Emergence of new SARS-CoV-2 variant under investigation in Saudi Arabia

Majed F. ALGHORIBI (alghoribima@ngha.med.sa)
Infectious Diseases Research Department, King Abdullah International Medical Research Center, Riyadh, Saudi Arabia https://orcid.org/0000-0002-2122-8640

Abdulrahman ALSWAJI
Infectious Diseases Research Department, King Abdullah International Medical Research Center, Riyadh, Saudi Arabia

Liliane OKDAH
Infectious Diseases Research Department, King Abdullah International Medical Research Center, Riyadh, Saudi Arabia

Sadeem ALHAYLI
Infectious Diseases Research Department, King Abdullah International Medical Research Center, Riyadh, Saudi Arabia

Zinab BU ALI
Infectious Diseases Research Department, King Abdullah International Medical Research Center, Riyadh, Saudi Arabia

Maha A. ALZAYER
Infectious Diseases Research Department, King Abdullah International Medical Research Center, Riyadh, Saudi Arabia

Reem E. ALARFAJ
Infectious Diseases Research Department, King Abdullah International Medical Research Center, Riyadh, Saudi Arabia

Reema BAWAZEER
Infectious Diseases Research Department, King Abdullah International Medical Research Center, Riyadh, Saudi Arabia

Mashael ALAHMADI
Infectious Diseases Research Department, King Abdullah International Medical Research Center, Riyadh, Saudi Arabia

Sara ALGHAMDI
Infectious Diseases Research Department, King Abdullah International Medical Research Center, Riyadh, Saudi Arabia

Sarah ALHASSINAH
Infectious Diseases Research Department, King Abdullah International Medical Research Center, Riyadh, Saudi Arabia

Aiman El-SAED
Infection Prevention and Control Department, King Abdulaziz Medical City (KAMC), Ministry of National Guard Health Affairs (MNGHA), Riyadh, Saudi Arabia
Fayssal M. FARAHA
Infection Prevention and Control Department, King Abdulaziz Medical City (KAMC), Ministry of National Guard Health Affairs (MNGHA), Riyadh, Saudi Arabia
Majid M. ALSHAMRANI
Infection Prevention and Control Department, King Abdulaziz Medical City (KAMC), Ministry of National Guard Health Affairs (MNGHA), Riyadh, Saudi Arabia
Sameera M. AL JOHANI
Department of Pathology and Laboratory Medicine, King Abdulaziz Medical City (KAMC), Ministry of National Guard Health Affairs (MNGHA), Riyadh, Saudi Arabia
Michel DOUMITH
Infectious Diseases Research Department, King Abdullah International Medical Research Center, Riyadh, Saudi Arabia

---

Research Article

Keywords: COVID-19, Genome Sequencing, Virus Evolution, Spike protein, variant of interest, variant of concern, Saudi Arabia

Posted Date: August 26th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-847384/v1

Abstract

Genomic surveillance helps public health tracking the path of a pandemic, understand its transmission route, and how quickly the virus is spreading and adapting through mutation and recombination. In late 2020, the Infectious Diseases Research Department (IDRD) at King Abdullah International Medical Research Center (KAIMRC) initiated a surveillance program focused on monitoring the dissemination of SARS-CoV-2 variants by sequencing the receptor-binding domain (RBD) of 20 to 32% of all community-acquired SARS-CoV-2 positive cases identified from November 2020 to February 2021 at King Abdulaziz Medical City (KAMC) in Riyadh. Sequence analysis detected among sequenced isolates a SARS-CoV-2 variant harboring an N501T substitution in the receptor-binding domain (RBD). COVID-19 cases linked to this new variant under investigation, first detected in isolates from November, grew exponentially in 2021 to account alone for more than 62 % (142/228) of all SARS-CoV-2 positive cases studied in February, thus suggesting that this variant might have increased transmissibility. Genome sequencing showed that this variant has evolved from the pangolin lineage B.1.1 and diverged from the original strain of Wuhan in China by ten amino acid changes located in ORF1a (P3359L and Q3729K), ORF1b (P314L), spike (S) protein (F157S, N501T, D614G), ORF6 (F2S) and nucleocapsid (N) protein (I84V, R203K and G204R). Isolates belonging to the new variant were further split into two sub-groups based on additional changes in the spike. Of these, one sub-group carried the Y144*, G257S, T859N and A899S variations combined, while one carried the H49Y variation alone. Patients infected with this new variant did not show any increase in the severity of infections. The rapid rise of the new variant among community-acquired SARS-CoV-2 cases suggests a national spread, although this needs to be further assessed carefully.

