Microbes and Infectious Diseases 2025; 6(2): 414-428

Microbes and Infectious Diseases

Journal homepage: https://mid.journals.ekb.eg/

Original article

Identifying SARS-CoV-2 lineage and spike protein mutations: A single center cross-sectional study

Ali N. M. Gubran[1*], Dalia Elsayed Metwally [2], Ezz Eldein Anwar Saleh[2], Reham Abo Elwafa[3], Rasha Emad[4], Iman S. Naga[2]

1- *Diagnostic and molecular microbiology, Department of Health Sciences, Faculty of Medicine and Health Sciences, University of Science and Technology, Aden-Yemen*
2- *Department of Microbiology, Medical Research Institute, University of Alexandria, Alexandria, Egypt*
3- *Department of Clinical Pathology, Faculty of Medicine, Alexandria University, Alexandria, Egypt*
4- *Clinical Pharmacy, Alexandria Main University Hospital, Alexandria, Egypt*

ARTICLE INFO

Article history:
Received 16 July 2024
Received in revised form 15 September 2024
Accepted 17 September 2024

Keywords:
COVID-19
D614G mutation
Pango lineage
SARS-CoV-2
Spike protein

ABSTRACT

Background: Since the onset of SARS-CoV-2 pandemic, it has become a hot spot for research. **Aim:** This study aimed to detect variations in the spike protein of SARS-CoV-2 isolated from Egyptian patients and correlate them with laboratory data. **Methods:** Fifty patients with positive nasopharyngeal swabs by PCR were enrolled in this cross-sectional study. Partial spike protein was successfully amplified and sequenced for 23 isolates. This study determined the clade, variant and lineage. Phylogenetic analysis was performed to detect heterogeneity among our isolates. **Results:** The partial spike protein belonged to 3 clades and 3 Pango lineages. The most common lineage was C.36.3 (56.5%), followed by B.1.1.7 (39%) then B.1.1 (4.5%). D614G mutation was present in 100% of isolates, while the second most frequent mutation was Q677H (60.9%) followed by L452R (52.2%). B.1.1.7 lineage was associated with higher WBC and lymphocytes compared to other lineages (p= 0.02, p=0.012, respectively). **Conclusion:** Level of WBC may significantly differ according to the infecting lineage. This finding may impact progression of SARS-CoV-2 infection.

Introduction

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) is the causative agent of Coronavirus disease 2019 (COVID-19). The World Health Organization (WHO) has designated the disastrous spread of COVID-19 that has already claimed millions of lives worldwide as a public health emergency of international concern. Globally, as of March 7, 2023, there have been 759,408,703 confirmed cases of COVID-19, including 6,866,434 deaths and in Egypt 515,759 confirmed cases of COVID-19 with 24,812 deaths as reported to the WHO. As of March 4, 2023, a total of 13,229,673,098 vaccine doses have been administered [1].

Spike (S) protein is the most important antigen of coronavirus particle that is targeted by the host neutralizing antibodies; therefore, it is the major determinant of virulence of coronaviruses as it mediates virus's attachment to the specific host receptor [2, 3]. This protein is cleaved into separate polypeptides by the host cell proteases such as furin-like protease [4]. These polypeptides are S1 and S2 where S1 serves as a large receptor binding domain (RBD) while S2 forms the stalk of the spike

DOI: 10.21608/MID.2024.304807.2084
* *Corresponding author:* Iman Salah Naga
E-mail address: *iman.naga@alexu.edu.eg*

molecule which mediates fusion of viral envelope with the host cell membrane [5-7], thereby allowing the entry of the viral genome into the cell cytoplasm of the host [8]. The glycan molecules that coat the S protein may help the virus to escape the host immune system [7]. The S protein is a valuable target for development of vaccine against SARS-CoV-2 and antiviral therapy efforts. It may also be used for diagnostic purposes [9,10].

The rate of mutations in SARS-CoV-2 genome leads to genomic variability that enable virus to evade immune defenses or to become resistant to antiviral therapy [11]. In S protein of SARS-CoV-2, there are various mutations that, can cause alteration in transmissibility, antigenicity of the virus and promoting its evasion of immune system [12]. Examples for mutations in S protein of SARS-CoV-2, is D614G that may facilitate the spread of virus rapidly [13,14]. In addition to pattern of spread, it also increases transmission and infectivity of virus by giving the virus some fitness advantage [15]. Other mutations such as Q677H, L452R, N501Y, A570D and P681H are also recognized to play a role to increase pathogenesis of SARS-CoV-2 [16].

This is a preliminary study due to the small sample size of the included isolates. The aim of this study was to investigate variations in the spike protein including genotypes and/or mutations in SARS-CoV-2 isolated from Egyptian patients by comparing them with the first reported isolate from Wuhan, China and to correlate these variations with the patients' laboratory data. Here, we hypothesized that maybe some of patients' laboratory data may differ based on different lineage of SARS-CoV-2.

Patients and Methods

Study settings

A total of 50 patients who tested positive for COVID-19 infection by real-time PCR were included in this cross-sectional study. Their demographic, clinical symptoms and laboratory data were collected for correlation analysis. Written informed consent was signed by each patient before enrollment in the study with ethical approval number IORG0008812.

Twenty-three isolates were successfully amplified and sequenced then subjected to genotyping, phylogenetic analysis and mutation detection. This is a preliminary work due to the small sample size of the study. Age and gender are included in the study as potential confounder variables.

The 23 nucleotide sequences were deposited in GenBank NCBI with the following accession numbers: QQ346382:QQ346403 and QQ347969. Also, they are present in Global Initiative on Sharing All Influenza Data (GISAID) with the following accession numbers: EPI_ISL_16713842: EPI_ISL_16713862 and EPI_ISL_16743975: EPI_ISL_16743976.

Sampling

Nasopharyngeal (NP) swabs and single 5 ml blood samples were collected from 50 Egyptian patients attending the Clinical Pathology Department, Faculty of Medicine, Alexandria, Egypt during a 3-month period between March to May 2021.

Laboratory investigations

Blood samples were used for performing complete blood count (CBC) for all 50 patients, then centrifuged at 1500 rpm and sera were separated. D-dimer, serum ferritin, lactate dehydrogenase (LDH), procalcitonin (PCT), C-reactive protein (CRP) and interleukin-6 (IL-6) levels were also determined.

This work was conducted in the Microbiology Department, Alexandria, Egypt, according to the ethical considerations of the Medical Research Institute, University of Alexandria and conforms to the ethical guidelines of the Declaration of Helsinki, 1975.

RNA extraction and real-time PCR

We used the NP swabs for RNA extraction of SARS-CoV-2 using Qiagen QIAamp viral RNA mini spin Kit (Qiagen®, Germany) according to manufacturer's instructions. The extracted viral RNA was amplified using Real-time PCR (QuantStudio 5 Real-Time PCR System, Applied Biosystems) for detection of SARS-CoV-2.

cDNA synthesis

We used a total of 1 µg of extracted RNA for reverse transcription using High-Capacity cDNA reverse transcription kit (Applied Biosystems, USA) with random hexamer primers provided with the kit. The RNA concentration and purity was estimated by Nanodrop 2000 (Thermo Scientific, USA).

