



Prevalence of COVID-19 and Possible Antigenic Drifts in SARS-CoV-2 Spike Protein in Kurdistan Region-Iraq

Karzan R. Sidiq*

Charmo Centre for Research, Training and Consultancy, Charmo University, 46023 Chamchamal, Sulaimani, Kurdistan Region, Iraq

Received 03 April 2021; revised 24 May 2021;
accepted 24 May 2021; available online 02 July 2021

[doi:10.24271/psr.27](https://doi.org/10.24271/psr.27)

ABSTRACT

The COVID-19 is an on-going viral pandemic that has been affecting the public health, routine life and global economy. The disease is caused by a novel strain of coronavirus, called SARS-CoV-2 virus. The COVID-19 outbreak has been reporting in Kurdistan region of Iraq since March 2020. However, few studies investigated the epidemiology of COVID-19 and SARS CoV-2 virus in the region. This study aims at investigating the epidemiological situations of COVID-19 and SARS-CoV-2 variants in Kurdistan region over a year of the pandemic. The results revealed that the prevalence of COVID-19 is 1.9% in the region and is still in parallel to the neighbouring countries and the entire world. The mortality is 59 per 100,000 populations that may be related with age, as 25 % of the patients are older than 50 years old, and underlying health conditions of the patients might be another reason. Meanwhile, the recovery rate is high (90.5%), suggesting a standard medical management of COVID-19 in the region. It was observed that males comprise the greater number of COVID-19 patients. Like other countries, Kurdistan region passed through two waves of COVID-19 and currently tackling the third wave. On the other hand, the investigated spike proteins of nine isolates of SARS-CoV-2 in Kurdistan showed five isolates with single (D614G) mutation and four isolates with multiple amino acid substitutions (A348S, T478K and D614G), (L452R, E583D and D614G) and (N501Y, A570D and D614G), of which A348S, L452R and T478K and N501Y are in the receptor binding domain (RBD). Interestingly, all the isolates in Kurdistan contained D614G mutation. The D614G mutation alone and in combination with other mutations makes the SARS-CoV-2 virus more infectious and transmissible, so virulent variants of the virus is currently circulating and might be the cause of third wave of COVID-19 in Kurdistan region. The detection of N501Y and A570D mutations indicate the circulation of the UK variant of concern in Kurdistan region. Moreover, the results showed that the altered amino acids (A348S, L452R and T478K and N501Y) in the RBD of spike protein are located in the predicted B-cell epitopes. This could possibly reduce the sensitivity of some neutralizing antibodies, produced after infection with the previous variants or after vaccination. This and future investigations of COVID-19 epidemiology and SARS-CoV-2 variants definitely provide insights to the Kurdistan health officials to evaluate, control and predict the course of COVID-19 pandemic and to order the right version of vaccine.

© 2021 Production by the University of Garmian. This is an open access article under the LICENSE

<https://creativecommons.org/licenses/by-nc/4.0/>

Keywords: COVID-19, Prevalence, Mortality rate, Spike protein, Antigenic drift, SARS-CoV-2 variants

1. Introduction

The coronavirus disease 2019 (COVID-19) was first reported from Wuhan/China in December 2019. COVID-19 is a contagious viral infection of respiratory system that caused by a novel strain of Beta-coronavirus, called Sever Acute Respiratory Syndrome- Coronavirus-2 (SARS-CoV-2) [1, 2]. The disease has been threatening the lives of human beings worldwide, so it was recognized as a pandemic disease by World Health Organization (WHO) on 11 March, 2020 [3]. As of 10 March 2021, more than 117 million confirmed cases of the disease and more than 2.5

million deaths were reported by WHO [4].

The incubation period of COVID-19 is up to 14 days with estimated average of 5 to 6 days [5, 6]. The typical symptoms of COVID-19 are fever and dry cough, besides some also have shortness of breath, fatigue, muscle pain, confusion, headache, sore throat, diarrhea, and vomiting [1]. Most of the patients with COVID-19 have mild to moderate symptoms, but approximately 15% develop acute pneumonia and about 5% eventually develop acute respiratory distress syndrome, septic shock, many organ failures and death [7].

SARS-CoV-2 is an enveloped single-stranded RNA virus, which possesses four structural proteins; spike (S), small protein (E), matrix (M), and nucleocapsid (N) [8]. Both the spike (S) and nucleocapsid (N) are key immunogenic proteins of SARS-CoV-2 that induce host immune system [9]. The spikes (S) protein

* Corresponding author

E-mail address: karzan.sidiq@charmouniversity.org (Instructor).

Peer-reviewed under the responsibility of the University of Garmian.

contains receptor binding domain (RBD) that facilitates the SARS-CoV-2 to interact with the receptor angiotensin-converting enzyme 2 (ACE2) on the bronchial and alveolar epithelial cells up on entry^[10, 11]. The spike protein is made up of 1273 amino acid residues and distinguished into S1 and S2 sub-domains. The S1 involves the residues from 1 to 685, while the S2 sub-domain starts from 686 to 1273 residues and is the fusion machine for the virus entry into the target cells^[12, 13]. The S1 subdomain contains the receptor-binding domain (RBD) with 222 residues length that locates between 319 to 541 region of spike protein^[13, 14].

It is known that mutation rates are higher in RNA viruses than DNA viruses^[15, 16]. Unlike flu virus that has multiple-segmented genome, SARS-CoV-2 has a single-segment genome with approximately 30 kilobases length^[17].