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a novel coronavirus identified as the cause of a cluster of pneumonia cases at the end of 2019 in China [1]. This new emerging pathogen responsible for the coronavirus disease of 2019 (COVID-19) has spread worldwide, infected millions of people, and caused many deaths. In such pandemic events, genomic studies are essential to investigate and track the dissemination and evolution of the emerging virus. Since the start of the pandemic and by the time of submitting the paper, more than two million complete genome sequences of SARS-CoV-2 have been deposited and shared in GISAID (https://www.gisaid.org/). This large amount of genomic data played an essential role in providing invaluable insights into the virus genetic diversity, molecular evolution, and transmission dynamics [2–5]. Many variants of SARS-CoV-2 are currently circulating worldwide, and few numbers of those variants are now being monitored due to their high-risk characteristics and degree of spread. Of these, four variants of concern comprising the alpha variant (B.1.1.7) first detected in the UK [6], beta (B.1.351) in South Africa [7], gamma (P.1) in Brazil [8] and delta (B.1.617.2) in India [8] are being carefully mapped because of their mutations that have the potential to affect transmissibility or severity and interfere with diagnostics, treatments or vaccines efficacy. The current study reports the observation of a new SARS-CoV-2 variant in Saudi Arabia based on genomic characterization of samples tested positive at the clinical diagnostic laboratory of King Abdulaziz

medical city (KAMC) in Riyadh. The new variant under investigation accounted for only a few cases when first discovered in November 2020 but later grew rapidly in prevalence to account alone for over 62% of all screened isolates at KAMC in February 2020.

Materials And Methods

SARS-CoV-2 screening and patient information

The emergence of new SARS-CoV-2 variants that posed an increased risk to global public health prompted the Infectious Diseases Research Department (IDRD) to initiate in late 2020 a screening program to monitor mutations in the region encoding the receptor-binding domain (RBD) of the spike protein of strains circulating in Saudi Arabia, and in particular, those associated with SARS-CoV-2 Variants of Concern (VOCs) reported at that time. In total, 20 to 34% of all nasopharyngeal swab specimens tested positive at each month from November 2020 to February 2021 at King Abdulaziz Medical City (KAMC) in Riyadh were randomly selected for screening as part of the program. Total RNA was extracted from selected specimens using the QIAamp Viral RNA (Qiagen) and MGIEasy Nucleic Acid (MGI Tech Co.) extraction kits. Reverse transcription into cDNA was carried out using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). The RBD region was later amplified using the following primers (nCOV_F45 5¢- TCTCAGCCTTTTCTTATGGACCT-3¢), (nCOV_R46 5¢- TGTGGGTATGGCAATAGAGTT -3¢). and analyzed by Sanger sequencing. Information about the patient demographics and medical data was collected from the medical record systems of KAMC.

Genome sequencing

Viral RNA from selected samples were first quantified by qPCR using the 2019-nCoV CDC EUA kit (IDT, USA) in a StepOn Real-Time PCR System (Thermo Fisher Scientific) and those exhibiting a CT value below 25 were further analyzed by genome sequencing. Viral genomes were sequenced on the Illumina MiSeq or MGI MGISEQ-G400 platforms using in both cases an amplicon-based enrichment strategy prior to sequencing. Sequencing on the MGI platform was performed using viral genomes enriched directly from extracted RNA using the ATOPlex SARS-CoV-2 Full-Length Genome Panel v1.0 (MGI Tech Co.) according to the manufacturer's instructions. Viral genomes sequenced on the Illumina platform were amplified from cDNA templates by PCR using a set of 58 pairs of overlapping primers (data not shown) that were designed to cover the entire SARS-CoV-2 genome. The amplified segments thus generated were equally pooled and purified with the Qiaquick PCR purification kit (QIAGEN) prior to library preparation with the Nextera XT library preparation DNA kit (Illumina) according to the manufacturer's instructions.

