC settings, despite it being a hugely valuable resource and providing contextually useful information. This finding supports ongoing research, upscaling of sequencing capacity and highlights the importance of collaborative platforms such as ISARIC to draw firm conclusions about the impact of genetic variants across the sub-Saharan African region.

No patients in this cohort were fully vaccinated, with 18% of patients in the third wave having received one vaccine. Malawi introduced COVID vaccination in March 2021 between the second and third COVID waves. As of October 1st 2021, at the end of the third wave, 2.5% of the population of Malawi were fully vaccinated (available vaccines at that time were Oxford/AstraZeneca ChadOx1-S and Johnson and Johnson), with a further 2.5% having received a single dose of Oxford/Astra-Zeneca recombinant vaccine (Public Health Institute of Malawi publicly available data). Although numbers of vaccinated participants are low, there is a higher proportion of vaccinated individuals within the cohort than in the general population, and the reasons behind this are not clear. This may represent a more COVID-aware population attending the treatment centres or increased access/uptake of vaccines during the COVID wave by people

medRxiv preprint doi: https://doi.org/10.1101/2022.02.17.22269742; this version posted February 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

in urban centres. Given the small numbers and recent introduction of vaccines with intermittent availability it is difficult to draw conclusions from this dataset. With an overall rate of complete vaccination of 4% Malawi is below the continental fully vaccinated rate of 11% (20), these low rates illustrate the unique challenges and inequities in tackling COVID-19 in LMIC.

Vital to our success in establishing surveillance of SARS-CoV-2 in Malawi was the portability of the MinION sequencer; the public lab protocols (18); bioinformatics software from the scientific community (13); and the infrastructure and funding available to us as an international research institution. The MinION has become a vital part of outbreak response, as demonstrated for SARS-CoV-2 in Africa (19,20) and elsewhere, and also during previous emerging viral outbreaks such as Ebola (21) and Zika (22). However, even with a portable and low-maintenance sequencer (with no service contracts or engineer visits required), experienced molecular biologists and bioinformaticians, and considerable international support, it was still very challenging to establish sequencing capability. We found it difficult to procure reagents, and this barrier to establishing sequencing capacity was compounded by border closures and travel restrictions.

The pandemic has highlighted the inequity of health-related resource distribution and reinforced the need for prioritised distribution networks and more regional manufacturing of laboratory equipment and consumables. While the MinION sequencing platform is easily set up, the need for cold chain reagents and the short shelf life of flow cells makes maintaining a real-time sequencing service difficult. The development of more stable reagents, such as lyophilised enzymes, would increase the affordability and accessibility of this technology. Computationally, the inconsistent internet at the time of this study was a hurdle in setting up a server with the requisite software installed. The current bioinformatic trends of containerisation (i.e. where the software required is setup and packaged by a third party, alongside the operating system and dependencies required to run the software) and virtual environments are significant advantages for reproducibility, but they are “greedy” in terms of bandwidth. To install a single tool often requires the download of an entire operating system in the form of a Docker container. As our computer hardware was based in Blantyre, Malawi, once the initial setup was achieved, we did not need to transfer large amounts of data internationally, which was a significant advantage given the intermittent internet connection. Using a bioinformatics “lab-on-an-SSD” is one potential approach to solving the challenges of computational setup in settings with inconsistent internet connection.

Our study has several limitations. Firstly, we produced a relatively small number of sequences. This was partly due to the limited number of patients recruited into the study during each wave but also because patients frequently presented with Ct values that were too high to produce good quality sequence data. Secondly, our observations are limited to a single centre in the Southern region of Malawi, however they appear to be broadly consistent with the national picture. Finally, we may not be capturing the full diversity of SARS-CoV-2 circulating in the community, as our sampling of hospitalised patients represents a considerable bias towards people with severe disease, and there is likely to be significant under ascertainment nationally (21).

medRxiv preprint doi: https://doi.org/10.1101/2022.02.17.22269742; this version posted February 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

This inequity in the availability of clinical and preventative interventions was mirrored by the lack of timely sequencing data available to inform national public health measures and to contribute to international databases. The recent Omicron VOC was first described in South Africa in November 2021 because facilities were available to link clinical and laboratory observations – despite the barriers we faced, at the start of the fourth wave, we were able to confirm the presence of Omicron VOC within 4 weeks of its first detection globally and within three days of the swab being taken.

Acknowledgments

The authors thank all study participants and the staff of the Queen Elizabeth Central Hospital (QECH) for their support and co-operation during the study. We would like to thank all the people mentioned in Supplementary File 1 for sharing their data to GISAID.

This work was supported by the UK Foreign, Commonwealth and Development Office and Wellcome grants for SARS-CoV-2 diagnostics [220757/Z/20/Z] and the MLW Core grant [206545/Z/17/Z].

Conflict of interest statement

We have no conflicts of interest to declare.