Antigenic drift is minor changes in viral antigens as a result of sequential point mutations; whereas antigenic shift is major antigenic changes that occur as a result of genetic reassortment in viruses with multiple-segmented genome (e.g. influenza virus)^[18, 19]. So, spontaneous point mutations in antigen-encoding genes that cause antigenic drift, not antigenic shift, could be the only mean of antigenic variations in SARS-CoV-2 virus. Mutations in the coding-regions mostly change the corresponding amino acid in the proteins. Therefore, any change in the amino acids of the spike protein can produce a new variant and significantly alter viral infectivity and/or interactions with neutralizing antibodies^[12]. There have been several of SARS-CoV-2 variants based on the mutations in the gene encoding spike protein since the onset of the outbreak^[20]. In December 2020, a new variant of SARS-Cov-2 virus was reported in the United Kingdom and was popularly known B.1.1.7 variant (VOC-202012/01 or Kent variant). This variant has dominantly caused COVID-19 infections across the UK because of its enhanced transmissibility^[21]. The COVID-19 outbreak was officially reported in Kurdistan region shortly after the WHO announcement on the 20th of January 2020^[22]. The ministry of health soon started tight health and prevention measures to tackle the spread of the outbreak. The measurement procedures have been instructing people for wearing masks and Gloves, hand washing, social distancing, staying at home, general lockdown and curfew. Despite the above prevention measures, the cases of COVID-19 have been reporting in the region. This study investigates the epidemiological situation of COVID-19 and antigenic drifts in the spike protein of SARS-CoV-2 isolates in Kurdistan region of Iraq.

2. Materials and Methods

2.1. COVID-19 situation data

All the data about the situation of COVID-19 in Kurdistan region-Iraq were collected from the Kurdistan regional government (KRG) official website. These data are PCR confirmed cases of COVID-19 that are continuously reported by the ministry of health from March 2020 to March 2021^[23]. The raw data about population sizes and COVID-19 situations of the world and neighbouring countries were obtained from Worldometer website and WHO COVID-19 dashboard^[4, 24]. These data may be substantially different from the real number of the COVID-19 cases, because a fraction of cases are asymptomatic infections and do not require medical attention.

2.2. Prevalence, recovery rate and mortality rate

The prevalence was calculated by dividing the number of confirmed COVID-19 cases from March 2020 to March 2021 on the total number of the population, and then multiplied by 100. The recovery rate was worked out by dividing the number of COVID-19 recovered patients on the total number of confirmed COVID-19 cases from March 2020 to March 2021, and then multiplied by 100. The mortality rate was found through dividing the number of cases on the total number of population, and then multiplied by 100,000.

2.3. Amino acid sequence of spike proteins and multiple sequence alignment

The spike protein sequences (accession numbers: Wuhan_YP_009724390.1, QXX36542.1 and QRW43499.1) of SARS-CoV-2 isolates were downloaded from National Center for Biotechnological Information (NCBI)^[25]. The multiple alignment of the spike protein sequences was done by Clustal Omega bioinformatics tool (<https://www.ebi.ac.uk/Tools/msa/clustalo/>).

2.4. Epitope prediction

The BepiPred-2.0 web server (<http://tools.iedb.org/bcell/>) was used to predict the Sequential B-Cell Epitopes from spike protein sequence. BepiPred-2.0 is a valuable tool for sequence-based B-cell epitope prediction from both linear and 3D protein structures. It is believed that BepiPred-2.0 will be an important tool for the bioinformatics and immunology communities^[26].

3. Results and Discussion

3.1. Prevalence and Mortality rate of COVID-19 in Kurdistan Region- Iraq

Kurdistan is a self-governing region in the north of Iraq with an estimate population of 6,200,000^[27]. The population distributed over four governorates of Erbil, Sulaymaniyah, Duhok and Halabja (Figure 1). The COVID-19 outbreak was officially reported in Kurdistan region shortly after the WHO announcement in early 2020. Despite continuous prevention measures, the cases of COVID-19 have been reporting. The total number of confirmed COVID-19 cases in the region is 120319 from March 2020 to March 2021. Out of this, 108835 cases got full recovery, while 7831 cases are under treatment and 3653 patients unfortunately died (Figure 2).

Prevalence is the proportion of a population that has a condition or illness over a specified period of time. Prevalence is often expressed as a percentage and population units, such as per 100,000 people, can also be used^[28]. The broad clinical spectrum of COVID-19 from asymptomatic to detected symptomatic cases may limit our ability to get real prevalence of COVID-19 in all the countries around the world.

The prevalence of COVID-19 in Kurdistan region is 1.9% from March 2020 to March 2021 (Table 1). This means that approximately two people infected with COVID-19 in every 100 people. This is almost similar to the prevalence of COVID-19 in Iraq (2.1%), Iran (2.2 %) and the entire world (1.6%), whereas it is lower than the prevalence of COVID-19 in Turkey (3.8%) (Table 1)^[4, 24]. Although differences in the total number of COVID-19 tests carried out in each country may limit making real comparison, these results suggest that the number of COVID-

19 cases is at a secured situation in Kurdistan region. The controlled prevalence of COVID-19 in Kurdistan region is probably associated with the implementation of prevention measures such as self-protection, lockdown and travel ban. It is also possible that the majority of Kurdish population asymptotically passed through herd immunity against COVID-19 before March 2020 [29].

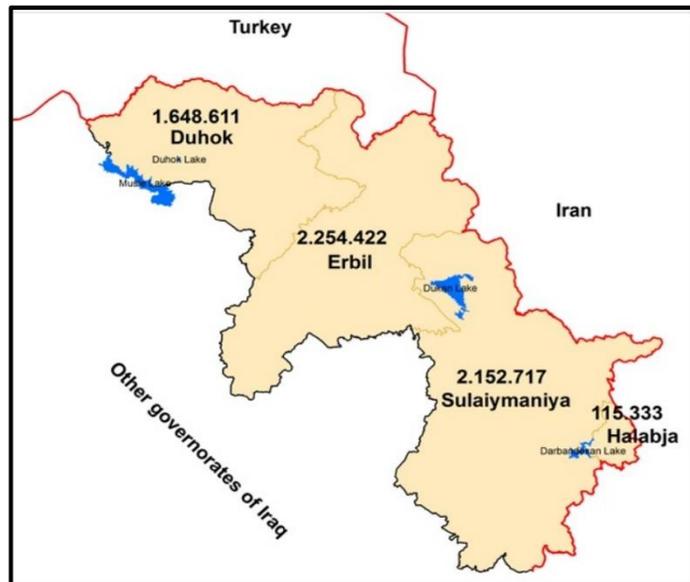


Figure 1: Map of estimated population of Kurdistan region at governorates level in 2020 [27].

