Received: 4 January 2021 | Revised: 22 January 2021 | Accepted: 23 January 2021

DOI: 10.1111/tbed.14004

ORIGINAL ARTICLE

Transboundary and Emerging Diseases WILEY

Genotype-phenotype correlation identified a novel SARS-CoV-2 variant possibly linked to severe disease

Tom Loney[1] | Hamda Khansaheb[2] | Sathishkumar Ramaswamy[3] | Divinlal Harilal[3] | Zulfa Omar Deesi[4] | Rupa Murthy Varghese[4] | Aydah Belal Al Ali[4] | Anees Khadeeja[4] | Hanan Al Suwaidi[1] | Abdulmajeed Alkhajeh[2] | Laila Mohamed AlDabal[5] | Mohammed Uddin[1] | Mubarak Al Faresi[6] | Madhvi Joshi[7] | Abiola Senok[1] | Norbert Nowotny[1,8] | Alawi Alsheikh-Ali[1] | Ahmad Abou Tayoun[1,3]

[1]College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates

[2]Medical Education & Research Department, Dubai Health Authority, Dubai, United Arab Emirates

[3]Al Jalila Genomics Center, Al Jalila Children's Hospital, Dubai, United Arab Emirates

[4]Microbiology and Infection Control Unit, Pathology and Genetics Department, Latifa Women and Children Hospital, Dubai Health Authority, Dubai, United Arab Emirates

[5]Medical Affairs Department, Rashid Hospital, Dubai Health Authority, Dubai, United Arab Emirates

[6]Microbiology Department, Sheikh Khalifa General Hospital, Umm Al Quwain, United Arab Emirates

[7]Gujarat Biotechnology Research Centre, Gandhinagar, India

[8]Institute of Virology, University of Veterinary Medicine Vienna, Vienna, Austria

Correspondence
Ahmad Abou Tayoun, Al Jalila Genomics Center, Al Jalila Children's Hospital, Dubai, United Arab Emirates.
Email: Ahmad.Tayoun@ajch.ae

Alawi Alsheikh-Ali, College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Building 14, Dubai Healthcare City, Dubai, United Arab Emirates.
Email: Alawi.Alsheikhali@mbru.ac.ae

Funding
This work was supported by internal funds from the College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences (to AAT, grant no. MBRU-CM-RG2020-04).

Abstract

The geographic location and heterogeneous multi-ethnic population of Dubai (United Arab Emirates; UAE) provide a unique setting to explore the global molecular epidemiology of SARS-CoV-2 and relationship between different viral strains and disease severity. We systematically selected (i.e. every 100th individual in the central Dubai COVID-19 database) 256 patients by age, sex, disease severity and month to provide a representative sample of laboratory-confirmed COVID-19 patients (nasopharyngeal swab PCR positive) during the first wave of the UAE outbreak (January to June 2020). Sociodemographic and clinical data were extracted from medical records and full SARS-CoV-2 genome sequences extracted from nasopharyngeal swabs were analysed. Older age was significantly associated with COVID-19-associated hospital admission and mortality. Overweight/obese or diabetic patients were 3–4 times more likely to be admitted to hospital and intensive care unit (ICU). Sequencing data showed multiple independent viral introductions into the UAE from Europe, Iran and Asia (29 January–18 March), and these early strains seeded significant clustering consistent with almost exclusive community-based transmission between April and June 2020. Majority of sequenced strains ($N = 60$, 52%) were from the European cluster consistent with the higher infectivity rates associated with the D614G mutation carried by most strains in this cluster. A total of 986 mutations were identified in 115 genomes, 272 were unique (majority were missense, $n = 134$) and 20/272 mutations were novel. A missense (Q271R) and synonymous (R41R) mutation in the S and N proteins, respectively, were identified in 2/27 patients with severe COVID-19 but not in patients with mild or moderate disease (0/86; $p = .05$, Fisher's Exact Test). Both patients were women (51–64 years) with no significant underlying health conditions. The same two mutations were identified in a healthy 37-year-old Indian man who was hospitalized in India due to COVID-19. Our findings provide evidence for continued community-based transmission of the European strains in the Dubai population and highlight new mutations that might be associated with severe disease in otherwise healthy adults.

KEYWORDS
COVID-19, molecular phylogeny, mutation, Q271R, SARS-CoV-2, whole genome sequencing

1 | INTRODUCTION

Since December 2019, the emergence of a novel virus later named SARS-CoV-2 causing respiratory disease (called COVID-19) has spread to 192 countries and infected more than 102 million people causing 2.2 million deaths (Johns Hopkins Center for Systems Sciences & Engineering, 2020; Uddin et al., 2020). All countries have upscaled their testing capabilities to identify infected individuals for isolation and contact tracing using the gold-standard reverse transcriptase quantitative PCR (RT-qPCR) to detect SARS-CoV-2 RNA in the nasopharyngeal swabs of suspected cases (Harilal et al., 2020). However, RT-qPCR cannot be used to determine which viral strains are circulating in the population, or the relationship between viral strain and disease severity. As such, it is vital for public health policy decisions to know the origin of, and predominant viral strains circulating in a population, and the association between these viral strains and clinical outcomes (Harilal et al., 2020). SARS-CoV-2 whole genome sequencing (cWGS) can be used to delineate the geographic origins and evolution of a virus through phylogenetic analysis (Butler et al., 2020; Tayoun et al., 2020) in combination with large volumes of viral sequence data accumulated from all over the world (Hadfield et al., 2018). Previous genetic sequencing and phylogenetic sequencing studies have shown the introduction, transmission and evolution of the virus within and between countries (Alm et al., 2020; Jesus et al., 2020; Stefanelli et al., 2020; Yang et al., 2020). However, these studies have been limited by the recruitment of opportunistic and convenient samples and lacked the clinical data that is required to explore the association between viral strain and disease severity.

In the United Arab Emirates (UAE), Dubai is a cosmopolitan city that has the busiest airport in the world connecting the east with the west and has quickly become a popular tourist destination for travellers (Tayoun et al., 2020; Uddin et al., 2020). Moreover, Dubai is a global centre for trade, finance and healthcare attracting a diverse multi-ethnic population from countries all over the world (Loney et al., 2013). Indeed, our earlier work on the index and early UAE cases showed multiple spatio-temporal viral introductions into the UAE from tourists and residents travelling from Asia, Europe and the Middle East (Tayoun et al., 2020). Dubai's geographic location and unique heterogeneous population make it an ideal setting to explore the global molecular epidemiology of SARS-CoV-2 and the relationship between viral strain and disease severity. In this study, we utilize a representative sample of patients with COVID-19 from the Dubai population to explore the relationship between various sociodemographic, clinical and viral genomic factors and disease severity.

2 | METHODS

2.1 | Ethics approval

The Dubai Scientific Research Ethics Committee—Dubai Health Authority (approval number #DSREC-04/2020_02) approved this study. All patients treated at a healthcare facility in the UAE provide written consent for their de-identified data to be used for research, and this study was performed in accordance with the relevant laws and regulations that govern research in the UAE. As this study was part of a public health surveillance and outbreak investigation in the UAE, the requirement for informed consent, specifically for this study, was waived by the Research Ethics Committee.

2.2 | Study design, setting and participants

Sociodemographic and clinical data were extracted from the electronic medical records of 256 patients with laboratory-confirmed SARS-CoV-2 (nasopharyngeal swab PCR positive) from 29 January to 30 June 2020 using the WHO case report form. In order to provide a representative sample, participants were systematically selected from a database containing the details for all cases in Dubai during the study time period stratified by age (<40 years, ≥40 years), sex (male, female), disease severity (asymptomatic/mild, moderate, severe/critical) and stage of the pandemic (<01 April, ≥01 April). The systematic sampling frame involved one researcher selecting every 100th individual in the central Dubai COVID-19 database to check whether they satisfied the inclusion criteria for one of the stratified groups and if not, then the consecutive case was reviewed for eligibility. Due to a limited number of females in the moderate and severe/critical categories in Dubai, additional samples and accompanying data were selected from the major public hospital managing COVID-19 patients in the neighbouring emirate of Umm Al Quwain. Cases were categorized into groups based on disease severity: asymptomatic and mild cases with either no symptoms or mild non-life-threatening symptoms (e.g. dry cough, mild fever ≥ 38.0°C, respiratory rate < 30 breaths per minute, oxygen saturation > 93%); moderate cases with symptoms (e.g. breathlessness, persistent fever ≥ 38.0°C, signs of pneumonia/lower respiratory symptoms) requiring hospitalization and medical attention (e.g. supplementary oxygen therapy, intravenous fluids) and severe/critical cases with advanced disease and pneumonia (respiratory distress (respiratory rate > 30 breaths per minute adults, oxygen saturation < 93%, PaO_2 /FiO_2 300 mmHg, lung infiltrate > 50% of the lung fields in 24–48 hr) requiring admission to intensive care units and specialized life-support treatment (e.g. mechanical ventilation).

