Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Molecular epidemiology of COVID-19 in Oman: A molecular and surveillance study for the early transmission of COVID-19 in the country

Samira Al-Mahruqi\textsuperscript{a}, Adil Al-Wahaibi\textsuperscript{a}, Abdul Latif Khan\textsuperscript{b}, Amina Al-Jardani\textsuperscript{a}, Sajjad Asaf\textsuperscript{b}, Hanan Alkindi\textsuperscript{a}, Samiha Al-Kharusi\textsuperscript{a}, Ahmed N. Al-Rawahi\textsuperscript{b}, Ahmed Al-Rawahi\textsuperscript{b}, Majid Al-Salmani\textsuperscript{b}, Intisar Al-Shukri\textsuperscript{a}, Aisha Al-Busaidi\textsuperscript{a}, Seif Salem Al-Abri\textsuperscript{a,}\textsuperscript{a}, Ahmed Al-Harrasi\textsuperscript{a,}\textsuperscript{a}.

\textsuperscript{a} Directorate General for Disease Surveillance and Control, Ministry of Health, Sultanate of Oman, P.O. Box 393, PC 100 Muscat, Oman
\textsuperscript{b} Natural and Medical Sciences Research Center, University of Nizwa, P.O. Box 33, PC 616 Nizwa, Oman

\section*{A R T I C L E   I N F O}

Article history:
Received 15 November 2020
Received in revised form 13 December 2020
Accepted 17 December 2020

Keywords:
SARS-CoV-2
COVID-19 epidemiology
Phylogenomic analysis
Clade
Mutation
Oman

\section*{A B S T R A C T}

Background: Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) has been proven to be lethal to human health, which affects almost every corner of the world. The objectives of this study were to add context to the global data and international genomic consortia, and to give insight into the efficiency of the contact tracing system in Oman.

Methods: We combined epidemiological data and whole-genome sequence data from 94 samples of SARS-CoV-2 in Oman to understand the origins, genetic variation, and transmissibility. The whole-genome size of sequence data was obtained through a customized SARS-CoV-2 research panel. Amplifier methods ranged from 26 Kbp to 30 Kbp and were submitted to GISAID.

Findings: The study found that P323L (94.7\%) is the most common mutation, followed by D614G (92.6\%) spike protein mutation. A unique mutation, I280V, was first reported in Oman and was associated with a rare lineage, B.1.113 (10.6\%). In addition, the study revealed a good agreement between genetic and epidemiological data.

Interpretation: Oman’s robust surveillance system was very efficient in guiding the outbreak investigation processes in the country, the study illustrates the future importance of molecular epidemiology in leading the national response to outbreaks and pandemics.

© 2020 The Author(s). Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

\section*{Introduction}

The severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) (\cite{Paules2020, Wu2020}) outbreak began in Wuhan, Hubei province, China in late December 2019 and spread globally with extraordinarily high rates of morbidity and mortality. As of December 9, 2020, more than 67.5 million confirmed cases of coronavirus disease-2019 (COVID-19) have been reported, which caused more than 1.5 million deaths (\cite{WHO2020}). Coronavirus, subfamily Coronavirinae from family Coronaviridae, comprises a large group of viruses that have been mainly associated with birds and mammals (\cite{Khan2005, Fehr2015, Paules2020}). The subfamily includes four genera: Alphacoronavirus, Betacoronavirus, Gammacoronavirus, and Deltacoronavirus.

The coronaviruses were first identified and characterized before the 1960s and were attributed to cause a disease of the respiratory system (\cite{Khan2005, Paules2020}). SARS-CoV-2 has spread among humans to cause catastrophic devastation in health systems and economies. Given the possible high mutation rate of RNA viruses as compared to DNA viruses, it was not unexpected that the viral genome of SARS-CoV-2 would mutate more rapidly, which should allow to track the spread of the virus (\cite{Grubaugh2019}). However, with confirmed cases exceeding...
46.5 million, it has become difficult to track individual transmission chains due to the small size of sequences, notwithstanding more than 147 thousand whole-genome sequences, which have been deposited in GISAID, an open access global science initiative, as of October 23, 2020 (GISAID, 2020). Another striking feature of coronaviruses, when compared with other RNA viruses, is the likelihood to mutate slowly due to their ability to proofread; a fact accounted for 3’ to 5’ exoribonuclease (Minskaia et al., 2006). It has been reported that the estimated mutation rate of SARS-CoV-2 is 0.71–1.40 × 10−3 (Hill and Rambaut, 2020).

The first confirmed cases of COVID-19 was reported in Oman on 24 February of two Omani nationals who returned from a visit to Iran. Initially, a steady increase in the number of imported confirmed cases was observed in the following weeks, with the first suspected case of local transmission reported on March 23, 2020 in Mutrah, a city within Muscat, the capital of Oman (Al Wahaibi et al., 2020). To date, the number of infected cases has increased to more than 115,000 and approximately 1200 deaths (Ministry of Health, Oman, 2020) were recorded. The cases continue to increase up to 400 infections daily.

The aim to perform retrospective research using both whole viral genome sequencing and molecular and epidemiological data to understand the clusters of SARS-CoV-2 infection in Oman was to add context to the global data and international genomic consortiums, and to give insight into the efficiency of the contact tracing system in the country.