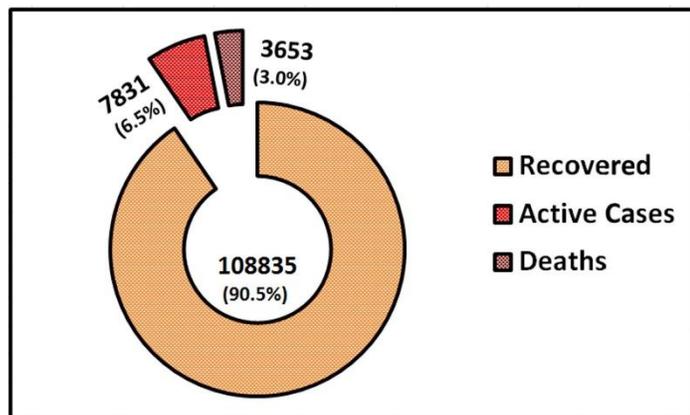


Figure 2: COVID-19 cases outcome of total number of 120319 confirmed cases in Kurdistan region from March 2020 to March 2021.

The mortality rate is the number of deaths in a fraction of population [28]. Here, mortality rate has been expressed as the number deaths due to COVID-19 per 100,000 populations in one year. It was found that the mortality rate of COVID-19 is 59 per 100,000 population of Kurdistan region from March 2020 to March 2021 (Table 1). However, the mortality rate per 100,000 populations of Iraq, Iran, Turkey and the world are 38,75, 37 and 35 respectively (Table 1) [4, 24]. Again, the differences in the total number of COVID-19 tests carried out in each country may limit making comparison between Kurdistan region and the other countries. The main causes of mortality due to COVID-19 in Kurdistan region presumably are chronic diseases such as diabetes, hypertension, heart disease, and hypercholesteremia [30]

and age of the patients, as a quarter of COVID-19 patients are older than 50 years old in Kurdistan region [23]. The virulence of the SARS-CoV-2 variants also plays role in higher case fatality rate among COVID-19 patients [31]. Moreover, the assessment of health facilities and medical management of COVID-19 patients in Kurdistan region can be done through finding the percentage of recovered patients. It was found that 90.5% of COVID-19 patients were recovered (Figure 2). This result suggests that the medical management of COVID-19 is at standard quality in Kurdistan region.

Table 1: The epidemiological situation of COVID-19 in Kurdistan region and neighbouring countries from March 2020 to March 2021. The raw data of population size and COVID-19 situation were abstained from worldometer and WHO COVID-19 dashboard [4, 24].

Region/Country	Prevalence %	Mortality per 100,000
Kurdistan region	1.9	59
Iraq	2.1	38
Iran	2.2	75
Turkey	3.8	37
The world	1.6	35

The reporting of the outbreak of COVID-19 at governorates level is important to detect differences in the number of cases between the four governorates (Erbil, Sulaymaniyah, Duhok and Halabja) of Kurdistan region (Figure 3). This is also important to assess the prevention measures and quickly control any factor that may increase the number of cases. The number of COVID-19 cases was (42270, prevalence 1.9%), (34750, prevalence 1.6%), (40172, prevalence 2.4%) and (3127, prevalence 2.7%) in Erbil, Sulaymaniyah, Duhok and Halabja respectively (Figure 3 and Table 2). The greatest number of cases was reported in Erbil. This might be due to the population density of Erbil, the capital city that accommodates the largest number of population (Figure 1). It is known that COVID-19 spreads fast in heavily populated area. Consistently, an early short-term study observed high morbidity of COVID-19 in Erbil city [32]. The low COVID-19 cases in Sulaymaniyah may be related with greater area of Sulaymaniyah governorate and the distribution of Sulaymaniyah population among Sulaymaniyah city, and Garmian and Raparin administrative districts (Figure 1). This reduces the population density in the governorate centre and decreases the chance of person to person contact. Perhaps differences in the tightness of prevention measures also caused differences in cases and prevalence of COVID-19 among governorates. The rate of patient recovery ranges from 88-95% and seems to be similar among the four governorates (Figure 3 and Table 2). This suggests that medical management is well organized and the medical services at the hospitals are at good standard. The results also showed different rates of mortality (per 100,000) in Erbil (44), Sulaymaniyah (81), Duhok (47) and Halabja (124) (Table 2). The mortality due to COVID-19 infection might be related to the patients' age [23] and underlying health conditions such as chronic illnesses of hypertension, diabetes and cancer [1, 30, 33]. Moreover, deaths due to COVID-19 may be related to the virulence of SARS-CoV-2 variant (Table 1). This can be supported by the detection of the N501Y and A570D mutations in SARS-CoV-2 isolates in Kurdistan region [34]. The N501Y and A570D

mutations were already observed in the variant of concern of the UK, which is claimed to be lethal [35, 36].