Data availability statement

All genome sequences are available in GISAID and INSDC databases – accessions are available in Supplementary Table 2.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.17.22269742; this version posted February 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

References

1. Abbas A, Abdukahil SA, Abdulkadir NN, Abe R, Abel L, Absil L, et al. The value of open-source clinical science in pandemic response: lessons from ISARIC. The Lancet Infectious Diseases. 2021 Dec 1;21(12):1623–4.

2. Drake TM, Riad AM, Fairfield CJ, Egan C, Knight SR, Pius R, et al. Characterisation of in-hospital complications associated with COVID-19 using the ISARIC WHO Clinical Characterisation Protocol UK: a prospective, multicentre cohort study. The Lancet. 2021 Jul 17;398(10296):223–37.

3. Clinical Characterisation Protocol (CCP) [Internet]. ISARIC. [cited 2021 Oct 28]. Available from: https://isaric.org/research/covid-19-clinical-research-resources/clinical-characterisation-protocol-ccp/

4. Morton B, Barnes KG, Anscombe C, Jere K, Matambo P, Mandolo J, et al. Distinct clinical and immunological profiles of patients with evidence of SARS-CoV-2 infection in sub-Saharan Africa. Nat Commun. 2021 Jun 11;12(1):3554.

5. Novel 2019 coronavirus genome - SARS-CoV-2 coronavirus [Internet]. Virological. 2020 [cited 2021 Oct 28]. Available from: https://virological.org/t/novel-2019-coronavirus-genome/319

6. CDC. CDC's Diagnostic Test for COVID-19 Only and Supplies [Internet]. Centers for Disease Control and Prevention. 2020 [cited 2021 Oct 28]. Available from: https://www.cdc.gov/coronavirus/2019-ncov/lab/virus-requests.html

7. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. New England Journal of Medicine. 2020 Dec 31;383(27):2603–15.

8. Baden LR, El Sahly HM, Essink B, Kotloff K, Frey S, Novak R, et al. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. New England Journal of Medicine. 2021 Feb 4;384(5):403–16.

9. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations - SARS-CoV-2 coronavirus / nCoV-2019 Genomic Epidemiology [Internet]. Virological. 2020 [cited 2021 Oct 28]. Available from: https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563

10. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa [Internet]. 2020 Dec [cited 2021 Dec 31] p. 2020.12.21.20248640. Available from: https://www.medrxiv.org/content/10.1101/2020.12.21.20248640v1

medRxiv preprint doi: https://doi.org/10.1101/2022.02.17.22269742; this version posted February 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

11. Why genomic sequencing is crucial in COVID-19 response [Internet]. WHO | Regional Office for Africa. [cited 2022 Jan 24]. Available from: https://www.afro.who.int/news/why-genomic-sequencing-crucial-covid-19-response

12. R: A language and environment for statistical computing. [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available from: https://www.R-project.org/

13. Simulundu E, Mupeta F, Chanda-Kapata P, Saasa N, Changula K, Muleya W, et al. First COVID-19 case in Zambia - Comparative phylogenomic analyses of SARS-CoV-2 detected in African countries. Int J Infect Dis. 2021 Jan;102:455–9.

14. ARTIC field bioinformatics pipeline [Internet]. GitHub. [cited 2021 Oct 28]. Available from: https://github.com/artic-network/fieldbioinformatics

15. O'Toole Á, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, et al. Assignment of Epidemiological Lineages in an Emerging Pandemic Using the Pangolin Tool. Virus Evolution. 2021 Jul 5;veab064.

16. Banda NP, Hara W, Cocker D, Musasa S, Burke RM, Brown C, et al. First case report of a successfully managed severe COVID-19 infection in Malawi. Malawi Med J. 2020 Dec;32(4):226–8.

17. Twohig KA, Nyberg T, Zaidi A, Thelwall S, Sinnathamby MA, Aliabadi S, et al. Hospital admission and emergency care attendance risk for SARS-CoV-2 delta (B.1.617.2) compared with alpha (B.1.1.7) variants of concern: a cohort study. The Lancet Infectious Diseases. 2022 Jan 1;22(1):35–42.

18. Bager P, Wohlfahrt J, Rasmussen M, Albertsen M, Krause TG. Hospitalisation associated with SARS-CoV-2 delta variant in Denmark. The Lancet Infectious Diseases. 2021 Oct 1;21(10):1351.

19. Bast E, Tang F, Dahn J, Palacio A. Increased risk of hospitalisation and death with the delta variant in the USA. The Lancet Infectious Diseases. 2021 Dec 1;21(12):1629–30.

20. COVID-19 Vaccination [Internet]. Africa CDC. [cited 2022 Feb 9]. Available from: https://africacdc.org/covid-19-vaccination/

21. Mandolo J, Msefula J, Henrion MYR, Brown C, Moyo B, Samon A, et al. SARS-CoV-2 exposure in Malawian blood donors: an analysis of seroprevalence and variant dynamics between January 2020 and July 2021. BMC Medicine. 2021 Nov 19;19(1):303.