Material and methods

Outbreak investigation and early transmission of SARS-CoV-2 in Oman

Data of laboratory-confirmed COVID-19 surveillance of Oman were included in this study from February 24 to May 23, 2020. Samples were selected based on epidemiological characteristics that represent various cluster types (company versus family), geographical location, and nationality. The epidemiological investigation was conducted by public health teams throughout the country, who fed data into a national electronic database. All confirmed cases underwent epidemiological investigations and included the information of each patient’s demographics (age, gender, residency, and nationality), epidemiology (source of infection if known, date of onset, primary case [infector] or secondary case [infected], and designation [cluster or sporadic]). The data were then analyzed to find out the relationship between the patients using the EpicontactR library; further the description of the methods is found in a recent article by Al Wahaibi et al. (2020).

Specimen collection and inclusion criteria

Respiratory specimens collected from confirmed cases per the Ministry of Health national case definition were used in this study. Samples were selected from main clusters since the start of local transmission for a period of two months, from March 23 to May 5, 2020.

RNA extraction, library preparation, and sequencing

A new method for SARS-CoV-2 sequencing by Thermo Fisher Scientific was followed (Supplementary information 1). Briefly, the RNA extraction of samples was carried out through MagMAX™ Viral Pathogen Extraction Kit (Thermo Fisher Scientific, Waltham, MA, USA) or Viral RNA Isolation Kit with LifeRiver EX3600 (LifeRiver Biotech, Hangzhou Bay, China), following manufacturer protocol. For the detection of the SARS-CoV-2 virus by real-time polymerase chain reaction (RT-PCR) system, Novel Coronavirus (2019-nCoV) Nucleic Acid Diagnostic Kit, CE-IVD, FDA-EUA (Sansure Biotech, Changsha, China) was used in accordance with the manufacturer’s instructions. The assay targets two genomic regions of the SARS-CoV-2 (N and ORF1ab). All samples included in the study had a cycle threshold (Ct) value of the gene target of less than 30. SuperScript VILO cDNA Synthesis Kit (Thermo Fisher, USA) was used to reverse transcribe the SARS-CoV RNA with the qPCR program (42 °C for 30 min, 85 °C for 5 min, and hold on 10 °C). To ensure enough cDNA content for NGS workflow, we quantified it on Qubit 3.0 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). The cDNA (60 ng/ul) was used to prepare libraries with Ion Xpress barcodes (Thermo Fisher Scientific, Waltham, MA, USA) through Ion AmpliSeq Library Kit Plus and predefined Ion AmpliSeq SARS-CoV-2 Research Panel (Thermo Fisher Scientific, Waltham, MA, USA). The panel provides 99% coverage to viral genomes with fewer copy numbers and consists of two primer pools comprising 237 amplicons of SARS-CoV-2 genomes. The panel comprises five additional primer pairs that target human expression control. Workflow according to manufacturer guidelines was used to prepare libraries and quantify, was template-enriched, and sequenced on Ion 530 Chip Ion Torrent S5 (Thermo Fisher Scientific, Waltham, MA, USA).

Genome assembly, annotation, and analysis

The Torrent Suite Software (version 12) (Thermo Fisher Scientific, Waltham, MA, USA) with SARS-CoV-2 plugins (COVID19AnnotateSnEff, IRMAreport, and AssemblerTrinity) were installed and preoptimized with the reference sequence (ion_ampisiseq_sars-cov2) and target regions (ion_ampisiseq_SARS-CoV-2.2020323). Designed.bed) to trim, filter, quality check, assemble, and annotate the samples. Additional built-in plugins such as coverage analysis and variant caller were also used to understand the SNPs and mutations. Specifically, iterative refinement meta-assembler plug-ins were used to identify low-frequency variants for highly variable RNA viruses as revealed by the manufacturer’s guidelines. To further validate, the obtained sequenced data were remapped with Wuhan-Hu-1 (GenBank accession number NC_045512.2).

Phylogenomic analysis

The sequence alignment and phylogenetic trees were generated by following the Rob Lanfear’s method (Roblafand and Richard, 2020) for a global phylogeny of SARS-CoV-2. All the sequences were downloaded from GISAID websites and created a global sequence alignment. First, every sequence was aligned to the Wuhan reference genome (NC_045512.2) with the help of the script, global_profile_alignment.sh, then joining the individually aligned sequences into a global alignment at the end. To construct a phylogenetic tree, the maximum likelihood (ML) method was used using the FastTree software (MicrobesOnline, Berkeley, CA, USA) with the best setting determined by GISAID. We further optimized that tree with a series of minimum evolution SPR moves and ML NNI moves in FastTree. The Letunic and Bork’s tool (2007) was used to visualize the trees. In the second phylogenetic analysis, 349 complete SARS-CoV-2 genome sequences (including 203 sequences from Oman) were downloaded from GISAID. These sequences were aligned to the Wuhan reference genome (NC_045512.2) using the above script. The above phylogenetic inference method was used to infer the tree as described above.