Spike protein amplification using PCR

PCR reaction components

PCR reactions for amplification of partial spike protein of SARS-CoV-2 were performed in 30 µl total volume containing 15 µl (2X) Hot start PCR

master mix (BioLine Scientific, London, UK), 10 pmol of each forward primer "5′-GATGATTTTACAGGCTGCGTTATAGC-3′" and reverse primer "5′-CAAAAGATTGCTGCATTCAGTTGAATC-3′" specifically designed in this study and 1 µl of cDNA. The reaction was performed using Veriti thermal cycler (Applied Biosystems, USA).

PCR thermal profile

The PCR thermal profile consisted of an initial Taq polymerase activation step at 95°C for 1 min followed by 40 cycles of denaturation step at 95°C for 15 seconds, then annealing step at 50°C for 20 sec, and an extension step at 72°C for 45 sec, each with a final extension cycle at 72°C for 10 minutes. The PCR products (950 bp) were visualized through a 1.5% agarose gel stained with ethidium bromide (0.4 µg/ml).

DNA Sequencing

The PCR products were sequenced using the BigDye Terminator V 3.1 Cycle Sequencing Kit (Applied Biosystems) on the automated sequencer ABI prism 310 genetic analyzers (Applied Biosystems). Two-directional Sanger sequencing was performed for 23 successfully amplified PCR products.

Phylogenetic analysis

A phylogenetic analysis was conducted to detect variability and/or similarity among our sequences compared to Wuhan reference strain (Accession no. NC-045512) (GenBank: MN908947.3). Initially, the sequences were cleaned and trimmed through BioEdit software (v.7.2.5) and then aligned by ClustalW multiple alignment. After cleaning and trimming, the final genomic length of sequences ranged from 785 to 920 bp.

Tree construction

A phylogenetic tree of 23 sequences of partial spike (S) protein was created using MEGA X software [17]. The tree was generated by Maximum Likelihood (ML) method [18] with 1000 bootstrap replicates [19]. The evolutionary distances were computed using the Tamura-Nei model [20].

Sequence genotyping

The sequences were genotyped by Audacity*instant* (v.5.0.1) (https://www.epicov.org/) to determine most related genome and Nextclade (https://clades.nextstrain.org/) online tools while mutation detection was performed through CoVsurver mutation online tool endorsed by GISAID.

Statistical analysis

Data were analyzed using R platform in RStudio (v.2.3.4). The main used packages are gt_summary, finalfit, corrplot, qgraph and ggplot2. Categorical data are presented in count and percent while continuous data are summarized in mean and standard deviation (SD) for normal distributed data and in median and interquartile range (IQR) for not normal distributed data. Normality of data was tested statistically by *Shapiro-Wilk's* test. Spearman correlation coefficient was used to assess the relation between age, clinical symptoms, O_2 saturation and laboratory data and correlation between detected mutations. For categorical data, chi-square or Fisher Exact tests were as appropriate. For continuous data, Wilcoxon rank sum test to compare 2 groups or Kruskal-Wallis to compare 3 or more groups were used for non-parametric data while T-test to compare 2 groups or ANOVA to compare 3 groups were used for parametric data. *P* values at alpha level <0.05 was considered statistically significant. Spearman's correlation was used to test correlation between different parameters.

Results

Among the 50 COVID-19 patients, 18 (36%) were males with a median age of 47.0 (31.2-59.0) years and 32 (64%) were females with a median age of 29.0 (22.0-50.0), while the median (IQR) of age for all patients was 34.0 (22.0-54.8). There was a statistically significant difference in age median between males and females (***p*= 0.036**). Regarding clinical symptoms, the most common (94%) presenting symptom was fever followed by cough (76%). Concerning laboratory data and oxygen (O_2) saturation, there was no statistically significant difference between males and females except for hemoglobin level (***p*= 0.003**) (**Table 1)**.

Correlation analysis between patients' parameters

Correlation analysis by Spearman's method (Figure 1) showed that several parameters are significantly correlated among the patients. Age and O_2 saturation are significantly negative correlated (rho= -0.57, *p*= **0.0022**). O_2 saturation and lymphocytes are positively correlated (rho= 0.56, *p*= **0.0034**). D-dimer is statistically positive correlated with several parameters including ferritin (rho= 0.66, p= **0.000029**), IL-6 (rho= 0.81, *p* **<0.000001**), LDH (rho= 0.66, *p*= **0.00003**), and CRP (rho= 0.64, p= **0.000007**). In addition to that, ferritin is statistically positive correlated with IL-6 (rho= 0.57,

p= **0.002**), CRP (rho= 0.58, p= **0.0014**, and LDH (rho= 0.6, *p*= **0.00052**). Moreover, IL-6 is positively correlated with LDH (rho= 0.7, *p*= **0.000002**) and CRP (rho= 0.61, *p*= **0.0003**).

Sequence genotyping

The partial spike sequences were uploaded to Audacity*instant* (v.5.0.1) endorsed by GISAID (https://www.epicov.org/) to determine their clade, variant and lineage based on the closest related genome. The tool could not assign the nearest lineage for only two sequences while the remaining 21 spike sequences were divided under 4 known Pango lineages (C.36.3, C.36, B.1.1.7, and B1). On the other hand, Nextclade (v.2.12.0) online platform (https://clades.nextstrain.org/) was able to determine lineages for all 23 sequences. There were some discrepancies in lineage assignment between Audacity and Nextclade tools (**Table S1**). According to Nextclade, the most frequent lineages were C.36.3 (13/23, 56.5%) followed by B.1.1.7 (9/23, 39%).

According to GISAID, most sequences (17/23, 74%) were described as old variant meaning that this variant last occurred 100 days ago or more. Only four sequences were labelled recent which means that this variant was detected during the past 100 days but no signs of spreading in the past 30 days. The remaining 2 sequences were referred to as rare meaning that variant occurred once or none (**Table S3**).

Mutation detection by CoVsurver

A total of 17 amino acid changes (mutations) with a total frequency of 102 were detected among our 23 partial spike sequences (V445G, L452R, N501Y, R567K, A570D, D614G, E619K, D627N, Y660H, Q677H (674), P681H (674), T716I, T719S, I720N, S721T, T747P, and S750K) according to the CoVsurver online tool (https://corona.bii.a-star.edu.sg/) enabled by GISAID. Among sequenced isolates, the mutation counts per isolate ranges from 2 to 8 mutations/isolate (**Table S2**). The most frequent mutations are D614G (23/23, 100%), followed by Q677H (14/23, 61%), and L452R (10/23, 43.5%).

Among the isolates, a smaller number of mutations (≤ 2 per sequence) was observed only in the amplified region in case of recent variants, on the other hand, the variants described as old showed multiple mutations (≥ 3 mutations) per sequence (**Table S3**). This difference was found statistically significant between old variants versus recent variants (^{FE}**p= 0.003**).