TABLE 1 Sociodemographic and health characteristics of the study patients ($N = 256$) stratified by disease severity

Variable	All Patients ($N = 256$)[b]	Stratified by Disease Severity		
		Asymptomatic/ Mild ($N = 98$)	Moderate ($N = 94$)	Severe/Critical ($N = 60$)
Age				
Mean (*SD*), years	40.3 (14.7)	35.4 (13.5)	40.9 (14.3)	47.2 (14.3)*
Distribution—no./total no. (%)				
0–14 years	4/253 (1.6%)	3/98 (3.1%)*	1/94 (1.1%)	0/60 (0.0%)
15–49 years	187/253 (73.9%)	81/98 (82.7%)*	66/94 (70.2%)	40/60 (66.7%)
50–64 years	47/253 (18.6%)	13/98 (13.3%)	21/94 (22.3%)*	12/60 (20.0%)
≥65 years	15/253 (5.9%)	1/98 (1.0%)	6/94 (6.4%)	8/60 (13.3%)*
Female Sex—no./total no. (%)	108/252 (42.9%)	35/98 (35.7%)	42/94 (44.7%)	31/60 (51.7%)
Body Mass (kg)—Mean (*SD*)	72.3 (16.0)	71.4 (18.8)	72.5 (15.7)	73.2 (12.5)
Height (cm) - Mean (*SD*)	165.1 (11.0)	164.8 (13.3)	165.8 (7.8)	164.7 (8.9)
Body mass index[a]				
Mean (*SD*), (kg/m^2)	27.68 (5.80)	26.70 (6.23)	28.51 (5.61)	28.73 (4.96)
Overweight (25.00–29.99 kg/m^2)—no./total no. (%)	53/144 (36.8%)	21/68 (30.9%)	15/41 (36.6%)	17/34 (50.0%)
Obesity class I (30.00–34.99 kg/m^2)—no./total no. (%)	24/144 (16.7%)	8/68 (11.8%)	10/41 (24.4%)	6/34 (17.6%)
Obesity class II (35.00–39.99 kg/m^2)—no./total no. (%)	11/144 (7.6%)	6/68 (8.8%)	2/41 (4.9%)	3/34 (8.8%)
Obesity class III (≥40.00kg/m^2)—no./total no. (%)	5/144 (3.5%)	2/68 (2.9%)	2/41 (4.9%)	1/34 (2.9%)
Current smoker—no./total no. (%)	2/122 (1.6%)	1/22 (4.5%)	1/70 (1.4%)	0/60 (0.0%)
Region of origin—no./total no. (%)				
Africa	11/256 (4.3%)	3/98 (3.1%)	8/94 (8.5%)	0/60 (0.0%)
Americas and caribbean	2/256 (0.8%)	1/98 (1.0%)	1/94 (1.1%)	0/60 (0.0%)
Australasia	2/256 (0.8%)	2/98 (2.0%)	0/94 (0.0%)	0/60 (0.0%)
East Asia	37/256 (14.5%)	12/98 (12.2%)	11/94 (11.7%)	14/60 (23.3%)**
Europe	18/256 (7.0%)	14/98 (14.3%)**	0/94 (0.0%)	2/60 (3.3%)
Middle East	64/256 (25.0%)	24/98 (24.5%)	26/94 (27.7%)	13/60 (21.7%)
South Asia	122/256 (47.7%)	42/98 (42.9%)	48/94 (51.1%)	31/60 (51.7%)
Residency status—no./total no. (%)				
UAE national	35/256 (13.7%)	5/98 (5.1%)	19/94 (20.2%)**	11/60 (18.3%)
Expatriate resident	203/256 (79.3%)	79/98 (80.6%)**	75/94 (79.8%)	45/60 (75.0%)
Tourist	18/256 (7.0%)	14/98 (14.3%)**	0/94 (0.0%)	4/60 (6.7%)
Exposure to source of transmission within past 14 days—no./total no. (%)				
Travel history	47/255 (18.4%)	37/98 (37.8%)*	2/94 (2.1%)	5/60 (8.3%)
Contact with positive case	40/256 (15.6%)	26/98 (26.5%)**	10/94 (10.6%)	4/60 (6.7%)
Healthcare worker	9/250 (3.6%)	7/98 (7.4%)*	2/94 (2.1%)	0/60 (0.0%)
Laboratory worker	0/250 (0.0%)	0/98 (0.0%)	0/94 (0.0%)	0/60 (0.0%)
Pregnant—no./total no. (%)	6/108 (5.6%)	3/98 (3.1%)	0/94 (0.0%)	3/60 (5.0%)
Blood Type—no./total no. (%)				
A positive	24/112 (21.4%)	7/29 (24.1%)	8/43 (18.6%)	9/40 (22.5%)
A negative	3/112 (2.7%)	2/29 (6.9%)	0/43 (0.0%)	1/40 (2.5)
B positive	46/112 (41.1%)	7/29 (24.1%)	18/43 (41.9%)	21/40 (22.5%)
B negative	1/112 (0.9%)	0/29 (0.0%)	1/43 (2.3%)	0/60 (0.0%)
AB positive	3//112 (2.7%)	2/29 (6.9%)	0/43 (0.0%)	1/40 (2.5%)
AB negative	0/112 (0.0%)	0/29 (0.0%)	0/43 (0.0%)	0/60 (0.0%)
O positive	34//112 (30.4%)	11/29 (37.9%)	15/43 (34.9%)	8/40 (13.3%)

(Continues)

TABLE 1 (Continued)

Variable	All Patients ($N = 256$)[b]	Stratified by Disease Severity		
		Asymptomatic/ Mild ($N = 98$)	Moderate ($N = 94$)	Severe/Critical ($N = 60$)
O negative	1/112 (0.9%)	0/29 (0.0%)	1/43 (2.3%)	0/60 (0.0%)
Coexisting Disorder—no./total no. (%)				
Any	65/244 (26.6%)	14/93 (15.1%)	21/94 (22.3%)	30/59 (53.6%)**
Chronic cardiac disease (not Hypertension)	10/244 (4.1%)	1/93 (1.1%)	3/94 (3.2%)	6/56 (10.7%)*
Hypertension	38/245 (15.5%)	10/93 (10.8%)	10/94 (10.6)	18/57 (31.6%)**
Chronic pulmonary disease	1/245 (0.4%)	5/98 (5.1%)	0/94 (0.0%)	1/57 (1.8%)
Asthma	5/245 (5.0%)	2/93 (2.2%)	2/94 (2.1%)	1/57 (1.8%)
Chronic renal disease	4/245 (1.6%)	0/93 (0.0%)	0/94 (0.0%)	4/57 (7.0%)**
Chronic hepatic disease	11/256 (4.3%)	0/93 (0.0%)	0/94 (0.0%)	0/57 (0.0%)
Chronic neurological disorder	1/245 (0.4%)	0/93 (0.0%)	0/94 (0.0%)	1/57 (1.8%)
Diabetes mellitus	44/245 (18.0%)	7/93 (7.5%)	15/94 (15.0%)	22/57 (38.6%)**
Immunodeficiency	1/245 (0.4%)	0/93 (0.0%)	1/94 (1.1%)	0/57 (0.0%)
Malignant neoplasms	2/245 (0.8%)	0/93 (0.0%)	1/94 (1.1%)	1/57 (1.8%)

[a]Children (<18 years) excluded from this analysis.

[b]Some missing data for four patients, therefore, sum of disease severity categories $N = 252$.

*Denotes $p \leq .05$; **Denotes $p \leq .01$.

2.3 | SARS-COV-2 whole genome sequencing

All 256 COVID-19 patients tested positive for SARS-CoV-2 by RT-qPCR on RNA extracted from nasopharyngeal swabs using the QIAamp Viral RNA Mini or the EZ1 DSP Virus Kits (Qiagen). SARS-CoV-2 whole genome sequencing was performed using either shotgun transcriptome sequencing (Jain et al., 2021; Tayoun et al., 2020) or viral target enrichment and deep sequencing (Harilal et al., 2020). For shotgun transcriptomic sequencing, RNA libraries were prepared using the TruSeq Stranded Total RNA Library kit from Illumina, following manufacturer's instructions. Libraries were sequenced using the NovaSeq SP Reagent kit (2 × 150 cycles) from Illumina. For target enrichment, double stranded DNA was first synthesized using the QuantiTect Reverse Transcription Kit (Qiagen) and then amplified using 26 overlapping primer sets covering most of the SARS-CoV-2 genome as recently described by our group (Harilal et al., 2020). PCR products were then fragmented by ultra-sonication (Covaris LE220-plus series) and prepared for sequencing using the SureSelectXT Library Preparation kit (Agilent). The target enrichment libraries were sequenced using the Illumina MiSeq Reagent Kit V2 (2 X 150 cycles) or the NovaSeq SP Reagent kit (2 × 150 cycles).