Results

Epidemiology of SARS-CoV-2 infection in Oman

In this study, 21 clusters were identified with a total of 456 laboratory-confirmed cases of which 94 representative cases of
different clusters with high viral load were selected for whole-genome sequencing. Figure 1 describes selected clusters with their month of transmission, genotype, geographical location, and type of cluster. Significant correspondence between the epidemiological and genetic data was observed. Five were major intrafamilial clusters, three from the workplace, and two clusters were a mixture of both. The results also showed that the initial SARS-CoV-2 infection cases were detected in March in the capital city of Muscat within a major trade and tourist destination (Mutrah) and were classified epidemiologically as cluster B (Figure 1). This cluster expanded during the months of April and May and spread to North Sharqiyah and Dhahira as a result of community transmission and trade. Cases selected from this cluster belonged to B.1.1.27 (GR) genotype, and most of the clusters were distributed across different geographical locations because the origin of the cluster emerged from major cluster B extended to other locations. In addition, cluster P occurred in a company, and labor camps belonged to clade B.1.127 (GR) genotype and spread the epidemic into four geographical locations (Muscat, Dakhiliyah, Al Batinah South, and Dhofar governorate). Family clusters have also contributed to the spread of the epidemic to the entire country, as seen in cluster V, C, and I. As shown in Table 1, around 50% of the cases were males, Omani nationals, and from Muscat governorate. The majority (78.7%) were from the working age group, (15–50

Figure 1. Location of the governorates of the genotyped sample with the corresponding cumulative incidence of lab-confirmed COVID-19 per 100,000 population (part A), and the weekly distribution of the genotyped sample with the total weekly confirmed cases (part B) in Oman from March 15 to May 23, 2020.
years). The results showed that the most frequent genotype was B.1.1.27 (GR), 58.5%, followed by B.1.1 (GR), 19.1%.

Whole-genome sequencing of SARS-CoV-2 samples

An amplicon-sequencing approach developed by Ion Torrent for COVID-19 was adopted for the project that yielded a total of 13 Gb of sequence data and over 56 million reads from 94 SARS-CoV-2 samples (Supplementary Table S1). The genome coverage was 99.8%, and the average genome size of SARS-CoV-2 was 29,834 bp (Supplementary Table S2). SARS-CoV-2 genome sequences generated in this study have been deposited to GISAID.1 A list of all sequences with their accession numbers is provided in Supplementary Table S3.

SARS-CoV-2 genomes from Oman

The results of whole-genome sequencing of SARS-CoV-2 revealed the presence of common mutations when compared with the Wuhan reference sequence (hCoV-19/Wuhan/WIV04/2019). In this study, around 66 variants were identified with variable frequencies and across different gene regions. The most prevalent mutation was P323L (94.7%) found in the non-structural protein 12 followed by the D614G (92.6%) in the Spike glycoprotein (Supplementary Figure S1). Another mutation (G715) in the non-structural protein 5 region was present at a frequency of 70% in the study sample (Supplementary Figure S1). Some of the identified mutations were defining a specific clade such as G204R and D614G in the Spike protein that constitutes the B.1.1.27 and the B.1.1 (GR). Interestingly, the common G715 mutation was not seen in the family cluster I, which belonged to B.1.1 (GR).

Two unique missense mutations were detected in this study; I280V in the NSP15 and R502C in the NSP13, which existed at low frequencies of 6 (5.3%) and 2 (2.1%), respectively (Supplementary Figure S1). The mutation I280V belonged to lineage B.1113 (GH), it was detected in cluster V in Al Batinah South and Muscat. The second mutation was detected in two cases that belonged to B.11 (GR) lineage within the E cluster of the Al Batinah South region (EPI_ISL_491143 and EPI_ISL_491144) as shown in Figure 2.

Phylogenomic analysis

We performed a detailed phylogenomic analysis through ML of the 94 sequenced samples. The whole viral genomes were aligned through a multiple alignment analysis and compared with the Wuhan reference genome (Gbench: MN908847). The phylogenetic analysis was also supplemented with epidemiological datasets of epi week of infection onset and distribution of the identified clusters across different geographical locations in Oman (Figure 3). Generally, there was a good correlation between the epidemiologically defined clusters and their phylogenomic relationship as shown with cluster B, V, I, G, R, X, and W. The phylogenetic tree in Figure 3 showed that cluster B started early, under a month following the first laboratory-confirmed cases identified in Oman. The outbreak continued for around six weeks. Sequences from this cluster belonged to B.1.1.27 (GR) and all were closely related to each other except for one case lineage B.0 (O) but was found to be epidemiologically linked to the index case and was within the same geographical location. It is also observed from Figure 4 that during weeks 16 and 17, multiple outbreaks were detected that affected different geographical locations; clusters: B, C, H, I, J, O, V, and W. The clusters were found to belong mostly to B.1.1 lineage except for one cluster (V) that was B.1.36 (GH). Some clusters were localized in specific governorates that shared a monophyletic clade within the phylogenetic tree such as cluster I, which belonged to Dhabira and cluster V that was found in Al Batinah South having >90% bootstrap support.

To compare the distribution of SARS-CoV-2 sequences from Oman with other countries in the region and globally, we obtained a subset of sequences from neighboring countries and cases with travel history within the same period of this study. In the second phylogenetic tree (Figure 4), we analyzed 349 complete SARS-CoV-2 genomes sequences (including 203 sequences from Oman) that were downloaded from GISAIDS on September 1, 2020. The results revealed that most sequences from Oman are closely related to each other and shared the B.1.1 (GR) clade with related genomes within the tree (Figure 4). Those sequences were closely related to sequences from the United Arab Emirates, the Kingdom of Saudi Arabia, Kuwait, Bahrain, Bangladesh, Southeast Asia, and Europe. Another clade, the B.1.36 (GH), was closely related to sequences obtained from KSA, Bahrain, Bangladesh, Iran, Tunisia, and UK. Additionally, it was observed from the tree that Oman had sequences from most of the identified clades or Pangolin lineages. (Figure 4).