Distribution of mutations versus lineages

Figure 2 shows a heatmap to demonstrate the occurrence of detected mutations among the Pango lineages according to Nextclade. Most of our spike sequences (22/23, 95.6%) fall under 2 main lineages named C.36.3 (13/23, 56.5%) and B.1.1.7 (9/23, 39%). The most common mutations detected among the C.36.3 lineages are D614G and Q677H (100% each), followed by L452R (77%). On the other hand, the most common mutations detected among the B.1.1.7 lineages are N501Y, A570D, D614G, P681H, and T716I (100% each), followed by D627N (67%) then R567K (22%).

Correlation analysis between mutations

Figure S1 shows network analysis to demonstrate correlation between different mutations in spike protein. The green lines represent positive correlation while the red lines represent negative correlations. The color intensity represents the degree of correlation. The more intense color indicates a higher correlation while faded colors describe low or no correlation.

The second most frequent mutation "Q677H" was perfectly negative correlated (rho= -1, ***p*<0.00001**) with the 4 following mutations: N501Y, A570D, P681H, and T716I, while those 4 mutations were perfectly positive correlated to each other (rho= 1, ***p*= <0.00001**). Similarly, the third most frequent mutation "L452R" was strongly negative correlated (rho= -0.7, ***p*= -0.019**) with the same 4 mutations. The D627N was positively correlated (rho= 0.74, ***p*= 0.0057**) with the following mutations: N501Y, P681H, and T716I. Moreover, several strong positive correlations were found among the following pairs of mutations, Q677H and L452R (rho= 0.7, ***p*= 0.019**), R567K and E619K (rho= 0.69, ***p*= 0.026**).

Phylogenetic analysis

According to the ML phylogenetic tree, the 23 partial spike sequences were clustered in two main domains: group 1 encompassing 14 branches (all C.36.3 Pango lineage and B.1.1) and group 2 encompassing 9 branches (all B.1.1.7 Pango lineage) **(Figure 3)**.

We compared mutation occurrence between these two groups. We found that the occurrences of L452R and Q677H were significantly (*p*= 0.002 and <0.001, respectively) associated with group 1. On the other hand, we

revealed that the occurrence of several other mutations including N501Y, A570D, D627N, P681H, and T716I were significantly (***p* <0.001,** each) associated with group 2 **(Table 2)**.

Association between genotypes and laboratory data

Upon comparing the patients' parameters including laboratory data between group 1 and group 2. Group 1 of isolates (mainly C.36.3) showed significantly lower level of lymphocytes (p=0.012) and WBC (*p*=0.02) compared to group 2 of isolates (B.1.1.7 lineage) **(Table 3)**. Moreover, logistic regression analysis **(Table 4)** confirms that lineage of SARS-CoV-2 is significantly associated with WBC. Majority of patients diagnosed with B.1.1.7 lineage (88.9%) expressed normal WBC (4-10 x 109/L) while most patients diagnosed with C.36.3 lineage (63.4%) expressed abnormal WBC (<4 x 109/L).

Table 1. Summary of demographic and clinical factors for COVID-19 patients included in the study (n= 50).

Variable	**Levels**	**Female,** (n= 32)	**Male,** (n =18)	**Total,** (N= 50)	***p*-value**
Age (in years)	Med. (IQR)	29.0 (22.0-50.0)	47.0 (31.2-59.0)	34.0 (22.0-54.8)	^{W}P=**0.036***
Clinical symptoms					
Fever	No	2 (6.2)	1 (5.6)	3 (6.0)	^{FE}P=1.000
	Yes	30 (93.8)	17 (94.4)	47 (94.0)	
Cough	No	7 (21.9)	5 (27.8)	12 (24.0)	^{X2}P=0.901
	Yes	25 (78.1)	13 (72.2)	38 (76.0)	
Anosmia	No	23 (71.9)	13 (72.2)	36 (72.0)	^{X2}P=1.000
	Yes	9 (28.1)	5 (27.8)	14 (28.0)	
Dyspnea	No	18 (56.2)	9 (50.0)	27 (54.0)	^{X2}P=0.897
	Yes	14 (43.8)	9 (50.0)	23 (46.0)	
O_2 Saturation	Med. (IQR)	98.0 (96.0-98.0)	97.0 (95.2-98.0)	98.0 (96.0-98.0)	^{W}P= 0.699
Laboratory data					
Procalcitonin (ng/mL)	Med. (IQR)	0.4 (0.2-0.6)	0.3 (0.2-0.8)	0.4 (0.2-0.7)	^{W}P= 0.943
D-dimer (µg/mL)	Med. (IQR)	0.4 (0.3-0.6)	0.5 (0.4-0.8)	0.4 (0.3-0.7)	^{W}P= 0.1
Ferritin (ng/mL)	Med. (IQR)	88.5 (64.8-277.0)	93.0 (41.0-404.0)	89.5 (50.5-324.2)	^{W}P= 0.793
IL-6 (pg/mL)	Med. (IQR)	3.0 (2.0-5.4)	4.2 (2.0-11.0)	3.0 (2.0-6.9)	^{W}P= 0.745
Hb (g/dL)	Mean ± SD	11.9 ± 1.4	13.3 ± 1.6	12.4 ± 1.6	^{T}P= **0.003***
PLTs (x 10^3/µL)	Mean ± SD	227.5 ± 79.5	192.2 ± 77.6	214.8 ± 79.9	^{T}P= 0.135
WBC (x 10^3/µL)	Med. (IQR)	4.7 (4.2-7.5)	5.6 (4.5-7.9)	5.3 (4.2-7.5)	^{W}P=0.385
Lymphocytes (x 10^3/µL)	Med. (IQR)	1.1 (0.8-1.7)	1.4 (0.9-1.8)	1.2 (0.9-1.7)	^{W}P=0.412
LDH (U/L)	Med. (IQR)	302.0 (240.5-384.8)	288.0 (247.5-381.2)	301.5 (244.2-394.2)	^{W}P=0.895
CRP (mg/L)	Med. (IQR)	8.0 (5.6-15.2)	6.3 (4.9-13.1)	7.6 (5.0-14.6)	^{W}P=0.708

IL-6: Interleukin-6, WBC: White Blood Count, Hb: Hemoglobin, PLTs: Platelet count, LDH: Lactate dehydrogenase, CRP: C-reactive protein. Data is summarized in count and percent for categorical variables. In case of continuous variables, mean and SD are used for normal distributed data and median (interquartile range) for not normal distributed data. **Bold with asterisk *** is statistically significant with *p*-value < 0.05. ^{W}P: Wilcoxon rank sum test, ^{FE}P: Fisher Exact test, ^{X2}P: Chi-square test, ^{T}P: T-test.

Table 2. Association of spike mutations and genotypes.