2.4 | SARS-COV-2 genome assembly and bioinformatics analysis

The adapter sequences were trimmed using BCL2fastq v.2.20.0, and then, high-quality sequencing reads were aligned to the reference SARS-CoV-2 genome from Wuhan, China (GenBank accession number: NC_045512.2) using BWA v0.7.17. Consensus sequences were generated using GATK v3.8–1–0 and BCFtools v1.3.1 for samples where > 80% of nucleotide positions (56–29,797) had at least 10X coverage. A total of 115 UAE isolates met this coverage criterion (Table S1) and were submitted to the Global Initiative on Sharing All Influenza Data (GISAID) database under accession numbers listed in Table S2.

2.5 | Phylogenetic analysis

Multiple sequence alignment of UAE SARS-CoV-2 isolates ($n = 115$) was performed using MAFT v7.455 (Katoh et al., 2002) and phylogenetic tree was built using IQtree v1.6.12 with a bootstrap of 1,000 replicates (Chernomor et al., 2016). In addition, the evolutionary time scale was estimated by using BEAST v.1.10.4 (Suchard *et al*., 2018) with strict clock mode. The Hasegawa–Kishino–Yano substitution model was used with 4 rate categories drawn from a gamma distribution (GTR4G). Finally, we estimated the mutation rate from the resulting BEAST tree. The final tree was visualized by Figtree v1.4.4 (Rambaut, 2018).

2.6 | Statistical analysis

Continuous variables were expressed as means with standard deviations or medians and interquartile ranges or simple ranges (as appropriate) and categorical variables as counts and percentages. No imputations were made for missing data. One-way ANOVA's were used to determine differences between group means for continuous variables, while Pearson chi-square tests were used for categorical variables. Univariate and multivariate regression models were used to quantify the association between various sociodemographic and clinical characteristics and the

TABLE 2 Symptoms on admissions and clinical characteristics of the study patients ($N = 256$) stratified by disease severity

Variable	All Patients ($N = 256$)[a]	Stratified by Disease Severity		
		Asymptomatic/ Mild ($N = 98$)	Moderate ($N = 94$)	Severe/Critical ($N = 60$)
Signs and symptoms on admission—no./total no. (%)				
History of fever ≥ 38.5°C	131/247 (53.0%)	28/93 (30.1%)	55/94 (58.5%)	48/59 (81.4%)**
Dry cough	98/246 (39.8%)	1/93 (1.1%)	46/94 (48.9%)	23/58 (39.7%)
Cough with sputum	1/246 (0.4%)	0/93 (0.0%)	0/94 (0.0%)	1/58 (1.7%)
Cough with haemoptysis	94/246 (38.2%)	29/93 (31.2%)	43/94 (45.7%)	22/58 (37.9%)
Sore throat	18/246 (7.3%)	16/93 (17.2%)**	1/94 (1.1%)	1/58 (1.7%)
Rhinorrhoea	1/246 (0.4%)	0/93 (0.0%)	1/94 (1.1%)	0/58 (0.0%)
Wheezing	6/246 (2.4%)	0/93 (0.0%)	4/94 (4.3%)	2/58 (3.4%)
Chest pain	83/246 (33.7%)	15/93 (16.1%)	44/94 (46.8%)**	24/58 (41.4%)
Myalgia	5/246 (2.0%)	1/93 (1.1%)	3/94 (3.2%)	1/58 (1.7%)
Arthralgia	41/247 (16.6%)	4/93 (4.3%)	12/94 (12.8%)	25/59 (42.4%)**
Malaise	85/247 (34.4%)	3/93 (3.2%)	61/94 (64.9%)**	21/59 (35.6%)
Shortness of breath	2/247 (0.8%)	0/93 (0.0%)	0/94 (0.0%)	2/59 (3.4%)*
Inability to walk	1/247 (0.4%)	0/93 (0.0%)	0/94 (0.0%)	1/59 (1.7%)
Lower chest wall indrawing	23/247 (9.3%)	14/93 (15.1%)	5/94 (5.3%)	4/59 (6.8%)
Headache	3/247 (1.2%)	0/93 (0.0%)	0/94 (0.0%)	3/59 (5.1%)**
Altered consciousness/ Confusion	2/247 (0.8%)	0/93 (0.0%)	0/94 (0.0%)	2/59 (3.4%)*
Seizures	8/247 (3.2%)	1/93 (1.1%)	4/94 (4.3%)	2/59 (3.4%)
Abdominal pain	11/247 (4.5%)	3/93 (3.3%)	2/94 (2.1%)	5/59 (8.5%)
Vomiting/Nausea	16/247 (6.5%)	2/93 (2.2%)	4/94 (4.3%)	10/59 (16.9%)**
Diarrhoea	1/247 (0.4%)	0/93 (0.0%)	1/94 (1.1%)	0/59 (0.0%)
Lymphadenopathy	1/247 (0.4%)	0/93 (0.0%)	0/94 (0.0%)	1/59 (1.7%)
Intracerebral haemorrhage	1/247 (0.4%)	0/93 (0.0%)	0/94 (0.0%)	1/59 (1.7%)
Other symptoms	14/247 (5.7%)	1/93 (1.1%)	7/94 (7.4%)	6/59 (11.9%)
Hemiparesis	1/247 (0.4%)	0/93 (0.0%)	0/94 (0.0%)	1/59 (1.7%)
Ageusia/Anosmia	3/247 (1.2%)	1/93 (1.1%)	2/94 (2.1%)	0/59 (0.0%)
Urinary tract infection	1/247 (0.4%)	0/93 (0.0%)	1/94 (1.1%)	0/59 (0.0%)
Constipation	1/247 (0.4%)	0/93 (0.0%)	0/94 (0.0%)	1/59 (1.7%)
Oedema/Decreased urine output	3/247 (1.2%)	0/93 (0.0%)	1/94 (1.1%)	2/59 (3.4%)
Dizziness	1/247 (0.4%)	0/93 (0.0%)	1/94 (1.1%)	0/59 (0.0%)
Syncope	1/247 (0.4%)	0/93 (0.0%)	1/194 (1.1%)	0/59 (0.0%)
Palpitations	1/247 (0.4%)	0/93 (0.0%)	1/94 (1.1%)	0/59 (0.0%)
Lower Leg paraesthesia/Weakness	2/247 (0.8%)	0/93 (0.0%)	0/94 (0.0%)	2/59 (3.4%)
Temperature				
Mean (*SD*), °C	37.8 (0.9)	37.3 (0.7)	38.0 (0.9)	38.2 (1.0)**
Distribution—no./total no. (%)				
<37.5°C	97/245 (39.6%)	57/91 (62.6%)**	26/94 (27.7%)	14/59 (23.7%)
37.5–38.0°C	56/245 (22.9%)	18/91 (19.8%)	24/94 (25.5%)	14/59 (23.7%)
38.1–39.0°C	65/245 (26.5%)	15/91 (16.5%)	32/94 (34.0%)**	18/59 (30.5%)
>39.0°C	27/245 (11.0%)	1/91 (1.1%)	12/94 (12.8%)	13/59 (22.0%)**
Heart rate (beats/min), Mean (*SD*)	93 (19)	92 (17)	88 (19)	100 (21)**
Respiratory rate (breaths/min), Mean (*SD*)	20 (5)	18 (2)	21 (6)	22 (6)**

(Continues)

TABLE 2 (Continued)