Discussion

To add context to the global data and international genomic consortiums, understand the spread of the virus, and support the epidemiological surveillance for pandemic management, in Oman a sequencing initiative for SARS-CoV-2 was established. Herein, we describe the sequencing of genomes of 94 samples of SARS-CoV-2 selected based on epidemiological and laboratory characteristics, collected from initial cases from the early outbreak clusters between March 22 and May 23, 2020. The consistency between the epidemiological and genetic information highlights the effectiveness of epidemiological surveillance and outbreak investigations. The epidemiological data illustrate the extensive interaction between different types of clusters (company and family) and the vast spread of the infection between other governorates in Oman. This could be explained by the origin of the epidemic that started from the main trade and tourist area in Muscat. Furthermore, the spread of infections between governorates was
propagated by the gatherings among extended Omani families as well as in crowded labor camps, thereby resulting in infection among nonnationals.

In this study, common and unique variants were identified from the genomic analysis of SARS-CoV-2. The D614G mutation in the Spike protein has been reported in 116 countries mostly of B1 lineages (GR, G, and GH clades). In Oman, it has been detected since March 2020 among patients with travel history to Europe (UK, Turkey, Spain, and Netherlands), India, Tanzania, and Qatar (Al-Mahruqi et al., unpublished data). The D614G was found to be the second-highest mutation with a frequency of 92.6% (87/94) in our study. Sallam et al. (2020) showed that the D614G mutation appeared to be taking over COVID-19 infections in the Middle East and North Africa (MENA) region as a significant increase in the proportion was noticed from 63.0% in February 2020 to 98.5% in June 2020 ($p < 0.001$). Two large phylogenetic clusters were identified through the ML analysis, which showed the evidence of intercountry mixing of sequences dating back to February 8, 2020.
and March 15, 2020. Another commonly observed mutation in this study, P323L, was described as the most common in other countries; furthermore, the increasing ratio of P323L indicates that this type of mutation may favor and enhance the transmission capacity of SARS-CoV-2 (Wang et al., 2020). Both D614G and P323L were prevalent and coexisted in this study with almost the same frequency. It was suggested that these coevolving mutations and how they could impact viral fitness, breadth, and complexity of clinical symptoms may be associated with new mutations and adaptations (Kannan et al., 2020). According to GISAID, this G71S common mutation was first reported in February 2020 in Germany (hCoV-19/Germany/BW-ChVir-1577/2020) and belongs to the B.1 lineage (G and GR clades) (https://www.gisaid.org/, 6–11–2020). This mutation was reported in 26 countries, Oman had the second-highest prevalence rate of 125/203 (61%) of the globally reported cases (COVID-19 Genomics UK Consortium).

The unique mutation NSP15-I280V detected in this study is in the endoRNase of the NSP15 genomic region of SARS-CoV-2. The I280V mutation has been reported in only three countries, Oman was the first to report this mutation from a strain (hCoV-19/Oman/11374/2020) collected on April 7, 2020 among the V cluster that was observed in Al Batinah South Governorate. The mutation was later reported in UK in a strain collected in April 2020, B.1. lineages (G), hCoV-19/England/CAMB-1AE373/2020 (CVR Bioinformatics, Glasgow, UK). Unfortunately, the origin of this mutation is not known as no travel history could be linked to the index case of this cluster. This unique mutation is worth further investigation for its biological effect on the pathogenesis of the virus. It is of grave importance to know the level of impact novel variants have on disease severity (Hodcroft et al., 2020).

Phylogenetic analysis of all SARS-CoV-2 Oman sequences deposited in GISAID and compared with a subset of sequences obtained from other countries revealed the presence of several lineages in the country initially. This lined up with the multiple introductions due to travel and before the travel restrictions on March 29. Similar findings were observed in UAE (Tayoun et al., 2020; Alandaljay et al., 2020). Early cases detected in Oman in February 2020 with a travel history from Iran belonging to B.4 (O) (Al-Mahruqi et al., unpublished data) were also closely related to cases from UAE, Lebanon, Bahrain, Pakistan, and Kuwait. However, this clade was absent in the selected clusters of this study. Similar findings were observed from sequences in other clades, except for the dominant clade B.1.1 and B.1.1.27 lineage (GR clade) that caused major outbreaks in the country. The B1.1 lineage (GR) comprising both Spike D614G and nucleocapsid RG203KR mutations, was the major clade found in Oman, which is inconsistent with the current globally predominant clade in Europe, Asia, South America, and Africa.

In general, and according to data deposited to GISAID, countries tend to resemble the clades of their continents, with a few exceptions. In China, the L clade (original) still dominates but other clades were obviously introduced after the reopening of the country (Mercatelli and Giorgi, 2020).

Currently and globally, the predominant clade worldwide is G and its offspring GH and GR (74% of world sequenced genomes), which vary remarkably within continents, and this variation is well pronounced in Europe.