Mutations	Levels	Group 1 n= 14	Group 2 n= 9	^{FE}p- value
D614G	Absent	0 (0.0)	0 (0.0)	NA
	Present	14 (100.0)	9 (100.0)	
V445G	Absent	14 (100.0)	8 (88.9)	0.391
	Present	0 (0.0)	1 (11.1)	
L452R	Absent	4 (28.6)	9 (100.0)	**0.002**
	Present	10 (71.4)	0 (0.0)	
N501Y	Absent	14 (100.0)	0 (0.0)	**<0.001**
	Present	0 (0.0)	9 (100.0)	
R567K	Absent	14 (100.0)	7 (77.8)	0.142
	Present	0 (0.0)	2 (22.2)	
A570D	Absent	14 (100.0)	0 (0.0)	**<0.001**
	Present	0 (0.0)	9 (100.0)	
E619K	Absent	14 (100.0)	8 (88.9)	0.391
	Present	0 (0.0)	1 (11.1)	
D627N	Absent	14 (100.0)	3 (33.3)	**0.001**
	Present	0 (0.0)	6 (66.7)	
Y660H	Absent	12 (85.7)	9 (100.0)	0.502
	Present	2 (14.3)	0 (0.0)	
Q677H	Absent	0 (0.0)	9 (100.0)	**<0.001**
	Present	14 (100.0)	0 (0.0)	
P681H	Absent	14 (100.0)	0 (0.0)	**<0.001**
	Present	0 (0.0)	9 (100.0)	
T716I	Absent	14 (100.0)	0 (0.0)	**<0.001**
	Present	0 (0.0)	9 (100.0)	
T719S	Absent	13 (92.9)	9 (100.0)	1
	Present	1 (7.1)	0 (0.0)	
I720N	Absent	13 (92.9)	9 (100.0)	1
	Present	1 (7.1)	0 (0.0)	
S721T	Absent	13 (92.9)	9 (100.0)	1
	Present	1 (7.1)	0 (0.0)	
T747P	Absent	12 (85.7)	8 (88.9)	1
	Present	2 (14.3)	1 (11.1)	
S750K	Absent	13 (92.9)	9 (100.0)	1
	Present	1 (7.1)	0 (0.0)	

Data are presented in count and percent for categorical variables. NA: not applicable.
Group 1: all C.36.3 and B.1.1 lineages (n= 14), Group 2: all B.1.1.7 lineages (n= 9).
^{FE}p: Fisher Exact test for count data. Bold *p*-values are statistically significantly different between two groups at alpha level ≤ 0.05.

Table 3. Association between patients' parameters and genotypes.

Variable	Levels	Group 1	Group 2	*p*- value
	N= 23	n= 14	n= 9	
Gender	Female	9 (64.3)	6 (66.7)	^{FE}p=1
	Male	5 (35.7)	3 (33.3)	
Age (years)	Median (IQR)	22.0 (20.2 to 33.5)	29.0 (22.0 to 53.0)	^{W}p=0.216
Fever	No	2 (14.3)	0 (0.0)	^{FE}p=0.502
	yes	12 (85.7)	9 (100.0)	
Cough	No	2 (14.3)	3 (33.3)	^{FE}p=0.343
	yes	12 (85.7)	6 (66.7)	
Anosmia	No	11 (78.6)	8 (88.9)	^{FE}p=1
	yes	3 (21.4)	1 (11.1)	
Dyspnea	No	6 (42.9)	6 (66.7)	^{FE}p=0.4
	yes	8 (57.1)	3 (33.3)	
O_2 saturation	Median (IQR)	98.0 (96.2 to 98.0)	98.0 (96.0 to 98.0)	^{W}p=0.867
D-dimer	Median (IQR)	0.4 (0.3 to 0.9)	0.4 (0.3 to 0.5)	^{W}p=0.412
Ferritin	Median (IQR)	82.0 (40.8 to 397.8)	115.0 (75.0 to 255.0)	^{W}p=0.875
IL-6	Median (IQR)	3.0 (1.6 to 11.2)	3.0 (2.0 to 4.0)	^{W}p=0.799
Procalcitonin	Median (IQR)	0.3 (0.2 to 0.6)	0.7 (0.2 to 0.9)	^{W}p=0.278
Hb	Median (IQR)	12.2 (11.6 to 12.9)	12.6 (10.8 to 13.5)	^{T}p=0.840
PLTs	Median (IQR)	222.0 (196.2 to 282.0)	235.0 (180.0 to 338.0)	^{T}**p**=0.958
WBC	Median (IQR)	3.8 (3.2 to 4.7)	6.7 (4.8 to 7.8)	**^{W}p=0.02***
Lymphocytes	Median (IQR)	1.0 (0.5 to 1.4)	1.7 (1.4 to 1.8)	**^{W}p=0.012***
LDH	Median (IQR)	271.0 (241.2 to 468.8)	302.0 (245.0 to 350.0)	^{W}p=0.571
CRP	Median (IQR)	6.4 (4.4 to 26.0)	11.0 (5.0 to 16.0)	^{W}p=0.777

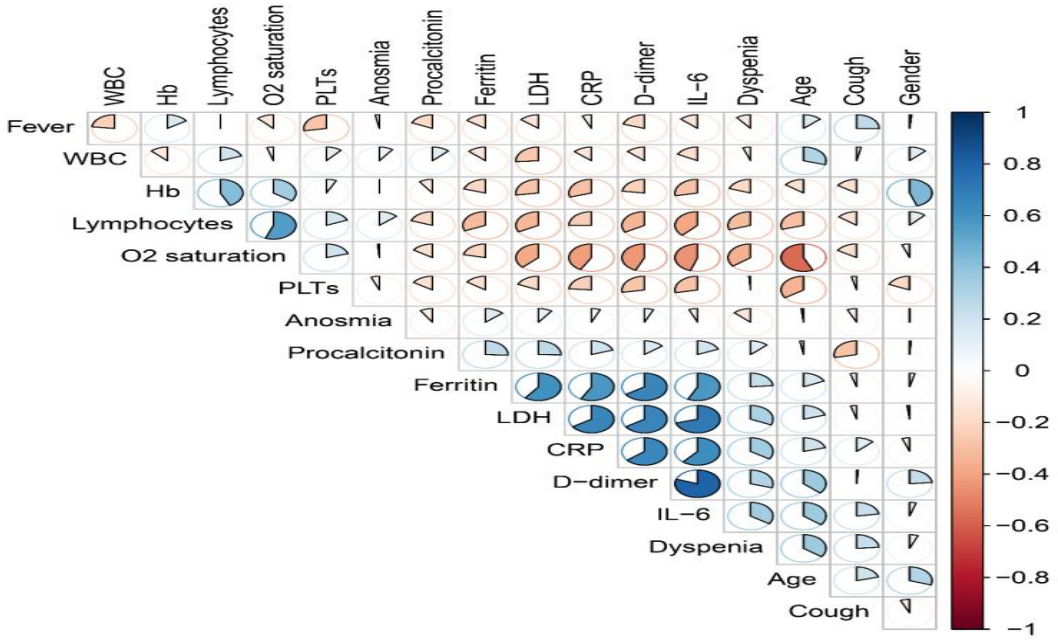
Data are presented in count and percent for categorical variables and in median (interquartile range) for continuous variables. IL-6: Interleukin-6, WBC: White Blood Count, Hb: Hemoglobin, PLTs: Platelet count, LDH: Lactate dehydrogenase, CRP: C-reactive protein. Group 1: all C.36.3 and B.1.1 lineages (n= 14)., Group 2: all B.1.1.7 lineages (n= 9).
^{W}P: Wilcoxon rank sum test, ^{FE}P: Fisher Exact test, and ^{T}P: T-test. **Bold** *p*-values are statistically significantly different between two groups at alpha level ≤ 0.05.