Bioinformatics and phylogenetic analyses

Paired reads were filtered out by mapping to the viral reference genome sequence (GenBank accession: MN908947.3) using Bowtie2 (version 2.3.4.1). The aligned reads were then supplied to the trim component of the iVar software (version 1.3) to remove the sequences of the primers used for amplification. Consensus sequences were called from the trimmed reads using Samtools (version 1.12)

and the iVar consensus component with a minimum read coverage of 10 and minimum base frequency of 85%. In order to put the analysis into an international context, genomes sequenced in this study were compared to representatives of SARS-CoV-2 genomes available in the GISAID database. Genomes representing all current SARS-CoV-2 lineages (n=3952) as well as all genomes harboring the N501T variation in the spike protein (n=3546) or originating from Saudi Arabia (n=957) deposited by 25 of June 2021 were retrieved. Phylogenetic trees were constructed with the NextStrain ncov [4] automated pipeline using the consensus sequences and associated metadata as inputs (i.e. collection date and patient demographics, and travel history). Assignment to the dynamic nomenclature proposed by Rambaut *et al.* was determined using the Pangolin software (github.com/hCoV-2019/pangolin) [9].

Ethical consideration

Nasopharyngeal swabs from COVID-19 infected patients were collected and stored in the Emerging Infectious Diseases Registry and Biobanking (EIRD) at the National Guard Health Affairs (NGHA) according to national and institutional standards. The study was approved by the Institutional Review Board at King Abdullah International Medical Research center (IRBC /0707/20).

Results

Demographic characteristics of patients infected with SARS-CoV-2 N501T variant.

Patients infected with the new N501T variant comprised of 51.2% males and 48.8% females with ages ranging from 0.4 to 96 years (median 40). Nearly two-thirds of the patients presented with cough (47/125, 37.6%) , fever (37/125, 29.6%), shortness of breath (28/125, 22.4 %), sore throat (19/125, 15.2 %), runny nose (16/125, 12.8%), headache (9/125, 7.2%), vomiting (7/125, 5.6%) or diarrhea (5/125, 4%) (Table 1). All cases were community-acquired infections, of which 24 required hospital admission. Nearly third (37.5%, 9/24) of hospitalized patients required ICU care and of which seven had fatal outcomes with ages ranging from 64 to 96 years, all with known comorbidities, including diabetes or hypertension. Local transmission through contacts with confirmed cases, mainly among household members, was established for 44% of the patients.

Table 1
Demographic characteristics of sequenced COVID-19 cases infected with the main N501T variant (n = 125) included in the study.

Baseline variables	Number	Percentage
Age		
• Median	40	-
• Range	0.4–96	-
Gender		
• Male	64	51.2%
• Female	61	48.8%
Hospital admission	24	19.2%
Mortality	7	5.6%
Comorbidities		
• Diabetes	20	16%
• Hypertension	23	18.4%
• DLP	13	10.4%
• Asthma	4	3.2%
Symptoms		
• Asymptomatic	39	31.2%
• Cough	47	37.6%
• Fever	37	29.6%
• Shortness of breath	28	22.4%
• Sore throat	19	15.2%
• Runny Nose	16	12.8%
• Headache	9	7.2%
• Vomiting	7	5.6%
• Diarrhea	5	4%

Screening of the RBD of the spike protein

Screened specimens represented 20.26% (110/543), 20.65% (114/552), 22.61% (90/398), and 32.43% (228/703) of all samples confirmed positive at KAMC from November 2020 to February 2021. Sequence analysis of the region encoding the BRD of the spike protein identified five samples with N501T variation in November and 12 in December. However, in 2021 the number of SARS-CoV-2 isolates harboring this variation increased exponentially to represent 32.22% (29/90) and 62.28% (142/228) of sequenced isolates in January and February, respectively. The majority of the isolates harbored the N501T variation alone (180/188), and only a few carried this variation in combination with either P330S (n=1), P337T (n=3), G476V (n=2) or V320F (n=2) substitution. Other modifications associated with variants of concerns such as N501Y (n=7) alone or in combination with K417N/E484K (n=3) or L452R (n=1) as well as the L452R substitution alone (n=3) were also detected. On few occasions, the N501Y was found associated with Y449H (n=2) or T523N (n=1). Sequence analysis also identified ten isolates carrying the N439K substitution and five with V367F. All remaining isolates (n=332) had no modifications in the RBD.

SARS-CoV-2 whole-genome sequences from Saudi Arabia.