Interestingly, GH is most prevalent in USA while GR is dominant in Europe and the most common worldwide. Regionally, in KSA, GH has the highest prevalence, which is possibly due to related travel history with USA. Not unexpectedly, UAE has a blend of all clades, but GR is dominant. In contrast to KSA, GR is dominating in Oman, which could be explained by related travel history with UK where GR is prevalent (Mercatelli and Giorgi, 2020).

According to GISAID, scarce published data are available on genetic characteristics of SARS-CoV-2 in the GCC region. Reviewing
these sequences from KSA (419), UAE (148), Bahrain (27), Qatar (16), Iran (18), and Kuwait (8) during the same period of our study revealed that GR clade was predominant in Oman and UAE, while GH clade was dominant in KSA. However, the mutation G715S is only present in Oman. The genome sequencing of SARS-CoV-2 has proceeded worldwide at an extraordinary rate with numerous published reports following the first published genome from Wuhan. This fact, however, does not prove true in the MENA region where the literature in this area is scarce. It is noteworthy mentioning that there are a few reports from UAE (49 isolates), Egypt (2 isolates), and Iran (7 isolates). However, in addition to the small number of samples used, these studies are based on the phylogenetic analysis only and no association with epidemiological surveillance was performed (Tabibzadeh et al., 2020; Tayoun et al., 2020; Kandeil et al., 2020).

From a timeline perspective, the current pandemic started with an original strain (L) then mutated in early 2020 to clade S and to a lesser extent, clade O. About two weeks later, clade V appeared that mutated in NSP6 and ORF3a. Then clade G appeared at the end of January 2020 and the first appearance of its subclades, GR and GH (mutated in Spike D614G, ORF3a, and Q57H), emerged about three weeks later (February 20, 2020). Since then, clade G and its derivatives have become the most dominant.

While infectiousness and transmissibility are closely related, they are not necessarily synonymous to one another, and therefore, detailed studies are needed to decide whether or not the D614G mutation has contributed to an increase in the number of infections and not simply higher viral loads during infection (Volf et al., 2021 and COVID-19 Genomics UK Consortium, 2020a). Interestingly the entire region is sharing the same common mutations within the same clades (COVID-19 Genomics UK Consortium, 2020b). As the pandemic is ongoing, so is the likely rise to further mutations and possible exhibition of phenotypic changes; our ability to assess and trace these variations will aid in apt and timely response measures.

To the best of our knowledge, this is the first large-scale molecular epidemiology study of COVID-19 in the MENA region, and the whole-genome sequencing of 94 samples of SARS-CoV-2 was coupled with epidemiological surveillance of the early transmission of COVID-19 in Oman. This has allowed us to link the surveillance information with the genetic analysis of the virus and enabled genotype tracking and identifying the mutations present in circulating strains. In addition, these results provide baseline information to which future genomes can be compared to study evolution.

The main limitation of our study is that it is retrospective and limited our ability in selecting certain epidemiological features for the genotyped samples such as travel-related and severe admissions. However, with the excess sample selected, we managed to involve sufficient samples to cover adequate epidemiological variations in our genotyped sample. This work needs to continue to get accumulated sequencing data and genomic analysis across the spectrum of the cases in the country for understanding diversity, future mutation, and fine-tuning primers used for the diagnosis based on local strains. The study confirms through genetic analysis the good quality of surveillance systems that prove to be robust during the early pandemic.

**Conclusion**

Oman’s robust surveillance system was very efficient in guiding outbreak investigation processes in the country, and illustrates the future importance of molecular epidemiology in guiding the national response to outbreaks and pandemics. Our work adds context to the global data and international genomic consortiums.

**Authors’ contributions**

Samira Al-Mahruqi, Amina Al-Jardani, Hanan Al-Kindi, Samiha Al-Kharusi, Intisar Al-Shukri and Aisha Al-Busaidi conducted the lab work and wrote the draft of the manuscript. Adil Al Wahabi conducted the epidemiological work and the sampling methodology for the clusters and wrote the draft manuscript. Sajjad Asaf, Ahmed N. Al-Rawahi, Ahmed Al-Rawahi, Abdul Latif Khan, Majid Al-Salmani conducted the WGS, the bioinformatics work, and the writing of the draft of the manuscript. Ahmed Al-Harassi and Seif Al-Abri supervised the study and participated in all stages of the manuscript.

**Conflict of interest**

The authors declare that they have no conflict of interest.

**Funding**

The authors would like to thank the University of Nizwa for funding this project through its internal research grant (IG/01-20-UoN/01/NMSRC).