Table 4. Logistic regression analysis to assess relation between lineages and WBC among patients with SARS-CoV-2.

	Abnormal WBC	Normal WBC	OR (univariable)	OR (multivariable)
Gender				
Female	7 (46.7)	8 (53.3)	-	-
Male	3 (37.5)	5 (62.5)	1.46 (0.26-9.31, p=0.673)	1.42 (0.15-13.85, p=0.751)
Age, Mean (SD)	28.2 (13.2)	37.0 (18.6)	1.04 (0.98-1.11, p=0.221)	1.03 (0.95-1.12, p=0.513)
Lineages				
Group 1	9 (64.3)	5 (35.7)	-	-
Group 2	1 (11.1)	8 (88.9)	14.40 (1.87-310.40, p=**0.026**)	13.22 (1.60-296.58, p=**0.035**)

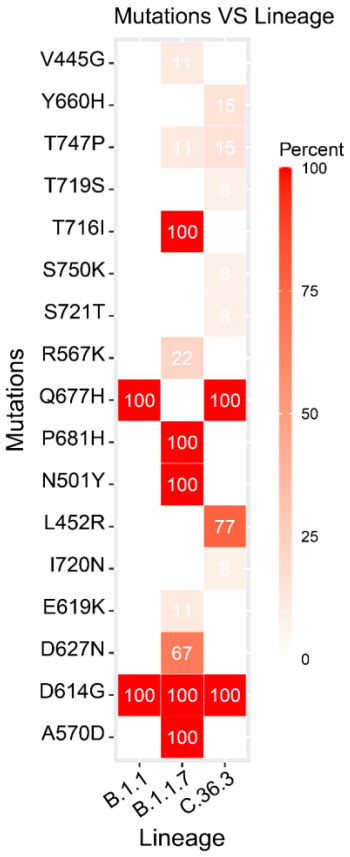
WBC: white blood count, SD: standard deviation, OR: odds ratio

Figure 1. Correlation matrix showing magnitude of correlation between multiple factors including age, gender, clinical symptoms, and laboratory data for all 50 COVID-19 patients included in the study.



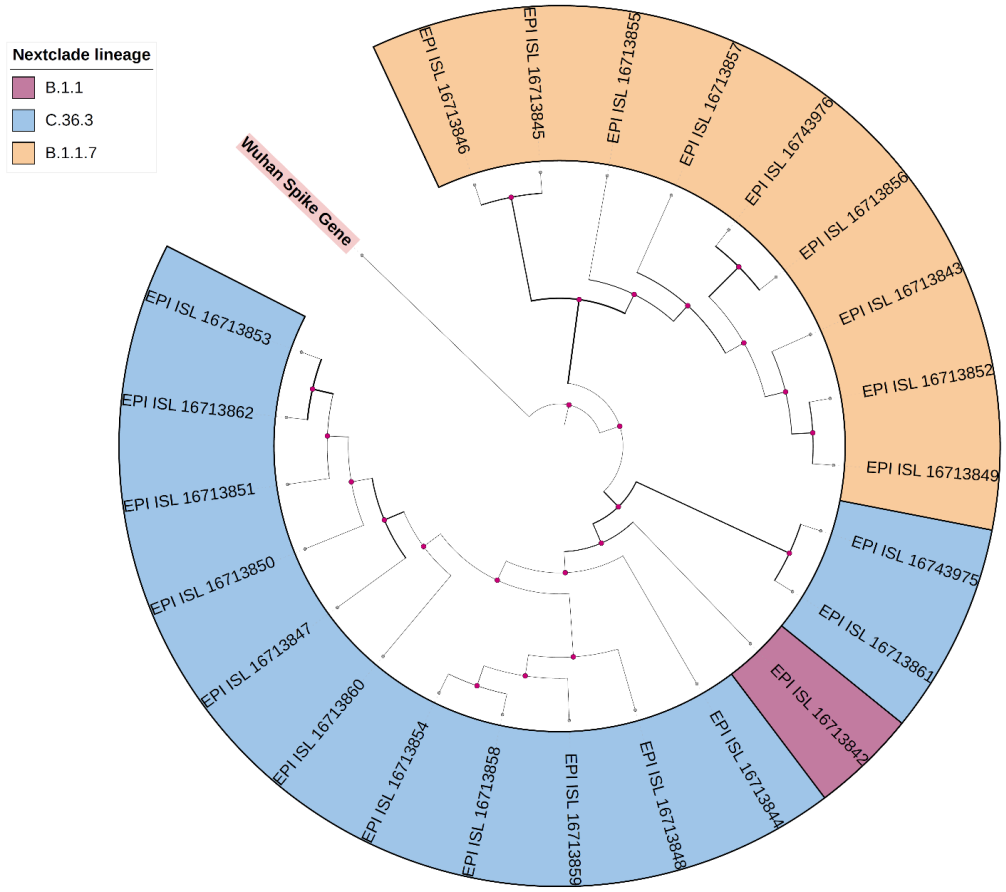
Blue circles indicate positive correlations while red circles indicate negative correlations between different parameters. The size of the pie represents magnitude of correlation. Abbreviations: IL-6: Interleukin 6, WBC: White Blood Count, Hb: Hemoglobin, PLTs: Platelet count, LDH: Lactate dehydrogenase, CRP: C-reactive protein.

Figure 2. Heat map showing the occurrence of mutations by percentage among Pango lineages determined by Next clade.



Data was last accessed in February 28, 2023 by Next clade. Darker red color represents higher percent of mutation occurrence.

Figure 3. A maximum likelihood (ML) phylogenetic tree demonstrating clustering of the 23 partial spike sequences.



Blue strip color indicates C.36.3 Pango lineage while Black strip color indicates B.1.1.7 Pang lineage and finally red strip color indicates B.1.1 Pango lineage according to Next clade online platform (https://clades.nextstrain.org/). Last accessed February 28, 2023.

Discussion

COVID-19 is a serious infection that has a significant impact on health worldwide. Globally, there are several variants of SARS-CoV-2 that have been emerged with multiple types of mutations [21]. Those variants have been reported with different types of mutations that may affect disease patterns, transmission and escaping the host immune defenses [22, 23]. The process of developing an effective antiviral drugs and vaccination may be hampered by such high production rate of variants and mutations.

In this study, most patients were females (64%) with a younger median age of 29.0 (22.0-50.0) years compared to males 47.0 (31.2-59.0) years. In contrast to **Alotaibi et al.** who reported a higher percentage of infection among males than females [24]. The controversy could be attributed to sex-based immunological changes that contribute to differences in the susceptibility to infectious diseases [25, 26]. In addition to the small sample size of and narrow time frame of the current study.