Screening of SARS-CoV-2 genome sequences deposited in the GISAID database by the end of June identified 957 genomes originating from Saudi Arabia. Nearly all (953/957, 99.58%) were deposited between February and August 2020 and belonged to 25 different pangolin lineages, but with pangolin lineage, B.1 being dominant (436/957, 45.56%), followed to less extent by B.1.36 (134/957, 14%), B.1.260 (107/957, 11.18%), B.1.1 (85/957, 8.88%) and B.1.1.161 (41/957, 4.28%). The remaining 20 lineages were each represented by less than 40 samples (range 1 to 38), none of which represented a variant of interest or a variant of concern. Only four genomes were deposited in April 2021 and included one alpha variant (i.e. B.1.1.7), one beta variant (i.e. B.1.351), while two harbored the N501T variation in the RBD of the spike protein.

Phylogenetic analysis of SARS-CoV-2 sequenced genomes.

Isolates carrying the N501T variation and exhibiting good CT values (n=132) were whole-genome sequenced and compared against the published SARS-CoV-2 phylogeny. The majority of sequenced genomes (125/132, 94.70%) grouped together in one clade (Figure 1) and only a handful clustered apart with other published genomes belonging to pangolin lineages A.28 (n=4), C.36 (n=2), and B.1.1.36 (n=1). The main cluster of sequenced genomes also comprised other 182 published genomes from 20 different countries, including Bahrain, Georgia, Qatar, Pakistan, India, Sri Lanka, Singapore, Indonesia, Philippines, Australia, Portugal, United Kingdom, France, Germany, Slovenia, Sweden, Latvia, Finland, United States, and Canada and which were all deposited after the end of December. The phylogeny indicated that the predominant N501T cluster has evolved from the pangolin B.1.1 lineage. In addition to B.1.1 substitutions in ORF1b (P314L), S (D614G), and N (R203K and G204R), these genomes accumulated additional substitutions in ORF1a (P3359L Q3729K), S (F157S, N501T), ORF6 (F2S) and N (I84V) (Figure 1). Isolates from the main cluster can be further split into two sub-groups based on extra modifications in the S protein where one sub-group carried four additional modifications (i.e. one deletion Y144* and three substitutions G257S, T859N and A899S) whereas one had only one additional substitution (H49Y) at

amino acid position 49. Mapping the N501T amino acid variation in the spike protein onto the branches of the global SARS-CoV-2 phylogeny revealed that genomes carrying this variation were scattered across 74 lineages (3462 sequences from 58 countries) comprising mainly pangolin lineages B.1.517 (1725/1808, 95%), A.28 (322/322, 100%), B.1.517.1 (164/175, 94%) and C.2.1 (120/125, 96%) and thus suggesting that this substitution originated independently at multiple occasions.

Discussion

Since the beginning of SARS-CoV-2 pandemic, high-throughput whole-genome sequencing (WGS) has been widely used as an invaluable and essential public health approach for sufficient and real-time tracking of the virus. More countries are currently utilizing the power of WGS technologies and contributing extensively and expeditiously by sharing and analyzing SARS-CoV-2 genome sequences using common platforms as but not exclusively to GISAID, Nextstrain, Mircoreact and Outbreak.info [2–5]. This impressive effort is a testament to the global collaboration to better understanding SARS-CoV-2 virus during the pandemic.

The exponential increase of COVID-19 cases associated with this new variant under investigation strongly suggests a substantially increased transmissibility [10–12]. On the other hand, the mortality and hospitalization rates associated with this variant indicated that these modifications are unlikely to have an effect on the severity of the disease. However, their impact on various aspects of the infection, diagnostics and vaccine protection needs to be further assessed more in-depth [12]. Sequenced isolates included in this study were all from one screening center located in Riyadh. Nevertheless, all samples included in this study were from community-acquired cases suggesting that these are representatives at the national level, although this speculation needs to be confirmed.

The SARS-CoV-2 S protein, which plays a key role in viral binding to host cell receptors (i.e., human angiotensin-converting enzyme 2 [hACE2]), contains three subunits, the N-terminus consisting of the signal peptide (amino acids 1–13), the S1 subunit (14–685 residues) and the S2 subunit (686–1273 residues). The variant under investigation from this study carried several modifications in the S protein, including the RBD. Previous studies suggest that variations at amino acid position 501 in the RBD of the spike may enhance the binding affinity of the S protein for hACE2 [13, 14].

Phylogenetic analysis grouped the majority of our SARS-CoV-2 sequenced isolates with S:N501T substitution in one clade differing by a set of amino acid differences from other S:N501T lineages s present in the GISAID databases. These variations might potentially lead to define a new pango lineage. The most striking result to emerge from our analysis is that the spike mutation at the 501 position might be a major driver of the increased transmissibility by allowing tight interaction of the spike RBD to the human ACE2 [15]. Published results showed that SARS-CoV-2 variants are constantly evolving, and therefore more recent SARS-CoV-2 specimens tested positive at KAMC are being checked to assess the evolution of this new variant and to detect new variants that might have evolved more recently.