As the COVID-19 is present as both asymptomatic and symptomatic infections and screening test was not done to the

entire populations of the Kurdistan governorates, these data might not represent the real epidemiological situation of COVID-19 in the Kurdistan governorates. However, these data can still provide analytical information about COVID-19 in the region

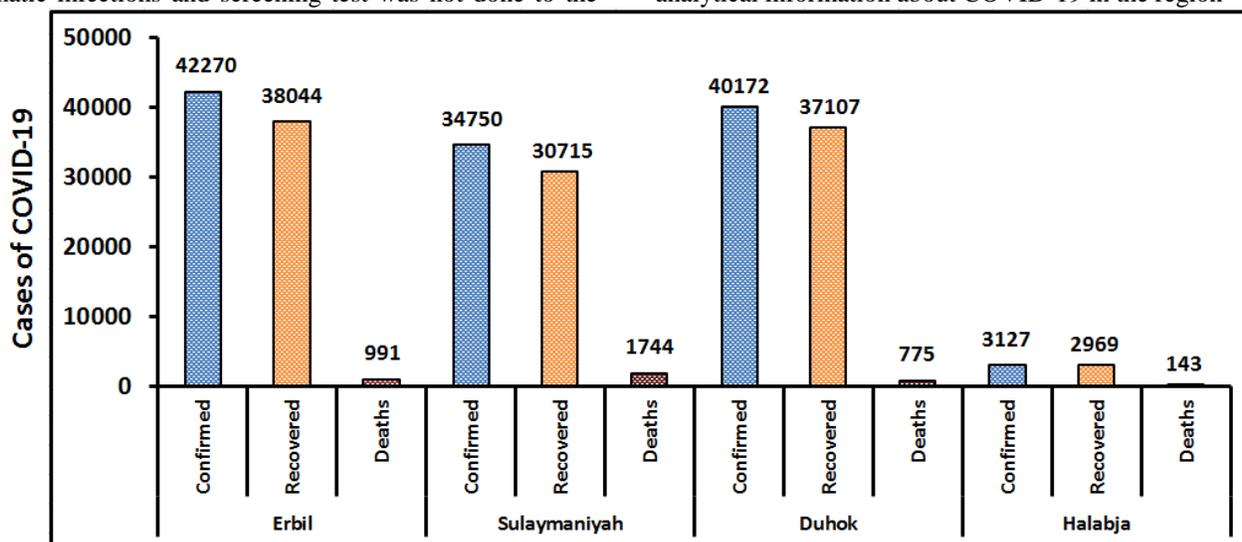


Table 2: The epidemiological situation of COVID-19 in Kurdistan region at Governorate level from March 2020 to March 2021.

Figure 3: The number of confirmed cases, recovered patients and deaths due to COVID-19 in the four governorates of Kurdistan region from March 2020 to March 2021.

Governorate	Prevalence %	Recovery rate %	Mortality per 100,000
Erbil	1.9	90	44
Sulaymaniyah	1.6	88	81
Duhok	2.4	92	47
Halabja	2.7	95	124

3. 2. Gender distributions of COVID-19 in Kurdistan Region

Overall the population of Kurdistan region appears gender-balanced [37]. However, higher percentage of COVID-19 cases was reported in males (61.6%) compared to females (38.4%) (Figure 4). This is consistent with other clinical and epidemiological studies in China, Italy and south Korea that showed males are more susceptible to COVID-19 than the females [1, 38, 39]. This may be determined by some possible factors such as higher expression of coronavirus receptor (ACE 2) in males, immunological differences driven by sex hormone and X chromosome and, smoking and drinking, which are more common among men than women. Moreover, females responsibly follow the prevention measures [40].

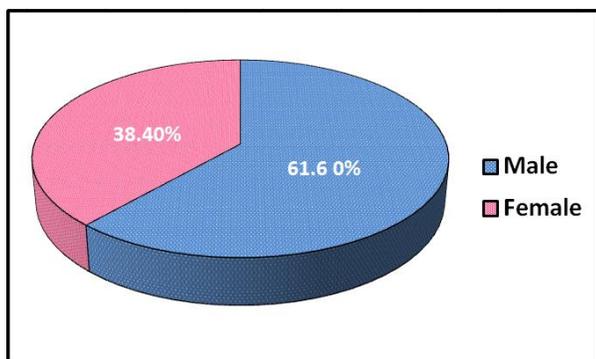


Figure 4: Gender distribution of COVID-19 patients in Kurdistan region of Iraq after one year (March 2020 to March 2021).

3. 3. The waves of COVID-19 in Kurdistan Region

Like the rest of the world, the COVID-19 occurred in three waves in Kurdistan region (Figure 5). The initial wave started in March 2020 with small number of cases. The outbreak was soon tackled through prevention measures and general lockdown. The protection measures such as quarantine of suspected cases, curfew and general lockdown was very effective in suppressing the first wave in Kurdistan region. However, the second wave happened with greater number of cases in June 2020 after loosening of the prevention measures and partial lifting of restrictions in the region [41]. The second wave stayed high for longer period from June to November 2020 (Figure 5). The reasons could be the gathering of people in the markets, malls, mosques, cafe, restaurants, social ceremonies and so on. Again, the cases of COVID-19 decreased in the cold months of December 2020, and January and February 2021 (Figure 5). It is clear that cold weather restricts the outdoors activities of people and reduces physical contact among them.

The third wave of COVID-19 has initiated in the early of March 2021 (Figure 5). The third wave may be due to ignoring of the health guidelines of protective measures against COVID-19 by people, normalizing the people’s routine life and opening of schools in February 2021. The normalizing of routine life activities in the region increases the close contacts among people and eases the circulation of the virus. In is worth mentioning that the number of COVID-19 cases doubled in the late of March

2021, perhaps because of gathering people in Nawroz and Spring festivals. The children are known to be carriers of COVID-19 virus, so the contacting of pupils at the schools results in the spreading of the virus among their families.