The clinical symptoms of COVID-19 range from asymptomatic presentations to dizziness, cough, fever, and dyspnea. The most common clinical symptoms in this study were fever (94%) and cough (76%). Similar results were reported in Saudi Arabia, fever was detected in 84.1% and cough in 64.4% among their patients [27]. Meanwhile in Egypt **Alotaibi et al.** reported fever in 92%, and cough in 66% among their study group [24]. According to meta-analysis done by **Rodriguez-Morales et al.,** fever (88.7%) and cough (57.6%) were reported as the most common presenting symptoms [28].

On the other hand, a study conducted in Turkey reported that cough (56.6%) was the most common symptom, followed by anosmia (35.7%), and fever (33.6%) [29]. In Iran, **Mamishi et al**. documented fever (91%) and abdominal pain (58%) among other symptoms as the main clinical symptoms among pediatric COVID-19 patients [30].

In this study, no significant association between clinical symptoms and gender was detected. Similarly, **Ahmed et al**. reported no significant association of clinical symptoms with gender among their cohort [31]. In contrast to this, **Ancochea et al**. reported significant association between clinical symptoms and male gender in Italy [32]. Such controversies may be attributed to physiological changes between male and female patients [33], or the small sample size and the fact that more females were enrolled in the current study.

According to the WHO, variants with amino acid substitutions associated with confirmed phenotypic impact are recognized as variants of interest (VOI) such as e.g., Eta, Iota, Kappa and Lambda., while variants associated with higher transmission rate, more virulence and severe clinical picture are referred to as variants of concern (VOC) such as Alpha, Beta, Gamma, Delta and most recently, Omicron [33,34]. Variant under monitoring (VUM) are variants with genetic changes that are suspected to affect virus characteristics compared to other circulating ones, however, their effect is still unclear [34,35]. The previous classification of variants is revised and updated regularly by WHO.

During the current era of COVID-19, it is essential to perform the whole genome sequencing (WGS), or at least to sequence either the complete or partial Spike gene to confirm the identity of a variant [36]. In fact, Sanger sequencing of the partial spike gene is considered a more feasible approach compared to WGS.

We analyzed the partial spike gene variants of 23 sequences from Egyptian patients during the first wave of the pandemic and categorized them into two groups according to phylogenetic analysis. Based on the branching of ML phylogenetic tree, the first group (group 1) included all C.36.3 and B.1.1 lineages while group 2 included all B.1.1.7 lineage.

The C.36 lineage has been circulating since the early phases of COVID-19 pandemic. It evolved into multiple sub-lineages, including C.36.1, C.36.3, and C.36.3.1, across the Egyptian patients' genome [37]. In the current study, according to Nextclade; the most frequent lineage was C.36.3 (56.5%) while the second most frequent lineage was B.1.17 (39%) which classified as alpha (VOC) by WHO.

Multiple studies were carried out to determine the SARS-CoV-2 protein mutations in different countries across the world including Egypt. Up to now, several mutations of interest have occurred in the RBD of the spike gene of SARS-CoV-2. That is why, in this study, we sequenced the partial spike gene to detect prevalence of these mutations among Egyptian COVD-19 patients.

In early 2020, the first known mutation associated with higher transmissibility was D614G. Since then, the virus has evolved; new mutations have occurred, and many variants have been described. All 23 sequences in this study expressed the spike aspartic acid 614 glycine (D614G) substitution which is highly prevalent in global SARS-CoV-2 strains associated with severe acute infection [38].

A total of 23 isolates were successfully amplified for sequencing. This may be attributed to several factors including faint bands during gel electrophoresis, low viral load or altered amino acids in the primer binding sites leading poor annealing of used primers.

Among the 23 samples sequenced in the current study, 17 mutations in S protein were detected. Generally, the multiple variants and increased number of mutations are due to the lack of replication fidelity and proof-reading mechanism that is pathognomonic of majority of RNA viruses [11, 39]. Despite of that, these mutations in SARS-CoV-2 were considered low compared to that occurring in other RNA viruses such as influenza viruses [40].

The emergence of a non-synonymous mutation at position 614 (D614G) that was rare before March 2020 and then became more common as the pandemic spread by June 2020. This mutation can lead to enhanced transmission, replication and infectivity of the virus in respiratory tract cells by enhancing viral entry and increasing the infectious titer [41]. This mutation was the most frequent mutation detected among the 23 sequenced partial spike protein (100%) in our study.

In agreement with this, D614G mutation was detected in 100% of studied cases in numerous countries such as Algeria, Morocco, Democratic Republic of Congo, and USA [23,38,42,43]. D614G was also reported by USA, Saudi Arabia, Qatar, Spain and Middle East and North Africa (MENA) with high frequencies, 99.3%, 98.3%, 84.2% 81.5%, 95% and 73.7%, respectively [44-48].

The second most frequently detected mutation in the present study was Q677H [49] which was found in 60.9% of sequenced partial spike

protein. This finding agrees with other studies conducted in different regions where Q677H mutation was the second most frequent mutation detected after D614G, but with a lower frequency of 8.4% in Egypt [46]. It was found that this mutation is responsible for increasing infectivity of SARS-CoV-2, in addition to its ability of escaping the host immune response by resisting the neutralizing antibodies [38,39,50]. It may also affect the cell entry due to its close contact to the polybasic cleavage site and its location is adjacent to the junction that connects S1 and S2 [43].

The third most frequent mutation detected in this study was L452R [51] in 52.2% of sequenced partial spike protein, in which leucine amino acid changed to arginine at position 452. Its prevalence was slightly lower compared to 65.9% reported in USA [52]. This mutation was observed to influence the transmission, infectivity of the virus and disease outcome by evading host immunity and enhancing adsorption of virus to the host cells.

A percentage of 39.1% of the sequenced cases were found for each of other mutations detected, namely, A570D, N501Y, P681H, and T716I. The A570D mutation was shown to increase the transmission rate and infectivity of the virus by switching on or off the RBD [53]. Sabir reported the presence of this mutation among 53% of Iraqi patients [54]. A570D was reported for the first time in UK and its prevalence increased from 0.1% to 49.7% in one month only [55] in South Africa and Brazil [56].

The following mutations: A570D, N501Y, P681H and T716I were positively correlated with each other in this study. In a previous study, conducted in Pakistan, a total of 93.5% of their sequences showed the same pattern. These mutations belong to B.1.1.7 variant in this study [57].

In the current study, WBC and lymphocytes were the only laboratory data associated with sequence genotypes. B.1.1.7 lineage was associated with higher WBC and lymphocytes compared to other lineages (p= 0.02, p=0.012, respectively). This was confirmed with logistic regression analysis. This may indicate that lineage may affect WBC and accordingly affect the immune response and prognosis to combat the infection. In the light of this, further studies with larger sample sizes are needed to confirm this association.