**Ethical approval**

The study was approved by the Directorate General for Disease Surveillance and Control, and there was no need for patients’ consent as the study was anonymous and used the data produced for public health purposes.
| Sr# | Authors | No of strains | Country | Originating Laboratory | Accession ID |
|-----|---------|--------------|---------|------------------------|--------------|
| 10  | Md. Murshed Hasan Sarkar et al | 1 | Bangladesh | National Institute of Laboratory Medicine and Referral Center | EPI_ISL_498800 |
| 11  | Senjuti Saha et al | 1 | Bangladesh | Child Health Research Foundation | EPI_ISL_477127 |
| 12  | Siyuan Yang et al | 4 | Beijing | Laboratory of Infectious Diseases Center of Beijing Ditan Hospital | EPI_ISL_452344, EPI_ISL_452333, EPI_ISL_452345, EPI_ISL_455693 |
| 13  | Lin Qi et al | 1 | Fujian | Fujian Center for Disease Control and Prevention | EPI_ISL_431783 |
| 14  | Shengyue Wang et al | 1 | Shanghai | Shanghai Public Health Clinical Center, Shanghai Medical College, Fudan University | EPI_ISL_416382 |
| 15  | Gao Q et al | 2 | Zhejiang | Department of Microbiology | EPI_ISL_455689, EPI_ISL_455688 |
| 16  | Rita Feghali et al | 2 | Lebanon | Rafik Hariri University Hospital | EPI_ISL_450512, EPI_ISL_450511 |
| 17  | Abi Habib, W et al | 1 | Lebanon | Lebanese American University | EPI_ISL_498551 |
| 18  | Issa Abu-Dayyah et al | 6 | Jordan | Biolab Diagnostic Laboratories | EPI_ISL_429998, EPI_ISL_430001, EPI_ISL_429997, EPI_ISL_450087, EPI_ISL_429994, EPI_ISL_429992 |
| 19  | Mohd Noor Mat Isa et al | 2 | Malaysia | National Public Health Laboratory | EPI_ISL_416885, EPI_ISL_528739 |
| 20  | Yoong Min Chong et al | 3 | Malaysia | Department of Medical Microbiology, University Malaya Medical Centre | EPI_ISL_501226, EPI_ISL_501206, EPI_ISL_501176 |
| 21  | Hajar Fauzan Ahmad et al | 1 | Malaysia | Microbiology Unit, Department of Pathology & Laboratory Medicine, IJUM Medical Centre | EPI_ISL_455313 |
| 22  | Suppiah J et al | 1 | Malaysia | Institute for Medical Research, Infectious Disease Research Centre, National Institutes of Health, Ministry of Health Malaysia | EPI_ISL_490089 |
| 23  | Mak TM et al | 12 | Singapore | National Public Health Laboratory, National Centre for Infectious Diseases | EPI_ISL_462415, EPI_ISL_443231, EPI_ISL_462350, EPI_ISL_428827, EPI_ISL_527370, EPI(ISL 462280, EPI(ISL 443233, EPI(ISL 435698, EPI(ISL 443240, EPI(ISL 462363, EPI(ISL 443223, EPI(ISL 475968 |
| 24  | Danielle E Anderson et al | 1 | Singapore | National Centre for Infectious Diseases | EPI_ISL_420104 |
| 25  | Chen YC et al | 1 | Singapore | Department of Laboratory Medicine Tan Tock Seng Hospital | EPI_ISL_492597 |
| 26  | Pilailuk, Okada et al | 3 | Thailand | Ramathibodi Hospital | EPI_ISL_451648, EPI_ISL_447921, EPI_ISL_430842 |
| 27  | Elizabeth Batty et al | 5 | Thailand | Ramathibodi Hospital | EPI_ISL_430075, EPI_ISL_447011, EPI(ISL 512861, EPI_ISL_438024, EPI_ISL_455915, EPI(ISL 455934 |
| 28  | Rodpan A et al | 2 | Thailand | Faculty of Medicine | EPI_ISL_437611, EPI_ISL_437624 |
| 29  | Samira Al-Maruqi et al | 30 | Oman | Oman-NIC | EPI_ISL_457701, EPI_ISL_457704, EPI_ISL_457706, EPI(ISL 457937, EPI(ISL 457938, EPI(ISL 457939, EPI(ISL 457974, EPI(ISL 457975, EPI(ISL 457976, EPI(ISL 457977, EPI(ISL 457978, EPI(ISL 457979, EPI(ISL 457980, EPI(ISL 491116, EPI(ISL 457987, EPI(ISL 457985, EPI(ISL 457986, EPI(ISL 457988, EPI(ISL 457989, EPI(ISL 457990, EPI(ISL 457991, EPI(ISL 457992, EPI(ISL 457993, EPI(ISL 457994, EPI(ISL 457995, EPI(ISL 457996, EPI(ISL 457997, EPI(ISL 457998, EPI(ISL 492065 |
| 30  | Fahad Zadjali et al | 68 | Oman | Oman-NIC | EPI_ISL_457702, EPI_ISL_457703, EPI_ISL_457705, EPI(ISL 457707, EPI(ISL 457707, EPI(ISL 457708, EPI(ISL 457981, EPI(ISL 457982, EPI(ISL 457983, EPI(ISL 457984, EPI(ISL 458116, EPI(ISL 458117, EPI(ISL 458118, EPI(ISL 458119, EPI(ISL 458120, EPI(ISL 458121, EPI(ISL 458122, EPI(ISL 458123, EPI(ISL 458124, EPI(ISL 458125, EPI(ISL 458126, EPI(ISL 458127, EPI(ISL 458128, EPI(ISL 458127, EPI(ISL 491668, EPI(ISL 491686, EPI(ISL 491697, EPI(ISL 491698, EPI(ISL 491698, EPI(ISL 491699, EPI(ISL 491698, EPI(ISL 491699, EPI(ISL 491698, EPI(ISL 491699, EPI(ISL 491699, EPI(ISL 491698, EPI(ISL 491698, EPI(ISL 491697, EPI(ISL 491697, EPI(ISL 491692, EPI(ISL 491693, EPI(ISL 491694, EPI(ISL 491695, EPI(ISL 491696, EPI(ISL 491697, EPI(ISL 491698, EPI(ISL 491699, EPI(ISL 492000, EPI(ISL 492001, EPI(ISL 492002, EPI(ISL 492003, EPI(ISL 492004, EPI(ISL 492005, EPI(ISL 492006, EPI(ISL 492007, EPI(ISL 492008, EPI(ISL 492009, EPI(ISL 492010, EPI(ISL 492011, EPI(ISL 492012, EPI(ISL 492013, EPI(ISL 492014, EPI(ISL 492015, EPI(ISL 492016, EPI(ISL 492017, EPI(ISL 492018, EPI(ISL 492019, EPI(ISL 492020,
31 Bas Oude 6 Bahrain Kingdom of Bahrain Ministry of Health EPI_ISL_492021, EPI_ISL_492022, EPI_ISL_492023, EPI_ISL_492024, EPI_ISL_492025, EPI_ISL_492026, EPI_ISL_483543, EPI_ISL_483546, EPI_ISL_483550, EPI_ISL_483560, EPI_ISL_483561, EPI_ISL_483639, EPI_ISL_486889