Variable	All Patients ($N = 256$)[a]	Stratified by Disease Severity		
		Asymptomatic/ Mild ($N = 98$)	Moderate ($N = 94$)	Severe/Critical ($N = 60$)
Blood pressure (mmHg), Mean (*SD*)				
Systolic	129 (17)	129 (16)	125 (14)	138 (20)**
Diastolic	79 (11)	79 (12)	78 (10)	78 (12)
Oxygen saturation (%), Mean (*SD*)	97 (6)	99 (1)	97 (4)	94 (11)**
AVPU Scale—no./total no. (%)				
Alert	244/247 (98.8%)	93/93 (100.0%)	94/94 (100.0%)	56/59 (94.9%)
Verbal	1/247 (0.8%)	0/93 (0.0%)	0/94 (0.0%)	1/59 (1.7%)
Pain	2/247 (0.8%)	0/93 (0.0%)	0/94 (0.0%)	2/59 (3.4%)
Unresponsive	0/247 (0.0%)	0/93 (0.0%)	0/94 (0.0%)	0/59 (0.0%)
GCS—Median (Min, Max)	15 (3, 15)	15 (15, 15)	15 (15, 15)	15 (3, 15)
ICU admission - no./total no. (%)	60/247 (24.3%)	0/93 (0.0%)	1/94 (1.1%)	60/60 (100.0%)**
Supportive Care—no./total no. (%)				
Oxygen therapy	93/209 (44.5%)	0/93 (0.0%)	33/94 (35.1%)	60/60 (100.0%)**
Non-invasive ventilation	45/140 (32.1%)	0/93 (0.0%)	0/94 (0.0%)	45/60 (75.0%)**
Invasive ventilation	54/79 (68.4%)	0/93 (0.0%)	0/94 (0.0%)	54/60 (90.0%)**
Inotropes/Vasopressors	34/141 (24.1%)	0/93 (0.0%)	0/94 (0.0%)	34/60 (56.7%)**
Extracorporeal	2/141 (1.4%)	0/93 (0.0%)	0/94 (0.0%)	2/60 (3.3%)
Prone position	27/50 (54.0%)	0/93 (0.0%)	12/94 (12.8%)	15/60 (25.0%)**
Death—no./total no. (%)	18/256 (7.0%)	0/93 (0.0%)	0/94 (0.0%)	18/60 (30.0%)**

[a]Some missing data for four patients; therefore, sum of disease severity categories $N = 252$. No patients presented with conjunctivitis, skin rash, or skin ulcers.

*Denotes $p \leq .05$; **Denotes $p \leq .01$.

outcomes of severe/critical disease and COVID-19-related mortality, separately. Factors associated with severe/critical disease or COVID-19-related mortality in the crude model were added to the multivariate model. Age and body mass index were utilized as both continuous variable (years; kg/m^2) and categorical variable (≥40 years, age ≥ 50 years, age ≥ 60 years; ≥25 kg/m^2, ≥30 kg/m^2). Due to the heterogenous Dubai population, nationality groups were collapsed to form regional groups (i.e. European descent (reference group) was composed of North Americans, Europeans and Australasians; Middle Eastern and African; South Asian and East Asian). Patients with moderate and severe/critical disease were collapsed to form one group COVID-19-related hospital admission, and intensive care unit admission and COVID-19-associated mortality contain only patients with severe/critical disease. Emirates airlines suspended flights to and from 30 global destinations from 18 March 2020 and Dubai airport was closed to passenger flights on 25 March 2020 (11). Patients after 18 March 2020 were more likely a result of community transmission as opposed to imported infections. We used 01 April 2020 as a cut-off for predominantly community-based transmission as this date was seven days after Dubai International Airport closed. The variable community transmission was based on receiving a first positive test either < 01 April or ≥ 01 April 2020. Crude odds ratio (OR) and adjusted odds ratios (aOR) with 95% confidence intervals (CI) were reported. IBM SPSS Statistics for Windows (Version 25.0: IBM Corp) was used to perform statistical analyses. Statistical significance was defined by a p-value less than or equal to 0.05 and 95% confidence intervals.

3 | RESULTS

3.1 | Sociodemographic and clinical characteristics

We obtained the nasopharyngeal swabs and data regarding the sociodemographic characteristics and clinical symptoms and outcomes of 256 patients selected systematically by age, sex, disease severity and month. The sociodemographic and clinical characteristics of the patients are shown in Table 1.

The majority of patients were males (57.1%), mean (*SD*) age was 40.3 ± 14.7 years and mean (*SD*) BMI was 27.7 ± 5.8 kg/m^2. Mean age increased significantly with disease severity while there was a pattern for the proportion of males to increase with disease severity (Table 1). Mean body mass index increased with disease severity; however, this did not reach statistical significance ($p = .096$). Majority of patients were expatriate residents (79.3%) from South

TABLE 3 Univariate and multivariate logistic regression for factors associated with, COVID-related hospital admissions, intensive care unit admission and COVID-associated mortality ($N = 256$)

	COVID-Related Hospital Admission		ICU Admission		COVID-Associated Mortality	
	Univariate	Multivariate	Univariate	Multivariate	Univariate	Multivariate
Variable	COR (95% CI)	AOR (95% CI)	COR (95% CI)	AOR (95% CI)	COR (95% CI)	AOR (95% CI)
Sex (Male)	0.62 (0.37, 1.04)	–	0.63 (0.35, 1.12)	–	4.07 (1.15, 14.44)	3.10 (0.78, 12.37)
Age (years)	1.04 (1.02, 1.06)	1.04 (1.01, 1.06)	1.04 (1.02, 1.06)	1.03 (1.00, 1.06)	1.06 (1.02, 1.09)	1.05 (1.01, 1.10)
Age ≥ 40 Years	2.57 (1.50, 4.40)	2.10 (1.15, 3.82)	2.23 (1.24, 4.03)	1.37 (0.70, 2.70)	1.71 (0.65, 4.48)	–
Age ≥ 50 Years	2.64 (1.36, 5.11)	2.15 (0.97, 4.74)	1.84 (0.97, 3.49)	–	3.43 (1.30, 9.01)	2.26 (0.65, 7.85)
Age ≥ 60 Years	3.26 (1.29, 8.22)	1.71 (0.59, 4.94)	2.38 (1.10, 5.13)	1.33 (0.50, 3.53)	3.85 (1.34, 11.11)	2.61 (0.62, 11.09)
Community Transmission (≥01 April 2020)	4.87 (2.77, 8.58)	4.20 (2.15, 8.19)	1.27 (0.67, 2.39)	–	0.55 (0.21, 1.46)	–
UAE National (Yes)	4.50 (1.68, 12.04)	4.57 (1.62, 12.86)	1.57 (0.72, 3.43)	1.29 (0.54, 3.09)	1.29 (0.35, 4.70)	–
Middle Eastern & African[a]	9.86 (2.65, 36.76)	12.60 (2.58, 61.49)	1.92 (0.40, 9.30)	2.03 (0.23, 18.03)	–	–
South Asian[a]	10.66 (2.95, 38.46)	13.26 (2.82, 62.33)	3.10 (0.68, 14.13)	4.29 (0.52, 35.4)	–	–
East Asian[a]	11.81 (2.89, 48.22)	13.40 (2.55, 70.44)	5.48 (1.10, 27.27)	7.01 (0.79, 62.51)	–	–
BMI (kg/m^2)	1.06 (1.00, 1.13)	1.05 (0.97, 1.13)	1.04 (0.98, 1.11)	–	1.03 (0.93, 1.13)	–
BMI ≥ 25 kg/m^2 (Yes)	2.47 (1.23, 5.00)	3.00 (1.21, 7.47)	2.51 (1.01, 6.28)	4.02 (1.30, 12.38)	0.97 (0.27, 3.44)	–
BMI ≥ 30 kg/m^2 (Yes)	1.53 (0.73, 3.21)	–	1.10 (0.47, 2.57)	–	1.54 (0.43, 5.58)	–
Any Chronic Disease (Yes)	2.91 (1.50, 5.63)	1.59 (0.73, 3.44)	5.01 (2.64, 9.51)	3.26 (1.54, 6.92)	5.24 (1.82, 15.08)	2.53 (0.68, 9.38)
Chronic Cardiac Disease (Yes)	5.87 (0.73, 47.13)	–	5.49 (1.49, 20.21)	5.49 (0.40, 76.01)	12.33 (3.07, 49.62)	4.43 (0.72, 27.37)
Hypertension (Yes)	1.89 (0.87, 4.10)	–	3.85 (1.87, 7.97)	0.90 (0.35, 2.28)	6.63 (2.32, 19.00)	3.15 (0.83, 12.03)
Diabetes Mellitus (Yes)	3.99 (1.70, 9.38)	2.52 (0.95, 6.70)	4.71 (2.35, 9.44)	3.30 (1.48, 7.38)	5.36 (1.89, 15.21)	3.09 (0.84, 11.23)
Temperature (°C)					1.25 (0.74, 2.09)	–
Heart Rate (beats/min)					1.03 (1.00, 1.05)	1.02 (0.98, 1.05)
Respiratory rate (breaths/min)					1.07 (1.01, 1.34)	1.00 (0.90, 1.11)
Systolic Blood Pressure (mmHg)					1.04 (1.02, 1.07)	1.02 (0.99, 1.04)
Oxygen Saturation (%)					0.95 (0.90, 1.00)	0.98 (0.92, 1.04)

Abbreviations: AOR, adjusted odds ratio; BMI, body mass index; CI, confidence interval; COR, crude odds ratio; ICU, intensive care unit.