Additionally, the reason of second and third waves of COVID-19 may be related with the emergence of new variants of SARS-CoV-2 and the immunity of Kurdistan population against COVID-19. It was reported that the spike protein of SARS-CoV-2 is subsequently undergoing changes over time [12, 42]. This makes some antibodies against the SARS-CoV-2 strain in the first wave may no longer be effective against SARS-CoV-2 strains in the next waves [43]. The early cases of COVID-19 were reported

in Kurdistan region in March 2020 and exactly one year passed until the time of this study. So, the antibody titers against SARS-CoV-2 probably decreased in the people, who initially contracted COVID-19. This results in re-infection of the people in third wave. Consistently, a study likely suggested the persistence of neutralizing antibody for one year for mild and severe cases [44], but with lower antibody titers and shorter duration of maintain in non-severe cases [45].

Interestingly, the cumulative cases and waves of COVID-19 in Kurdistan region (figure 5) suggest that the infectivity of SARS-CoV-2 might be increased in moderates and hot months, whereas it's infectivity is reduced in cold months of the year.

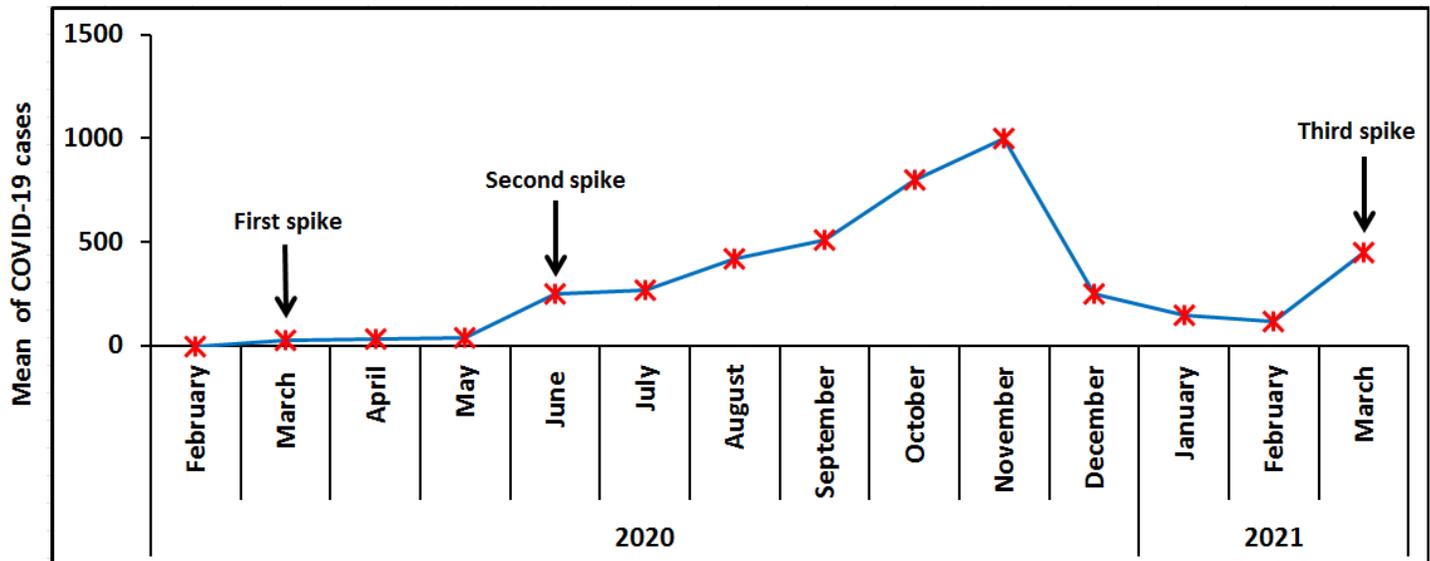


Figure 5: The cumulative cases and waves (spikes) of COVID-19 in Kurdistan region of Iraq after one year (March 2020 to March 2021). The data represent the average of COVID-19 cases in each month.

3. 4. Variants of SARS-CoV-2 in Kurdistan Region

Point mutations frequently occur in RNA viruses and result in antigenic drifts particularly in the immunogenic surface proteins. The surface spike protein of SARS-CoV-2 has been undergoing mutations [12]. Several point mutation has been reported in the gene, encoding the spike protein of SARS-CoV-2 [20]. Any antigenic drift in the spike protein produces a new variant of the virus.

Since the beginning of COVID-19 outbreak, the focus of Kurdistan health sectors was on the protection measures, PCR detection of the cases and medical management of the patients. However, the investigations of the variants of the virus have been paid little attention in Kurdistan region. There are only few studies that investigated mutations in the spike protein of Kurdistan isolates of SARS-CoV-2, using low-cost and rapid partial sequencing technique [34, 46]. The continuous monitoring of SARS-COV-2 variants is very important to find alterations in the transmissibility, infectivity and antigenicity of the virus. This definitely gives insights for evaluation, controlling and expecting the epidemiological situation of COVID-19 and for ordering the right version of COVID-19 vaccine.

In this study, the amino acid changes in the spike proteins of SARS-CoV-2 isolates were investigated in Kurdistan region. The NCBI bioinformatics tool was searched for any available full genome sequence data about clinical SARS-CoV-2 isolates in Kurdistan region, only two full sequences, QXX36542.1 and QRW43499.1, of the viral spike protein were found that were submitted in January and February of 2021 respectively. Besides, some other partial sequences of spike protein of Kurdistan clinical isolates (QNA42621.1, QTQ61374.1, QTQ61375.1, QTQ61376.1, QTQ61377.1, QTQ61378.1 and QTQ61379.1) were found, which were submitted in August 2020 for QNA42621.1 and April 2021 for the rest of the isolates.