32 AlHafaj Z et al 1 Bahrain National Influenza Center, Bahrain EPI_ISL_510351, EPI_ISL_423449

33 Zaatad et al 1 Bahrain Communicable Disease Laboratory, Public Health Directorate EPI_ISL_487273

34 AlWastah H. and AlHafaj Z. 1 Bahrain Communicable Disease Laboratory, Public Health Directorate EPI_ISL_510528

35 Shehadi F 1 Bahrain Communicable Disease Laboratory, Public Health Directorate EPI_ISL_471444

36 Zeinali S et al 1 Iran Human Genetic Research Center, Kiasar Biotech Company EPI_ISL_307007, EPI_ISL_514753, EPI_ISL_442044, EPI_ISL_416543, EPI_ISL_421652, EPI_ISL_422426, EPI_ISL_422427

37 Shahzad Z et al 3 Iran Yafatabad Hospital, Covid Lab Center EPI_ISL_416543, EPI_ISL_421652, EPI_ISL_422426, EPI_ISL_422427

38 Fahad AlMulla et al 4 Kuwait Dasman Diabetes Institute EPI_ISL_451958

39 Javed A et al 1 Pakistan Department of Healthcare Biotechnology, National University of Sciences and Technology (NUST) EPI_ISL_513925

40 Shakeel M et al 1 Pakistan Jamil-ur-Rahman Center for Genome Research, Dr. Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi EPI_ISL_468163

41 Massab Umair et al 1 Pakistan Department of Virology, Public Health Laboratories Division, National Institute of Health EPI_ISL_513925

42 Khan M.T et al 1 Pakistan Microbiology & Bioinformatics, Kohat University of Science and Technology (Pakistan) & Shanghai Jiao Tong University (China) EPI_ISL_463033, EPI_ISL_455761

43 Sunil Raghav et al 2 India Institute of Life Sciences, Bhubaneswar EPI_ISL_436434, EPI_ISL_435111, EPI_ISL_435110

44 Pramod Kumar et al 3 India National Centre for Disease control (NCDC) EPI_ISL_528383

45 Saurabh Kumar et al 1 India Translational Health Science and Technology Institute -ESIC medical college and hospital, Faridabad EPI_ISL_512475

46 Dinesh Kumar et al 1 India GMERS Medical College and Hospital, Dharapur, Patan EPI_ISL_512475

47 Arindam Maity et al 1 India Indian Institute of Science EPI_ISL_508310

48 Amit Kumar et al 2 Saudi Arabia Pathogen Genomics Lab King Abdullah University of Science and Technology(KAUST) EPI_ISL_513248, EPI_ISL_513241

49 Sharif Hala et al 2 Saudi Arabia Pathogen Genomics Lab King Abdullah University of Science and Technology(KAUST) EPI_ISL_437463, EPI_ISL_437756

50 Majed Alghoribi et al 1 Saudi Arabia Clinical Microbiology Lab EPI_ISL_416432

51 Sara Mfarraj et al 1 Saudi Arabia Pathogen Genomics Lab King Abdullah University of Science and Technology(KAUST) EPI_ISL_512978

52 Afrah Alsomali et al 1 Saudi Arabia Pathogen Genomics Lab King Abdullah University of Science and Technology(KAUST) EPI_ISL_513257

53 Ahmad Abou Tayoun et al 1 UAE Mohammed Bin Rashid University of Medicine and Health Sciences EPI_ISL_520695, EPI_ISL_463740, EPI_ISL_520718, EPI_ISL_520681, EPI_ISL_520708, EPI_ISL_520738

54 Ortwin Adams et al 2 Germany Center of Medical Microbiology, Virology, and Hospital Hygiene, University of Dusseldorf EPI_ISL_425123, EPI_ISL_425130

55 Max Muenchhoff et al 3 Germany Max von Pettenkofer Institute, Virology, National Reference Center for Retroviruses, LMU Munchen EPI_ISL_466905, EPI_ISL_451940, EPI_ISL_437284