[a]Compared with reference group (Americas and Caribbean, Europe, Australasia). Multivariate analysis controlling for age (years), nationality, presence of any chronic disease and sex (only for COVID–associated mortality).

Asia (47.7%) and the Middle East (25.0%). There was a significantly greater proportion of Europeans and East Asians in the asymptomatic/mild and severe/critical groups, respectively.

Nearly a fifth (18.4%) of patients had a positive travel history, 15.6% had known contact with a positive case, and only 3.6% were healthcare workers. The most common blood type was B positive (41.1%) followed by O positive (30.4%) and A positive (21.4%). A quarter (26.6%) of patients had at least one coexisting health disorder and a significantly greater proportion of severe/critical patients (53.6%) had at least one underlying health issue compared with other disease severity categories. The most common coexisting disorders were diabetes mellitus (18.0%) and hypertension (15.5%) which were both significantly more prevalent in patients with severe disease (38.6% and 31.6%, respectively). The prevalence of chronic cardiac disease (10.7%) and chronic renal disease (7.0%) was significantly greater in severe patients compared with other groups.

Half (53.0%) of patients had a fever (≥38.0°C) on admission and more than a third had a dry cough (39.8%), cough with haemoptysis (38.2%), malaise (34.4%) or chest pain (33.7%) (Table 2). Fever (81.4%), arthralgia (42.4%), headache, vomiting/nausea (16.9%), lower chest wall drawing (5.1%), altered consciousness/confusion (3.4%) and shortness of breath (3.4%) had a significantly greater prevalence amongst severe/critical patients while malaise (64.9%) and chest pain (46.8%) were significantly more common symptoms amongst patients with moderate disease, and a sore throat (17.2%) had a significantly higher prevalence amongst asymptomatic/mild cases (Table 2).

Mean temperature, heart rate, respiratory rate and systolic blood pressure were significantly higher and mean oxygen saturation significantly lower in the severe/critical disease category. Nearly, half of all patients required oxygen therapy and the mortality ratio was 30.0% in the patients with severe disease (Table 2).

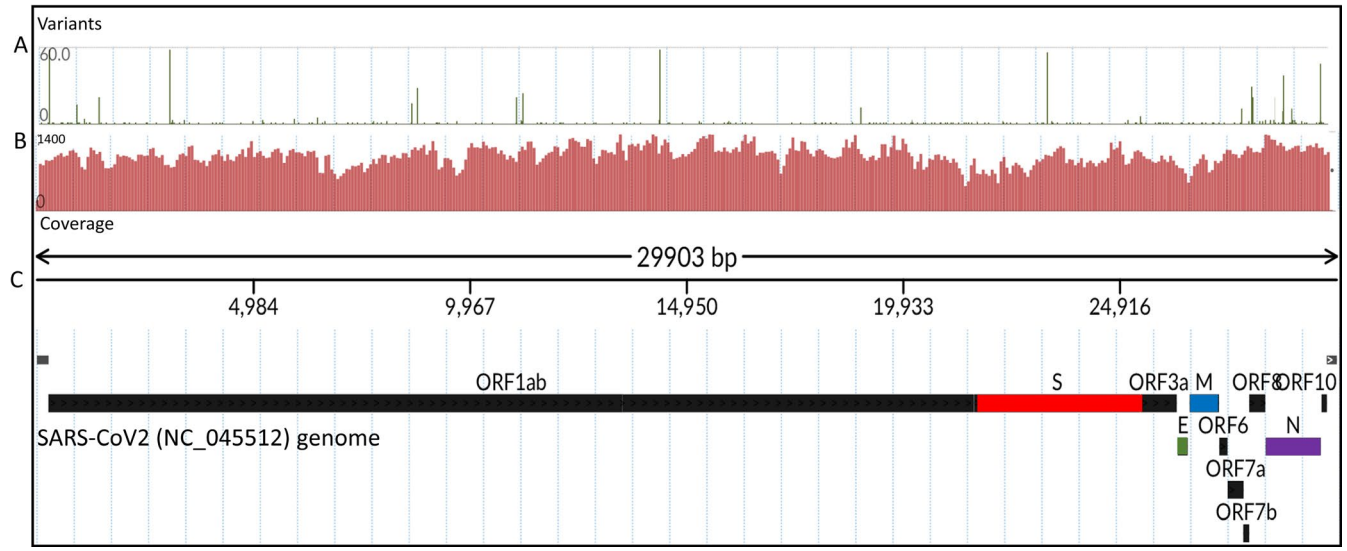

FIGURE 1 Distribution of UAE strains with a given variant (a) and the average sequencing coverage (b) across the SARS-CoV-2 genome sequence (NC_045512) (c). E = Envelope protein, *M* = Membrane, *N* = Nucleocapsid protein, ORF = open reading frame, S = Spike protein

3.2 | Associations between sociodemographic and clinical characteristics and hospital admission and mortality

Age was significantly associated with COVID-19-related hospital admission and COVID-19-associated mortality but not ICU admission, with patients aged 40 years and older twice as likely to be admitted to hospital compared with younger patients (Table 3). Compared with patients of European descent, patients from the Middle East and Africa, South Asia and East Asia regions were 12–13 times more likely to be admitted to hospital, but there was no association with ICU admission or mortality (Table 3). Overweight and obese patients (BMI ≥ 25.0 kg/m^2) were 3–4 times more likely to be admitted to hospital and ICU compared with healthy weight patients, and patients with diabetes mellitus were three times more likely to be admitted to ICU compared with patients without diabetes (Table 3). Patients that were more likely to have been infected from community transmission as opposed to imported infections were four times more likely to be admitted to hospital with either moderate or severe/critical disease (Table 3).

3.3 | SARS-COV-2 whole genome sequencing and mutation analysis

SARS-CoV-2 whole genome sequencing was performed using RNA extracted from nasopharyngeal samples of all 256 COVID-19 patients (Methods). A total of 115 genomes meeting our coverage criteria were assembled and used for phylogenetic analysis, mutational analysis and genotype-phenotype correlation. Overall, we were able to assemble full genomes from 115 patients with an average coverage of 890X where more than 97% of the SARS-CoV-2 genome had at least 10X coverage (Figure 1).

Of the sequenced genomes, 30 were obtained by shotgun transcriptome sequencing where, on average, 3% of the reads mapped to the SARS-CoV-2 genome with an average coverage of 1,756X and > 99% of the virus genome having at least 10X reads (Table S1). On the other hand, 85 full genomes were obtained by target enrichment followed by sequencing. On average, 96% of the reads mapped to the SARS-CoV-2 genome using this protocol with an average coverage of 582X where > 96% of the genome had ≥ 10X reads (Table S1).

In aggregate, a total of 272 unique mutations were identified in all 115 genomes relative to the reference SARS-CoV-2 Wuhan genome (GenBank Accession: NC_045512.2) (Figure 1) with an average of 8.6 (1–17) mutations per genome. Some mutations were more frequent than others; for example, the 241C>T, 3037C>T, 14408C>T, and the 23403A>G (also known as the D614G mutation in the S protein), were each seen in 58 of the genomes (Table S3). Of the 272 mutations, the majority of which were missense (*n* = 134) (Figure 2). More than half (162/272) were localized to the ORF1ab gene (Figure 2 and Table S3), while 29/272, 2/272, 10/272 and 27/272 were in the S, E, M and N genes, respectively. Notably, 20 out of the 272 mutations were novel as they were not identified in the Chinese National Center for Bioinformation Database (https://bigd.big.ac.cn/ncov/variation/annotation; last accessed January 03, 2021). The majority (11/20) were coding missense mutations including two (N657S and Q1036L) in the S protein (Table S3).