Declarations

Data availability.

The complete genome sequence of the SARS-CoV-2 sequenced isolates in this study has been deposited in GenBank (https://www.ncbi.nlm.nih.gov/genbank/) under the accession numbers MZ895658-MZ895789.

Funding

MFA acknowledges funding from King Abdulaziz City for Science and Technology (5-20-01-536-0003) as part of research funding program (COVID-19 Research Grant Program) that aims to support R&D on COVID-19 to better understand and elucidate the Coronavirus pandemic through in-depth etiological and epidemiological studies, disease pathogenesis, data modelling, developing rapid diagnostics.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Kong WH, Li Y, Peng MW, Kong DG, Yang XB, Wang L, et al. SARS-CoV-2 detection in patients with influenza-like illness. Nat Microbiol. 2020;5: 675–678. doi:10.1038/s41564-020-0713-1
2. Julia L. Mullen, Ginger Tsueng, Alaa Abdel Latif, Manar Alkuzweny, Marco Cano, Emily Haag, Jerry Zhou, Mark Zeller, Nate Matteson, Kristian G. Andersen, Chunlei Wu, Andrew I. Su, Karthik Gangavarapu, Laura D. Hughes and the C for VSB outbreak. inf. outbreak.info. 2020.
3. Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb genomics. 2016;2: e000093. doi:10.1099/mgen.0.000093
4. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. NextStrain: Real-time tracking of pathogen evolution. Bioinformatics. 2018;34: 4121–4123. doi:10.1093/bioinformatics/bty407
5. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Challenges. 2017;1: 33–46. doi:10.1002/gch2.1018
6. Andrew Rambaut, Nick Loman, Oliver Pybus, Wendy Barclay, Jeff Barrett, Alesandro Carabelli, Tom Connor, Tom Peacock, David L Robertson EV. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations - SARS-CoV-2 coronavirus / nCoV-2019 Genomic Epidemiology - Virological. In: virological [Internet]. 2020 [cited 3 Jul 2021]. Available: https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563

7. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al. Detection of a SARS-CoV-2 variant of concern in South Africa. Nature. 2021;592: 438–443. doi:10.1038/s41586-021-03402-9
8. Faria NR, Mellan TA, Whittaker C, Claro IM, Candido DDS, Mishra S, et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. Science (80-). 2021;372: 815. doi:10.1126/science.abh2644
9. Rambaut A, Holmes EC, Hill V, OToole A, McCrone J, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology. bioRxiv. 2020; 2020. 04.17.046086 . doi:10.1101/2020.04.17.046086
10. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. Science (80-). 2021;372: eabg3055. doi:10.1126/science.abg3055originally
11. Deng X, Garcia-Knight MA, Khalid MM, Servellita V, Wang C, Morris MK, et al. Transmission, infectivity, and neutralization of a spike L452R SARS-CoV-2 variant. Cell. 2021;184: 3426–3437.e8. doi:10.1016/j.cell.2021.04.025
12. Campbell F, Archer B, Laurenson-Schafer H, Jinnai Y, Konings F, Batra N, et al. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. Eurosurveillance. 2021;26: 2100509. doi:10.2807/1560-7917.es.2021.26.24.2100509
13. Fiorentini S, Messali S, Zani A, Caccuri F, Giovanetti M, Ciccozzi M, et al. First detection of SARS-CoV-2 spike protein N501 mutation in Italy in August, 2020. Lancet Infect Dis. 2021;21: e147. doi:10.1016/S1473-3099(21)00007-4
14. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, et al. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. Cell. 2020;182: 1295–1310.e20. doi:10.1016/j.cell.2020.08.012
15. Peacock TP, Penrice-Randal R, Hiscox JA, Barclay WS. SARS-CoV- 2 one year on: Evidence for ongoing viral adaptation. Journal of General Virology. Microbiology Society; 2021. p. 001584. doi:10.1099/jgv.0.001584

Figures

Figure 1

Time-scaled phylogeny of downsampled SARS-CoV-2 genomes representing the global phylogeny with all sequenced Saudi genomes, including those generated in the study for the new variant under investigation.