The spike protein sequences of QXX36542.1 and QRW43499.1 were compared to the initial isolate of SARS-CoV-2 in Wuhan/China (Wuhan_YP_009724390.1). The percent identity matrix of the spike protein showed that two Kurdistan isolates, QXX36542.1 and QRW43499.1, have 99.76% identity to Wuhan_YP_009724390.1 and 99.69% identity to each other (Table 3). This comparison suggests that the SARS-CoV-2 isolates of Kurdistan greatly conserved the spike protein sequence of the Wuhan Wuhan_YP_009724390.1 isolate and only few changes have occurred, which have been investigated below.

Table 3: Percent identity matrix of spike proteins between Wuhan and two Kurdistan isolates of SARS-CoV-2. The identity analysis created by Clustal 2.1.

Accession No.	Identity matrix (%)		
Wuhan_YP_009724390.1	100.00	99.76	99.76
QQX36542.1	99.76	100.00	99.69
QRW43499.1	99.76	99.69	100.00

The multiple alignment analysis was done for the spike protein sequences of the Kurdistan isolates, using SARS-CoV-2 in Wuhan/China (Wuhan_YP_009724390.1) as reference sequence. The results showed that the spike protein in each of QQX36542.1 and QRW43499.1 contains three amino acid substitutions (Figure 6A and Table 4). The mutations in the spike protein QQX36542.1 are A348S, T478K and D614G, whereas L452R, E583D and D614G in QRW43499.1. Moreover, the spike proteins of QNA42621.1, QTQ61374.1, QTQ61375.1, QTQ61376.1 and QTQ61377.1 only contained D614G substitution. However, the QTQ61378.1 and QTQ61379.1 showed three mutations of N501Y, A570D and D614G (Figure 6 B and Table 4). Apart from A570D, E583D and D614G, all other mutations are located in the receptor binding domain (RBD) of the spike proteins (Figure 6A and B, Table 4).

The impact of each mutation in the spike protein must be taken in to count in terms of infectivity, antigenicity and transmissibility. Previous studies showed some common amino acid substitutions in the spike protein that lead to increased infectivity and transmissibility of SARS-CoV-2. The D614G substitution is located outside the RBD and considered as a common mutation in nearly 70 % of the spike protein of COVID-19 virus isolates by June 2020 [42]. The D614G mutation presumably induces the conformational change in the spike protein and leads to increased infectivity due to strong binding to ACE2 [31, 42]. Studies also found that the D614G variant alone and other variants containing both D614G and another amino acid substitution were more infectious, stable, and transmissible, but did not cause more severe illness [12, 47]. The spike protein in all the isolates of Kurdistan region showed D614G either alone or with two more mutations (Figure 6 A and B, Table 4), suggesting the circulation of such dominate mutation in the virus in the region. The D614G substitution was already detected globally, including Iraq [46, 48, 49]. Although the mortality rate is low in Kurdistan region (Figure 2 and Table 1), a preliminary study suggested that the D614G dominant variant is related with the increased fatality rate [31], this definitely requires further confirmatory data.

Recently a variant of concern, called VOC-202012/01 or Kent variant, was reported in the UK in December 2020. This variant is characterised by having a cluster of mutations in the spike protein. The mutations are amino acid deletions of HV 69-70 and Y144, and amino acid substitutions of N501Y, A570D, P681H, T716I, S982A and D1118H. It is estimated that VOC-202012/01 variant is more transmissible than the wild-type SARS-Cov-2. The increasing transmissibility is related with the unusual N501Y mutation, because it is one of the six key contact residues in the receptor-binding domain (RBD) that increases the binding affinity of the spike protein to the human ACE2 [21]. However, the fatality of the UK variant is controversial that needs further

investigations [35, 36, 50]. Moreover, only two isolates (QTQ61378.1 and QTQ61379.1) in Kurdistan region showed N501Y and A570D mutations in their spike proteins (Figure 6 B and Table 4), suggesting the presence of VOC-202012/01 or Kent variant in Kurdistan region. This supports the claim of the Kurdistan health officials for the presence of the UK variant of SARS-CoV-2 in the region. This is also consistent with a study that used rapid and low-cost technique to detect N501Y and A570D mutations in Kurdistan clinical isolates [34]. The combination of N501Y, A570D and D614G in the isolates of QTQ61378.1 and QTQ61379.1 (Table 4) suggests that the Kurdistan variants of SARS-CoV-2 are more infectious and transmissible. This is supported by the studies that found D614G with other mutations in the same variant make the virus more infectious, stable, and transmissible [12, 47].

In addition, the L452R mutation was suggested to make the virus more transmissible and was initially detected in Denmark and recently in Californian [51] and Indian [52]. Only the Kurdistan isolate of QRW43499.1 contained L452R mutation (Figure 6A and Table 4), which is a co-mutation with E583D and D614G in the isolate. The combination of two infectivity-enhancing mutations in a single isolate of QRW43499.1, suggests tie circulation of a virulent variant of SARS-CoV-2 in the region. The E583D mutation was also detected in the isolate of QRW43499.1. The same mutation was already observed in American [53] and Asian isolates of SARS-CoV-2, including India [54, 55]. The E583D mutation is located away from the RBD

and it's relation with infectivity of SARS-CoV-2 remains clear.

Moreover, two of the amino acid substitutions in one of the Kurdistan isolate, QQX36542.1, are A348S and T478K (Figure 6A and Table 4), which were already detected in America [53, 56]. Both of the mutations are located in the RBD of the spike protein. A study suggested that the A348S causes negative binding affinity of the spike protein with ACE2, whereas T478K results in the positive binding affinity. The reason could be related with the location of the mutations in the receptor-binding motif (RBM), the region of RBD that directly connects the spike protein with the ACE2. The A348S is not located within the RBM, but T478K is [53, 56]. So, T478K can be considered as an infectivity-enhancing mutation. Thus the presence of D614G mutation with A348S and T478K in a single isolate, QQX36542.1 (Figure 6A and Table 4), indicates circulating of an infectious and transmissible variant of SARS-CoV-2 in Kurdistan region.