One limitation of the current study is the small sample size due to inability to amplify all included isolates. Therefore, we recommend a larger scale study with follow up of patients would be beneficial. Using different sets of primers for better amplification rate and to amplify larger part of S protein to cover further mutations in the future. Also, we recommend that future studies include more confounder variables such as common comorbidities like diabetes. We should implement sequencing on a routine basis to identify circulating variants in Egypt and to determine their impact on viral transmission and virulence.

Conclusions and recommendations

To the best of our knowledge, this is the first study to correlate SARS-CoV-2 spike protein mutations with laboratory data in Alexandria, Egypt. The main conclusions are low heterogeneity among the 23 sequenced isolates that was observed based on the phylogenetic tree. Recent circulating variants have fewer spike mutations (2 or fewer) compared to old variants (3 or more). The diversity of mutations in SARS-CoV-2 may enhance virulence and transmission of the virus. Level of WBC and/or lymphocytes may significantly differ according to infecting lineage. This finding may impact progression of SARS-CoV-2 infection.

Declarations

Funding

No funding was received for conducting this study.

Conflict of interest

The authors declare no conflict of interest

Availability of data and materials

The authors confirm that the data supporting the results and findings of this study are available within the article. The sequencing data are accessible through GenBank NCBI with the following accession numbers: QQ346382:QQ346403 and QQ347969. Also, we submitted them to GISAID database with the following accession numbers: EPI_ISL_16713842: EPI_ISL_16713862 and EPI_ISL_16743975: EPI_ISL_16743976.

Ethics statement and consent to participate

The authors confirm that the ethical policies of the journal, as noted on the journal's author guidelines page, have been adhered to and the Medical Research Institute ethical review committee approval has been received with the following approval number IORG0008812.

Standards of reporting

The study conforms to the Alexandria University guidelines.

Author's contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by [Ali N. M. Gubran], [Dalia Metwally Ragab], [Ezz Eldein Anwar Saleh] and [Iman S. Naga]. Samples collection was provided by [Reham Abo Elwafa]. Statistical analysis was performed by [Rasha Emad]. The first draft of the manuscript was written by [Iman S. Naga] and [Rasha Emad] and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

1- **WHO.** WHO COVID-19 dashboard. Available at : thttps://covid19.who.int/.

2- **Walls AC, Tortorici MA, Bosch B-J, Frenz B, Rottier PJM, DiMaio F, et al.** Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. Nature. 2016;531(7592):114-7.

3- **Yuan Y, Cao D, Zhang Y, Ma J, Qi J, Wang Q, et al.** Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. Nature Communications. 2017;8(1):15092.

4- **Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG, Decroly E.** The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. Antiviral Res. 2020;176:104742.

5- **Huang Y, Yang C, Xu X-f, Xu W, Liu S-w.** Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. Acta Pharmacologica Sinica. 2020;41(9):1141-9.

6- **Letko M, Marzi A, Munster V.** Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nature Microbiology. 2020;5(4):562-9.

7- **Watanabe Y, Allen JD, Wrapp D, McLellan JS, Crispin M.** Site-specific glycan analysis of the SARS-CoV-2 spike. Science. 2020;369(6501):330-3.

8- **Li F.** Structure, Function, and Evolution of Coronavirus Spike Proteins. Annu Rev Virol. 2016;3(1):237-61.

9- **Pillay TS.** Gene of the month: the 2019-nCoV/SARS-CoV-2 novel coronavirus spike protein. J Clin Pathol. 2020;73(7):366-9.

10- **Thanh Le T, Andreadakis Z, Kumar A, Gómez Román R, Tollefsen S, Saville M, et al.** The COVID-19 vaccine development landscape. Nat Rev Drug Discov. 2020;19(5):305-6.

11- **Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, et al.** Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. J Transl Med. 2020;18(1):179.

12- **Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, et al.** SARS-CoV-2 variants, spike mutations and immune escape. Nat Rev Microbiol. 2021;19(7):409-24.

13- **Zhang J, Cai Y, Xiao T, Lu J, Peng H, Sterling SM, et al.** Structural impact on SARS-CoV-2 spike protein by D614G substitution. Science. 2021;372(6541):525-30.

14- **Lamptey J, Oyelami FO, Owusu M, Nkrumah B, Idowu PO, Adu-Gyamfi EA, et al.** Genomic and epidemiological characteristics of SARS-CoV-2 in Africa. PLoS Negl Trop Dis. 2021;15(4):e0009335.

15- **Weissman D, Alameh MG, de Silva T, Collini P, Hornsby H, Brown R, et al**. D614G Spike Mutation Increases SARS CoV-

2 Susceptibility to Neutralization. Cell Host Microbe. 2021;29(1):23-31.e4.

16- **Muttineni R, R NB, Putty K, Marapakala K, K PS, Panyam J, et al.** SARS-CoV-2 variants and spike mutations involved in second wave of COVID-19 pandemic in India. Transbound Emerg Dis. 2022;69(5):e1721-e33.

17- **Kumar S, Stecher G, Li M, Knyaz C, Tamura K**. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018;35(6):1547-9

18- **Kimura M.** A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of molecular evolution. 1980;16(2):111-20.

19- **Felsenstein J.** CONFIDENCE LIMITS ON PHYLOGENIES: AN APPROACH USING THE BOOTSTRAP. Evolution. 1985;39(4):783-91.

20- **Tamura K, Nei M.** Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 1993;10(3):512-26.

21- **European Centre for Disease Prevention and Control (ECDC).** SARS-CoV-2 variants of concern. ECDC 2021. Available at https://www.ecdc.europa.eu/en/covid-19/variants-concern.

22- **Frampton D, Rampling T, Cross A, Bailey H, Heaney J, Byott M, et al .** Genomic characteristics and clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in London, UK: a whole-genome sequencing and hospital-based cohort study. . The Lancet Infectious diseases. 2021;21(9):1246–56.

23- **Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, et al**. Spike mutation D614G alters SARS-CoV-2 fitness. Nature. 2021;592(7852):116–21.

24- **Alotaibi B, El-Masry TA, Seadawy MG, Farghali MH, El-Harty BE, Saleh A, et al.** SARS-CoV-2 in Egypt: epidemiology, clinical characterization and bioinformatics analysis. Heliyon. 2022;8(2):e08864.

25- **Bwire GM.** Coronavirus: Why Men are More Vulnerable to Covid-19 Than Women? SN Compr Clin Med. 2020;2(7):874-6.

26- **Elgendy IY, Pepine CJ.** Why are women better protected from COVID-19: Clues for men? Sex and COVID-19. Int J Cardiol. 2020;315:105-6.

27- **Ibrahim ME, Al-Aklobi OS, Abomughaid MM, Al-Ghamdi MA.** Epidemiological, clinical, and laboratory findings for patients of different age groups with confirmed coronavirus disease 2019 (COVID-19) in a hospital in Saudi Arabia. PLoS One. 2021;16(4):e0250955.

28- **Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, Villamizar-Peña R, Holguin-Rivera Y, Escalera-Antezana JP, et al.** Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis. Travel Med Infect Dis. 2020;34:101623.