3.4 | Phylogenetic analysis

SARS-CoV-2 sequencing from patients during the early phase (29 January – 18 March) of the pandemic showed multiple independent introductions of the virus into the UAE. Specifically, early strains during that time period were genetically close to viruses from COVID-19 patients in Europe, Iran and Asia (3). Extending our

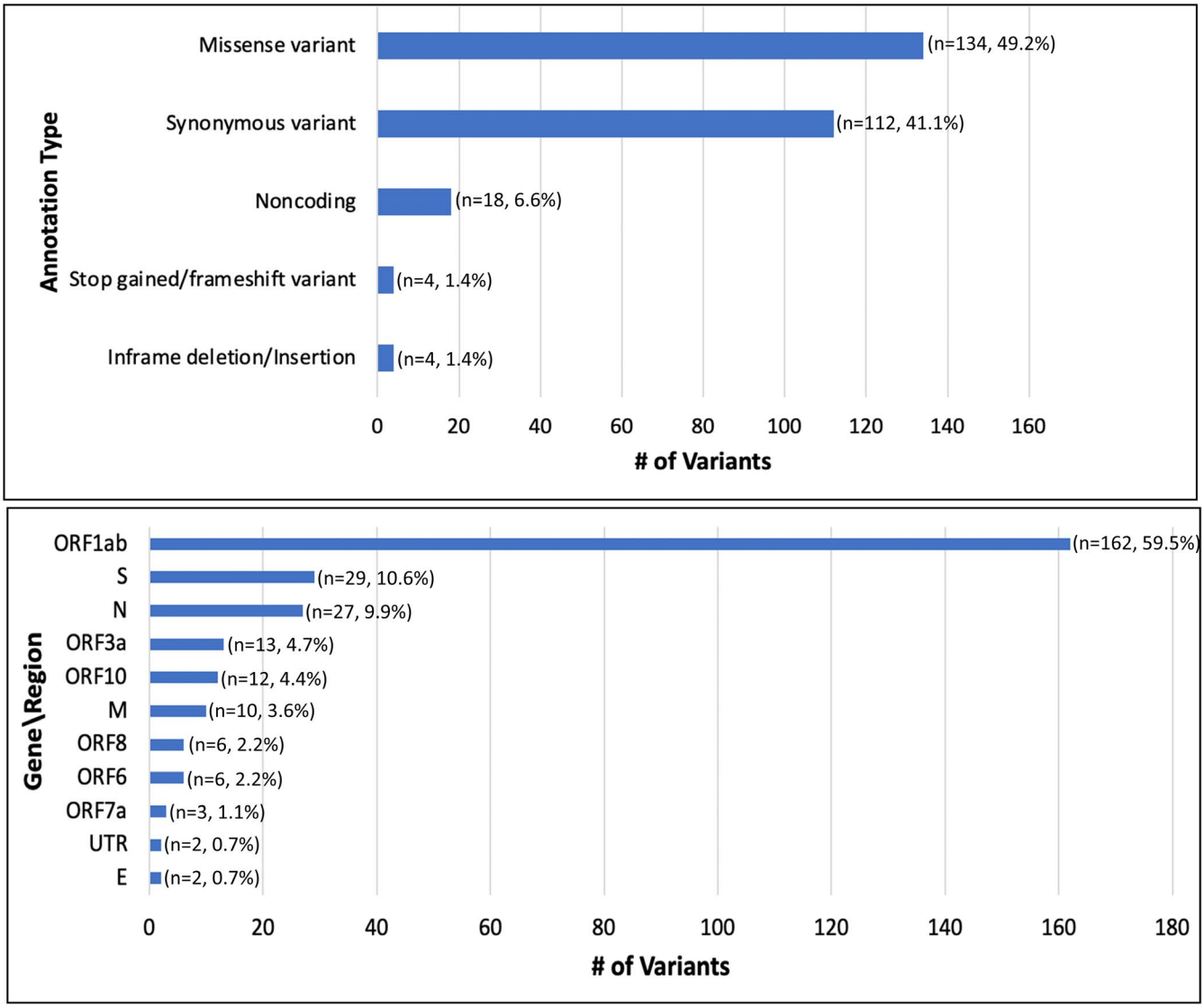

FIGURE 2 Distribution of the unique SARS-CoV-2 variants ($N = 272$) identified in 115 SARS-CoV-2 genomes in the UAE by type (a), and location (b)

genomic studies to include patients through June 2020 showed that those earlier strains seeded significant clustering consistent with almost exclusive community-based transmission in the time period between April and June 2020 (Figure 3). This clustering was exacerbated towards the end of March, which is when Dubai International Airport was closed to passenger flights (March 25, 2020).

The majority of the sequenced strains ($N = 60$, 52%) were in the European cluster, while 30% ($N = 34$) and 18% ($N = 21$) were in the Asian and Iranian clusters, respectively (Figure 3). The predominance of European cluster is consistent with the higher infectivity rates associated with the D614G mutation carried by most strains in this cluster (Figure S1).

3.5 | Genotype–phenotype correlation

The assembled 115 genomes were obtained from patients with mild ($N = 54$), moderate ($N = 32$) and severe ($N = 27$) disease, while clinical information was not available for two patients. The majority were males ($N = 77$, 67%) and the mean age was 41 years though age was highest in the severe disease group (50 years) (Table S4).

To determine if any viral mutation was associated with severe disease outcome, we counted the number of patients with mild and moderate disease carrying each sequence variant, and compared with those with severe COVID-19 (Figure 4). No statistically significant enrichments were identified (two tailed *t*-test), though missense (Q271R) and synonymous (R41R) mutations in the S and N proteins, respectively, were identified in 2/27 patients with severe COVID-19 but not in patients with mild or moderate disease (0/86; $p = .05$, Fisher's Exact Test). The two patients carrying this viral variant were both females, one of whom was 51 years old with a history of hypertension while the second was 64 years and otherwise healthy. The same two mutations were identified in a 37-year-old non-obese, non-diabetic, normotensive Indian man (EPI_ISL_426414) who had a travel history to China in January 2020 with no infection and then came into close contact

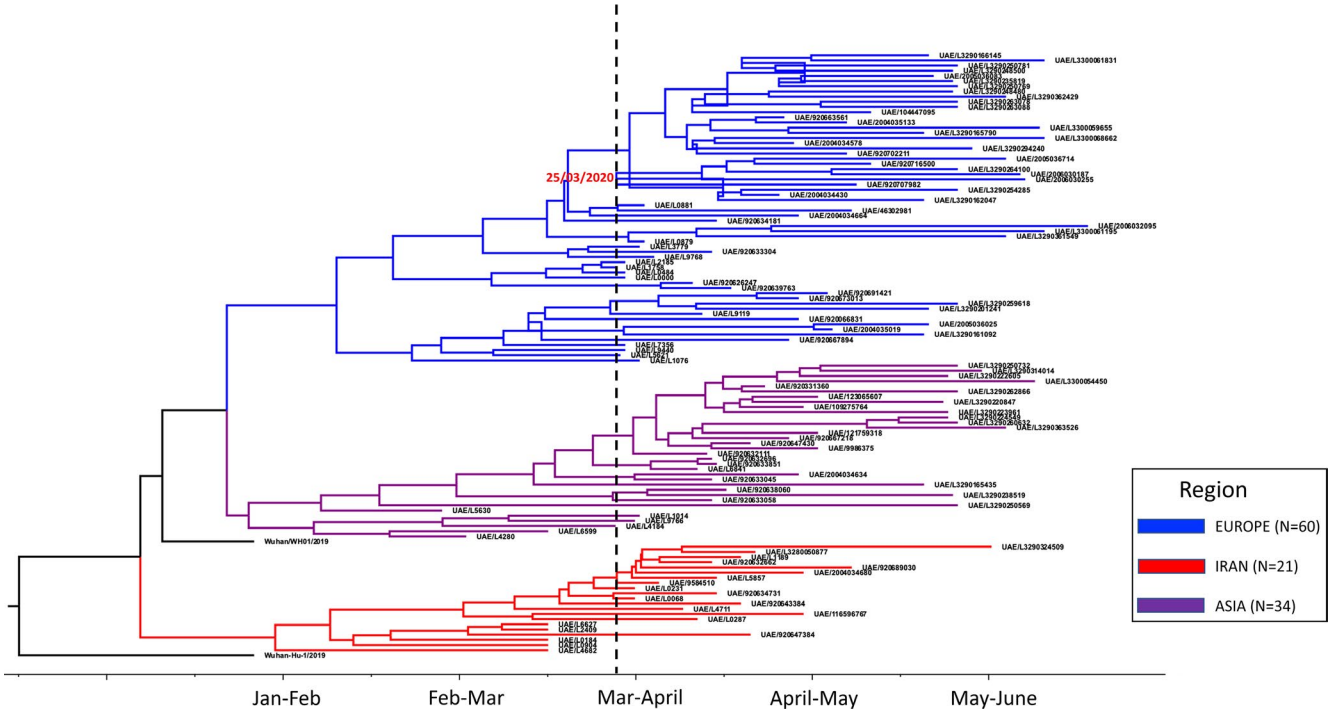

FIGURE 3 Maximum likelihood phylogenetic tree of SARS-CoV-2 sequences isolated from UAE ($n = 115$) was constructed using BEAST v.1.10.4 with strict clock mode, the Hasegawa–Kishino–Yano substitution model was used with 4 rate categories drawn from a gamma distribution (GTR4G). The two Wuhan genomes (Wuhan-Hu-1/2019, GISAID ID: EPI_ISL_402125 and Wuhan/WH01/2019, GISAID ID: EPI_ISL_406798) were used as reference genomes (branch colour = black). UAE SARS-CoV-2 isolates are indicated in blue (Europe, $n = 60$), green (Asia, $n = 34$) and red (Iran, $n = 60$). The dashed vertical line represents the date when Dubai International Airport closed to passengers

with someone who travelled to the Middle East. The 37-year-old Indian man (EPI_ISL_426414) subsequently developed a fever and cough on 26 March 2020 and was diagnosed with COVID-19 and admitted to hospital in India on 28 April 2020. There were no significant enrichments for variants in patients with moderate and severe disease versus those with mild disease suggesting no association between hospitalization and SARS-CoV-2 genomic variants (data not shown).