The presence of infectious and transmissible isolates of QQX36542.1, QRW43499.1, QTQ61378.1 and QTQ61379.1 suggests the strength of COVID-19 outbreak in Kurdistan region. These isolates may have caused the third wave of COVID-19 in March 2021.

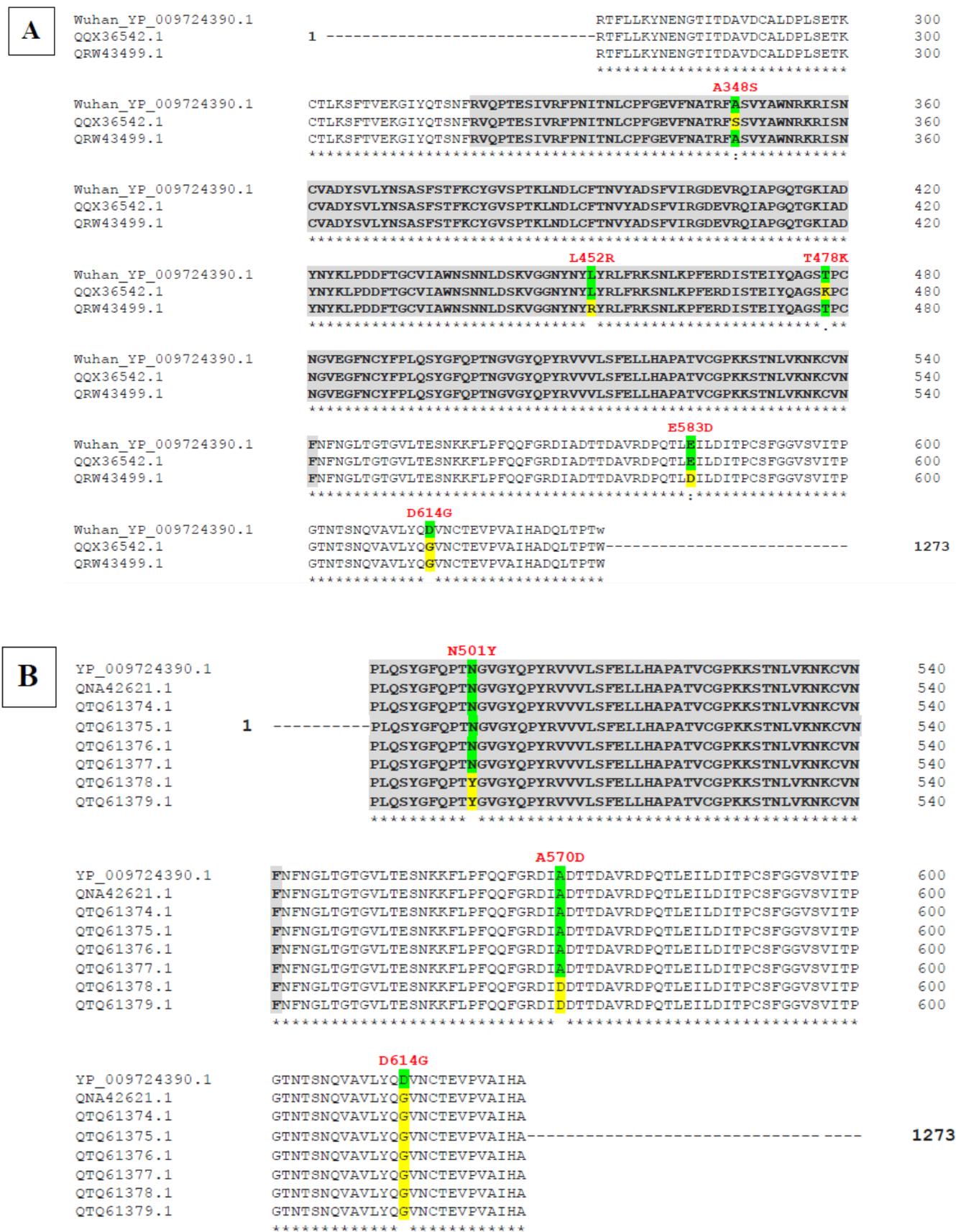


Figure 6: Multiple alignment of full spike protein amino acid sequences between first Wuhan SARS-CoV-2 isolate (Wuhan_YP_009724390.1) and two recently clinical isolates (QQX36542.1 and QRW43499.1) in Kurdistan region (A). Also, multiple alignment of full spike protein amino acid sequences of Wuhan SARS-CoV-2 isolate (Wuhan_YP_009724390.1) with some partial sequences of clinical isolates (QNA42621.1, QTQ61374.1, QTQ61375.1, QTQ61376.1, QTQ61377.1, QTQ61378.1 and QTQ61379.1) in the region (B). The point mutations are bolded and highlighted in yellow. The receptor binding domain (RBD) is bolded and highlighted in grey.

Table 4: The amino acid substitutions in the spike proteins of SARS-CoV-2 isolates in Kurdistan region. The mutations that located in receptor binding domain of the protein are bolded and underlined.