56 Peter Bauer and Krishna Kumar Kandaswamy 1 Germany Centogene AG EPI_ISL_459964, EPI_ISL_459962

57 Lorusso A et al 1 Italy Ospedale "Ss. Annunziata" EPI_ISL_529014

58 Antonio Mori et al 1 Italy IRCCS Sacro Cuore Don Calabria Hospital, Department of Infectious, Tropical Diseases & Microbiology EPI_ISL_492985

59 Paola Stefanelli et al 1 Italy Department of Infectious Diseases, Istituto Superiore di SanitA, Rome, Italy EPI_ISL_412974

60 Guillermo Martin Gutierrez et al 1 Spain Servicio de Microbiologia y Parasitologia clinica. UCEIMP, Hospital Universitario Virgen del Rocio(IBIS/CSIC/UAS) EPI_ISL_452570

61 Salud Rodriguez-Pallares et al 2 Spain Hospital Universitario Puerta del Mar de Cádiz - INIBICA EPI_ISL_467071, EPI_ISL_467080

62 Gustavo Cilla et al 1 Spain Servicio de Microbiologia. Hospital Universitario Donostia. OSI Donostialdea. Área de Enfermedades Infecciosas, Grupo EPI_ISL_442688

63 Antonio Rezusta Lopez et al 1 Spain Servicio de Microbiologia, Hospital Miguel Servet, Zaragoza EPI_ISL_468780

64 McHugh M et al 1 Schotland Virology Department, Royal Infirmary of Edinburgh, NHS Lothian | School of Biological Sciences, University of Edinburgh / Institute of Genetics and Molecular Medicine, University of Edinburgh EPI_ISL_426007

65 Sam Haldenby et al 1 England Liverpool Clinical Laboratories EPI_ISL_517180

66 Luke W Meredith et al 3 England Department of Pathology, University of Cambridge EPI_ISL_441294, EPI_ISL_442841, EPI_ISL_441294

67 Claire McMurray et al 2 England University of Birmingham EPI_ISL_529403, EPI_ISL_529603, EPI_ISL_529603
Acknowledgments

We would like to thank Dr Saqib Bilal and Mr Adil Khan for their experimental help in library preparation and sequencing. We also would like to thank Mr Ali Al-Abri for preparing the graphs and tables. We express our gratitude to the collective endeavor of the GISAID consortium and to the thousands of contributing laboratories that made comparative genomics possible. We would like to thank the Oman Animal and Plant Genetic Resource Center (OAPGRC) for their support of NGS. We also acknowledge Lesley Carson for her help in the English editing and proofreading of the manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ijijid.2020.12.049.

References

Al Wahabi A, Al Manji A, Al Maani A, Al Rawahi B, Al Harthy K, Abyaqobi F, et al. COVID-19 epidemic monitoring after non-pharmaceutical interventions: the use of time-varying reproduction number in a country with a large migrant population. Int J Infect Dis 2020;99:466–72, doi: http://dx.doi.org/10.1016/j.ijid.2020.08.039.

Alandijany TA, Faizo AA, Azhar EI. Coronavirus disease of 2019 (COVID-19) in the Gulf Cooperation Council (GCC) countries: current status and management practices. J Infect Public Health 2020;13(6):839–42, doi: http://dx.doi.org/10.1016/j.jiph.2020.05.020.

Al-Mahruqi, et al., unpublished data.

COVID-19 Genomics UK Consortium. CoV-GLUE enabled by data from GISAID. 2020. [Accessed 13 November 2020] http://cov-glue.cov-gis.aic.edu/.

COVID-19 Genomics UK Consortium. Report #9-25th June 2020. [https://www.gisaid.org/]. [6-11-2020] https://www.cogconsortium.uk/wp-content/uploads/2020/07/2.

Fehr AR, Perlman S. Coronavirus: an overview of their replication and pathogenesis. In: Maier HJ, Bickerton E, Britton P, editors. Coronavirus: methods and protocols. New York: Springer; 2015. p. 1–23.

GISAID.GISAID_IcoV-19_Analysis_Update_2020-10-23_1500UTC.pdf. 2020. [Accessed 13 November 2020] https://www.yunzhan365.com/basic/92100557.html.
Tayoun AA, Loney T, Khansaheb H, Ramaswamy S, Harilal D, Deesi ZO, et al. Multiple early introductions of SARS-CoV-2 into a global travel hub in the Middle East. Sci Rep 2020;10:17720, doi:http://dx.doi.org/10.1038/s41598-020-74666-w.

Volz E, Hill V, McCrone JT, Price A, Jorgensen D, O’Toole A, et al. Evaluating the effects of SARS-CoV-2 Spike mutation D614G on transmissibility and pathogenicity. Cell 2021;184(1):64–75, doi:http://dx.doi.org/10.1016/j.cell.2020.11.020 In this issue.

Wang R, Chen J, Gao K, Hozumi Y, Yin C, Wei GW. Characterizing SARS-CoV-2 mutations in the United States, Version 1. Res Sq Preprint 2020;[August], doi: http://dx.doi.org/10.21203/rs.3.rs-49671/v1.

WHO. WHO coronavirus disease (COVID-19) dashboard. 2020. [Accessed 9 December 2020] https://covid19.who.int/.

Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, et al. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. Cell Host Microbe 2020;27(3):325–8, doi:http://dx.doi.org/10.1016/j.chom.2020.02.001.