4 | DISCUSSION

Our study is the first in the Middle East region to explore the association between sociodemographic factors, SARS-CoV-2 viral strains and clinical outcomes in a representative sample of patients with COVID-19 in Dubai. Similar to other studies from around the world, age and underlying health conditions were positively associated with COVID-19-related hospital admission. Patients ≥ 40 years were twice as likely to be admitted to hospital compared with younger patients and overweight/obese or diabetic patients were 3–4 times more likely to be admitted to hospital and ICU.

Using two different sequencing protocols, we successfully obtained 115 full SARS-CoV-2 genomes from a representative patient sample in Dubai. Our analysis showed that 20 out of the 272 unique viral sequence variants identified in this cohort were novel. This data, which was submitted to GISAID, contributes to the growing variation catalogue of this novel virus and further delineates its mutational landscape.

Our phylogenetic analysis showed that multiple early introductions of SARS-CoV-2 into the UAE during the earlier phase of the pandemic (January – March 2020) (Alm et al., 2020; Harilal et al., 2020; Tayoun et al., 2020) seeded significant clustering due to community-based transmission subsequently between April and June 2020. This clustering overlapped with the closure of Dubai International Airport on March 25, 2020, further supporting the community-based mode of transmission. The three major clusters were an extension of the early European, Asian and Iranian strains (Harilal et al., 2020), though strains of European origin constitute the majority (60/115). Most of those strains (58/60) carried the D614G mutation previously associated with higher infectivity rates (Volz et al., 2020; Zhang et al., 2020). Thus, the closure of international flights and the strict community-based transmission between April – June 2020 in the UAE provides another epidemiological evidence that the D614G mutation increases infectivity (16, 17) (Figure S1).

Our genotype–phenotype analysis suggests a possible association between the Q271R missense and R41R synonymous mutations in the S and N genes, respectively, and disease severity. Those two mutations were very rare suggesting a strong negative selective pressure. They were simultaneously identified in only 4 patients (2 in our cohort and 2 in GISAID EPI_ISL_426414 and EPI_ISL_698928), 3 of whom did not have high risk of developing severe disease outcomes (i.e. were either young or females without

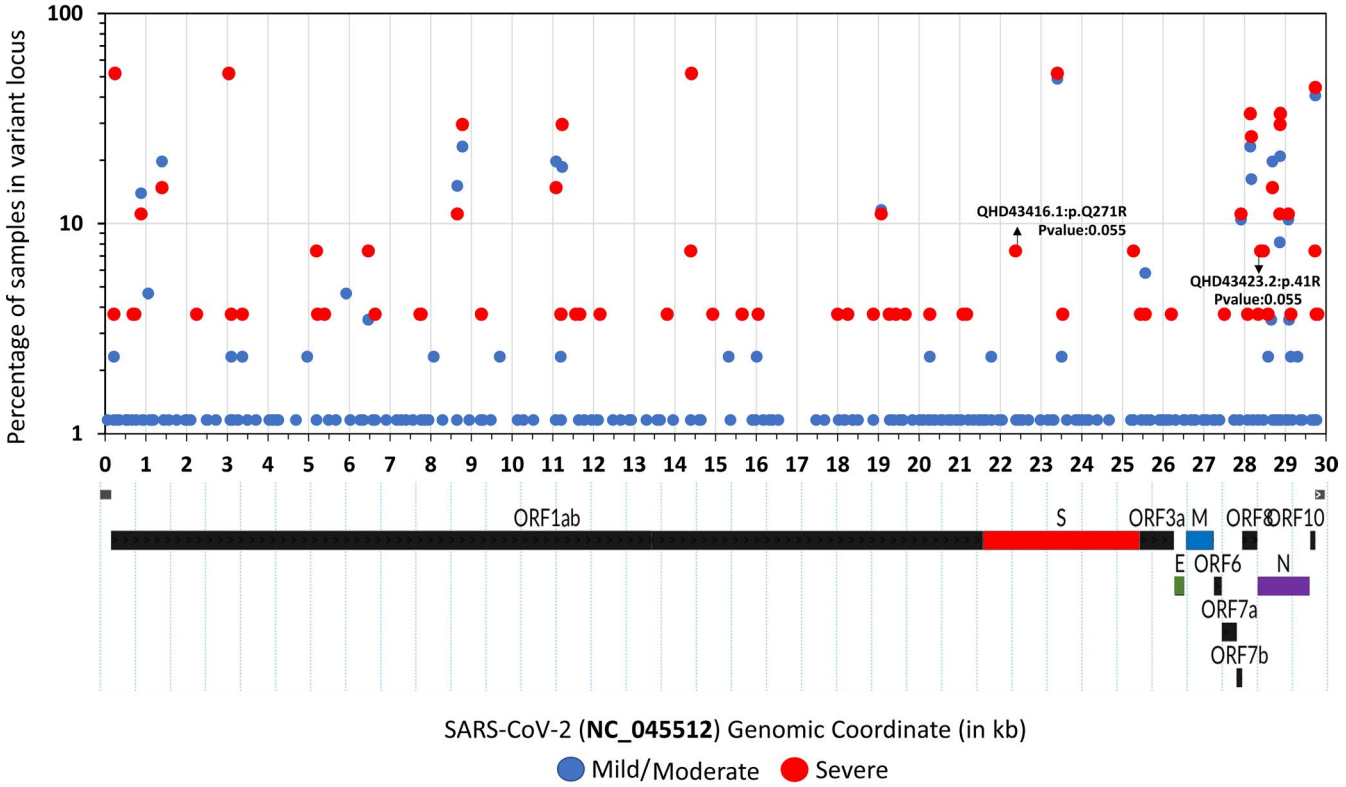

FIGURE 4 Association of SARS-CoV-2 sequence variants with COVID-19 clinical severity. X-axis indicates SARS-CoV2 genomic locus/mutation and Y-axis indicates the percentage of patients with severe (red circles) and mild/moderate (blue circles) disease at each mutation. Two mutations, Q271R and R41R in the S and N genes, respectively, were found in severely sick patients only ($p = .055$, two tailed *t*-test)

significant co-morbidities). No clinical information was available for patient EPI_ISL_698928 who, according to GISAID, was sampled in the UAE on 19th may 2020. Additional case level and/or functional data is needed to prove this association. The Q271R missense mutation is located in the N-terminal domain of S protein, outside the receptor binding domain (RBD), and its effect on protein structure or function cannot be deciphered without additional functional analysis.

Strengths of the study included the recruitment of a representative sample of the Dubai population by systematically selecting COVID-19 patients balanced by age, sex, disease severity and month during the first wave of the outbreak in the UAE (January to June 2020). Another major strength was the integration of detailed sociodemographic, clinical and viral genomic data to explore the relationship between different viral strains and disease severity. The main limitation was the inability to complete (full) whole genome sequencing on all samples most likely due to low viral load issues in the asymptomatic and mild cases. Overall, our findings contribute important clinical and novel molecular epidemiological data to the SARS-CoV-2 literature showing continued community-based transmission of the European strains in the Dubai population and new mutations that might be associated with severe disease in otherwise healthy adults. Our study has identified a missense (Q271R) and synonymous (R41R) mutation in the S and N proteins that might be associated with severe disease in otherwise healthy adults. The novel SARS-CoV-2 mutations identified in our representative sample of the Dubai population should be explored in other global populations, and we urge researchers conducting genetic and phylogenetic sequencing studies to collect detailed clinical data to explore the association between different viral strains and disease severity.