29- **Çalıca Utku A, Budak G, Karabay O, Güçlü E, Okan HD, Vatan A.** Main symptoms in patients presenting in the COVID-19 period. Scott Med J. 2020;65(4):127-32.

30- **Mamishi S, Movahedi Z, Mohammadi M, Ziaee V, Khodabandeh M, Abdolsalehi MR, et al.** Multisystem inflammatory syndrome associated with SARS-CoV-2 infection in 45 children: a first report from Iran. Epidemiol Infect. 2020;148:e196

31- **Ahmed MM, Sayed AM, El Abd D, Fares S, Said MSM, Elsayed Sedik Ebrahim E.**

Gender Difference in Perceived Symptoms and Laboratory Investigations in Suspected and Confirmed COVID-19 Cases: A Retrospective Study. J Prim Care Community Health. 2021;12:21501327211039718.

32- **Ancochea J, Izquierdo JL, Soriano JB.** Evidence of Gender Differences in the Diagnosis and Management of Coronavirus Disease 2019 Patients: An Analysis of Electronic Health Records Using Natural Language Processing and Machine Learning. J Womens Health (Larchmt). 2021;30(3):393-404.

33- **Tao K, Tzou PL, Nouhin J, Gupta RK, de Oliveira T, Kosakovsky Pond SL, et al.** The biological and clinical significance of emerging SARS-CoV-2 variants. Nat Rev Genet. 2021;22(12):757-73.

34- **Candido KL, Eich CR, de Fariña LO, Kadowaki MK, da Conceição Silva JL, Maller A, et al.** Spike protein of SARS-CoV-2 variants: a brief review and practical implications. Braz J Microbiol. 2022;53(3):1133-57.

35- **WHO.** World Health Organization: Updated working definitions and primary actions for SARS-CoV-2 variants 15 March 2023. Available at https://www.who.int/publications/m/item/updated-working-definitions-and-primary-actions-for--sars-cov-2-variants.

36- **European Centre for Disease Prevention and Control**. Sequencing of SARS-CoV-2: first update 2021. Available at: https://www.ecdc.europa.eu/sites/default/files/documents/Sequencing-of-SARS-CoV-2-first-update.pdf.

37- **Roshdy WH, Khalifa MK, San JE, Tegally H, Wilkinson E, Showky S, et al.** SARS-CoV-2 Genetic Diversity and Lineage Dynamics in Egypt during the First 18 Months of the Pandemic. Viruses. 2022;14(9)

38- **Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al.** Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell. 2020;182(4):812-27.e19

39- **Yao Z, Zhang L, Duan Y, Tang X, Lu J.** Molecular insights into the adaptive evolution of SARS-CoV-2 spike protein. J Infect. 2024;88(3):106121

40- **Manzanares-Meza LD, Medina-Contreras O.** SARS-CoV-2 and influenza: a comparative overview and treatment implications. Bol Med Hosp Infant Mex. 2020;77(5):262-73.

41- **Yurkovetskiy L, Wang X, Pascal KE, Tomkins-Tinch C, Nyalile TP, Wang Y, et al.** Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. Cell. 2020;183(3):739-51.e8.

42- **Grubaugh ND, Hanage WP, Rasmussen AL.** Making Sense of Mutation: What D614G Means for the COVID-19 Pandemic Remains Unclear. Cell. 2020;182(4):794-5.

43- **Lau SY, Wang P, Mok BW, Zhang AJ, Chu H, Lee AC, et al.** Attenuated SARS-CoV-2 variants with deletions at the S1/S2 junction. Emerg Microbes Infect. 2020;9(1):837-42.

44- **Pandey U, Yee R, Shen L, Judkins AR, Bootwalla M, Ryutov A, et al.** High prevalence of SARS-CoV-2 genetic variation and D614G mutation in pediatric patients with COVID-19. Open Forum Infectious Diseases 2021;8(6):ofaa551.

45- **Benslimane FM, Al Khatib HA, Al-Jamal O, Albatesh D, Boughattas S, Ahmed AA, et al**. One Year of SARS-CoV-2: Genomic Characterization of COVID-19 Outbreak in

Qatar. Front Cell Infect Microbiol. 2021;11:768883.

46- **Sallam M, Ababneh NA, Dababseh D, Bakri FG, Mahafzah A.** Temporal increase in D614G mutation of SARS-CoV-2 in the Middle East and North Africa. Heliyon. 2021;7(1):e06035.

47- **Troyano-Hernáez P, Reinosa R, Holguín Á.** Evolution of SARS-CoV-2 Envelope, Membrane, Nucleocapsid, and Spike Structural Proteins from the Beginning of the Pandemic to September 2020: A Global and Regional Approach by Epidemiological Week. Viruses. 2021;13(2).

48- **Ullah MF, Alnour TMS, Elssaig EH, Ahmed-Abakur EH.** Characterization of altered genomic landscape of SARS-CoV-2 variants isolated in Saudi Arabia in a comparative in silico study. Saudi J Biol Sci. 2021;28(12):6803-7

49- **Niu X, Xu J, Liu M, Tu H, Koenig SN, Saif LJ, et al.** Isolation and characterization of a SARS-CoV-2 variant with a Q677H mutation in the spike protein. Arch Virol. 2022;168(1):5

50- **Zeng C, Evans JP, Faraone JN, Qu P, Zheng YM, Saif L, et al.** Neutralization of SARS-CoV-2 Variants of Concern Harboring Q677H. mBio. 2021;12(5):e0251021.

51- **Kyung K, Ku J, Cho E, Ryu J, Woo J, Jung W, et al.** Fluorometric Detection of SARS-CoV-2 Single-Nucleotide Variant L452R Using Ligation-Based Isothermal Gene Amplification. Bioengineering (Basel). 2023;10(10).

52- **Tchesnokova V, Kulakesara H, Larson L, Bowers V, Rechkina E, Kisiela D, et al.** Acquisition of the L452R mutation in the ACE2-binding interface of Spike protein triggers recent massive expansion of SARS-Cov-2 variants. bioRxiv. 2021DOI 10.1101/2021.02.22.432189.

53- **Yuan M, Huang D, Lee CD, Wu NC, Jackson AM, Zhu X, et al.** Structural and functional ramifications of antigenic drift in recent SARS-CoV-2 variants. Science. 2021;373(6556):818-23.

54- **Sabir DK.** Analysis of SARS-COV2 spike protein variants among Iraqi isolates. Gene Rep. 2022;26:101420.

55- **Leung K, Shum MH, Leung GM, Lam TT, Wu JT.** Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. Euro Surveill. 2021;26(1).

56- **Weisblum Y, Schmidt F, Zhang F, DaSilva J, Poston D, Lorenzi JCC, et al.** Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. bioRxiv 2020; 9:e61312.

57- **Umair M, Salman M, Rehman Z, Badar N, Ali Q, Ahad A, et al.** Proliferation of SARS-CoV-2 B.1.1.7 Variant in Pakistan-A Short Surveillance Account. Front Public Health. 2021;9:683378.