CONFLICTS OF INTEREST

Authors have no conflicts of interest to disclose.

AUTHOR CONTRIBUTIONS

AAA, AAT, NN and TL conceived the study and drafted the protocol. All authors provided critical input into the protocol. AAA, AAK and HK coordinated the ethical approval and sample retrieval. RV, ZD, ABA and AK conducted the RT-qPCR analysis. AAT, SR and DH performed the whole genome sequencing and phylogenetic analysis. HK performed data extraction from the medical records and TL completed data analysis for the manuscript. AAT and TL drafted the manuscript, AAA and NN refined it before all authors provided comments and feedback on the first draft, and then read and approved the final version of the manuscript.

DATA AVAILABILITY STATEMENT

All data generated or analysed during this study are included in this published article (and its Supplementary Information files), and the

sequences are available on the GISAID database under the corresponding accession numbers.

ORCID

Tom Loney https://orcid.org/0000-0003-1687-6587
Norbert Nowotny https://orcid.org/0000-0002-3548-571X
Ahmad Abou Tayoun https://orcid.org/0000-0002-9134-1673

REFERENCES

Alm, E., Broberg, E. K., Connor, T., Hodcroft, E. B., Komissarov, A. B., Maurer-Stroh, S., Melidou, A., Neher, R. A., O'Toole, Á., Pereyaslov, D. & WHO European Region sequencing laboratories and GISAID EpiCoV group; WHO European Region sequencing laboratories and GISAID EpiCoV group* (2020). Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. *Euro Surveillance*, *25*, 2001410. https://doi.org/10.2807/1560-7917.ES.2020.25.32.2001410

Butler, D., Mozsary, C., Meydan, C., Dnako, D., Foox, J., Rosiene, J., Shaiber, A., Afshinnekoo, E., MacKay, M., Sedlazeck, F. J., Ivanov, N. A., Sierra, M., Pohle, D., Zietz, M., Gisladottir, U., Ramlall, V., Westover, C. D., Ryon, K., Young, B., … Mason, C. E. (2020). Shotgun transcriptome and isothermal profiling of SARS-CoV-2 infection reveals unique host responses, viral diversification, and drug interactions. *bioRxiv* [Preprint], 2020.04.20.048066. https://doi.org/10.1101/2020.04.20.048066

Chernomor, O., von Haeseler, A., & Quang Minh, B. (2016). Terrace aware data structure for phylogenomic inference from supermatrices. *Systems Biology*, *65*, 997–1008. https://doi.org/10.1093/sysbio/syw037

Hadfield, J., Megill, C., Bell, S. M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., & Neher, R. A. (2018). Nextstrain: Real-time tracking of pathogen evolution. *Bioinformatics*, *34*, 4121–4123. https://doi.org/10.1093/bioinformatics/bty407

Harilal, D., Ramaswamy, S., Loney, T., Al Suwaidi, H., Khansaheb, H., Alkhaja, A., Varghese, R., Deesi, Z., Nowotny, N., Alsheikh-Ali, A., & Tayoun, A. A. (2020). SARS-CoV-2 whole genome amplification and sequencing for effective population-based surveillance and control of viral transmission. *Clinical Chemistry*, *66*, 1450–1458. https://doi.org/10.1093/clinchem/hvaa187

Jain, R., Ramaswamy, S., Harilal, D., Uddin, M., Loney, T., Nowotny, N., Alsuwaidi, H., Varghese, R., Deesi, Z., Alkhajeh, A., Khansaheb, H., Alsheikh-Ali, A., & Tayoun, A. A. (2021). Host transcriptomic profiling of COVID-19 patients with mild, moderate, and severe clinical outcomes. *Computational and Structural Biotechnology Journal*, *19*, 153–160. https://doi.org/10.1016/j.csbj.2020.12.016

Jesus, J. G., Sacchi, C., Candido, D. D. S., Claro, I. M., Sales, F. C. S., Manuli, E. R., Silva, D. B. B. D., Paiva, T. M., Pinho, M. A. B., Santos, K. C. O., Hill, S. C., Aguiar, R. S., Romero, F., Santos, F. C. P. D., Gonçalves, C. R., Timenetsky, M. D. C., Quick, J., Croda, J. H. R., Oliveira, W., … Faria, N. R. (2020). Importation and early local transmission of COVID-19 in Brazil, 2020. *Revista do Instituo De Medicina Tropical De Sao Paulo*, *62*, e30. https://doi.org/10.1590/s1678-9946202062030

Johns Hopkins Center for Systems Sciences and Engineering (2020). COVID-19 Dashboard. Retrieved from https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*, 3059–3066. https://doi.org/10.1093/nar/gkf436

Loney, T., Aw, T. C., Handysides, D. G., Ali, R., Blair, I., Grivna, M., Shah, S. M., Sheek-Hussein, M., El-Sadig, M., Sharif, A. A., & El-Obaid, Y. (2013). An analysis of the health status of the United Arab Emirates: The 'Big 4' public health issues. *Global Health Action*, *6*, 20100. https://doi.org/10.3402/gha.v6i0.20100

Rambaut, A. (2018). *FigTree 1.4.2 software*. Institute of Evolutionary Biology, Univ. Edinburg. Retrieved from http://tree.bio.ed.ac.uk/software/figtree/

Stefanelli, P., Faggioni, G., Lo Presti, A., Fiore, S., Marchi, A., Benedetti, E., Fabiani, C., Anselmo, A., Ciammaruconi, A., Fortunato, A., De Santis, R., Fillo, S., Capobianchi, M. R., Gismondo, M. R., Ciervo, A., Rezza, G., & Castrucci, M. R., Lista, F. & On behalf of Iss Covid-Study Group (2020). Whole genome and phylogenetic analysis of two SARS-CoV-2 strains isolated in Italy in January and February 2020: Additional clues on multiple introductions and further circulation in Europe. *Euro Surveillance*, *25*, 2000305. https://doi.org/10.2807/1560-7917.ES.2020.25.13.2000305

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, *4*, vey016. https://doi.org/10.1093/ve/vey016

Tayoun, A. A., Loney, T., Khansaheb, H., Ramaswamy, S., Harilal, D., Deesi, Z. O., Varghese, R. M., Al Suwaidi, H., Alkhajeh, A., AlDabal, L. M., Uddin, M., Hamoudi, R., Halwani, R., Senok, A., Hamid, Q., Nowotny, N., & Alsheikh-Ali, A. (2020). Multiple early introductions of SARS-CoV-2 into a global travel hub in the Middle East. *Scientific Reports*, *10*, 17720. https://doi.org/10.1038/s41598-020-74666-w

Uddin, M., Mustafa, F., Rizvi, T. A., Loney, T., Suwaidi, H. A., Al-Marzouqi, A. H. H., Eldin, A. K., Alsabeeha, N., Adrian, T. E., Stefanini, C., Nowotny, N., Alsheikh-Ali, A., & Senok, A. C. (2020). SARS-CoV-2/COVID-19: Viral genomics, epidemiology, vaccines, and therapeutic interventions. *Viruses*, *12*, 526. https://doi.org/10.3390/v12050526

Volz, E., Hill, V., McCrone, J. T., Price, A., Jorgensen, D., O'Toole, Á., Southgate, J., Johnson, R., Jackson, B., Nascimento, F. F., Rey, S. M., Nicholls, S. M., Colquhoun, R. M., da Silva Filipe, A., Shepherd, J., Pascall, D. J., Shah, R., Jesudason, N., Li, K., … Connor, T. R. (2020). Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell*, *S0092–8674*, 31537–31543. https://doi.org/10.1016/j.cell.2020.11.020

Yang, X., Dong, N., Chan, E. W., & Chen, S. (2020). Genetic cluster analysis of SARS-CoV-2 and the identification of those responsible for the major outbreaks in various countries. *Emerging Microbes and Infections*, *9*, 1287–1299. https://doi.org/10.1080/22221751.2020.1773745

Zhang, L., Jackson, C. B., Mou, H., Ojha, A., Peng, H., Quinlan, B. D., Rangarajan, E. S., Pan, A., Vanderheiden, A., Suthar, M. S., Li, W., Izard, T., Rader, C., Farzan, M., & Choe, H. (2020). SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nature Communications*, *11*, 6013. https://doi.org/10.1038/s41467-020-19808-4

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Loney T, Khansaheb H, Ramaswamy S, et al. Genotype-phenotype correlation identified a novel SARS-CoV-2 variant possibly linked to severe disease. *Transbound Emerg Dis*. 2022;69:465–476. https://doi.org/10.1111/tbed.14004