Isolates of SARS-CoV-2	Amino acid substitutions in spike protein
QQX36542.1	<u>A348S</u> , <u>T478K</u> , D614G
QRW43499.1	<u>L452R</u> , E583D, D614G
QNA42621.1	D614G
QTQ61374.1	D614G
QTQ61375.1	D614G
QTQ61376.1	D614G
QTQ61377.1	D614G
QTQ61378.1	<u>N501Y</u> , A570D, D614G
QTQ61379.1	<u>N501Y</u> , A570D, D614G

3. 5. Antigenic drift in the SARS-CoV-2 variants in Kurdistan Region

The viral surface proteins (e.g. spike protein of SARS-CoV-2 virus) are usually immunogenic antigens due to having epitopes

on their surface. The B-cell epitope of viral surface protein can specifically bind to the host's B-cell antigen receptor and induce the body to produce protective antibody. The antibodies bind to the antigenic epitopes in a highly specific manner and prevent infection through neutralizing the viral particle. A study reported the escape of multiple variants of SARS-CoV-2 from vaccine-induced neutralizing antibody immunity [57]. Thus the investigation of epitopes in the spike protein is valuable for the development and ordering of SARS-CoV-2 vaccine, and for understanding of SARS-CoV-2's pathogenesis [26, 58]. In this study, the BepiPred-2.0 server was used to predict the possible antigenic epitopes in the spike protein of SARS-CoV-2. This web server is a reliable tool for recognizing the epitope and non-epitope amino acid residues even in the 3D structure of antigenic proteins. The amino acid residues that have the scores greater than the threshold (default value is 0.5) are predicted to be a part of antigen epitopes [26]. The results of analysis showed that the Spike protein of initial Wuhan virus (Wuhan_YP_009724390.1) contain 34 predicted epitopes in the form of residues and peptides (Figure 7 and Table 5).

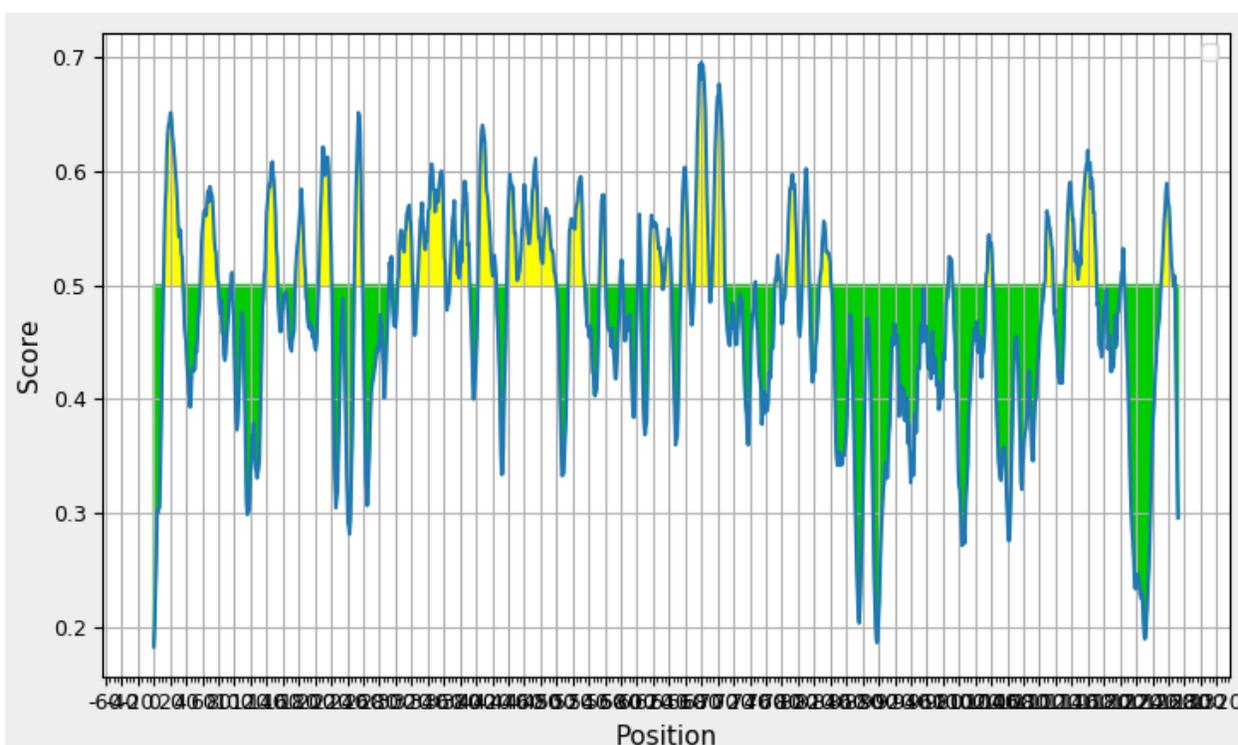


Figure 2: The predicted B-cell epitopes from Spike protein sequence of Wuhan SARS-CoV-2 strain (Wuhan_YP_009724390.1), The residues with scores above the threshold (default value is 0.5) are predicted to be part of an epitope and colored in yellow on the graph (where Y-axis depicts residue scores and X-axis residue positions in the sequence) [26].

The epitope regions of the spike proteins in all the Kurdistan isolates (Table 4 and Table 5) were then investigated. It was found that the isolate QQX36542.1 has two mutations (A348S and T478K) in the epitope number 10 and 13, which are located in the RBD of the spike protein (Table 5). However, the isolate QRW43499.1 showed two mutations (L452R and E583D) in the epitope number 13 and 16, and only L452R mutation in epitope 13 is located in the RBD. Moreover, N501Y mutation in two other Kurdistan isolates (QTQ61378.1 and QTQ61379.1) is also located in the epitope number 13 (Table 5). Interestingly, only the predicted epitope number 13 is located in the receptor binding motif (RBM), the region in the RBD of the spike protein that interact

with the ACE2 [53]. Thus, the occurrence of T478K, L452R and N501Y mutations in the RBM of three different isolates in Kurdistan region suggests changes in the antigenicity of the SARS-CoV-2 virus. The altered amino acid residues at the RBD of spike protein may make the SARS-CoV-2 virus resistant to the previously induced neutralizing antibodies, which block the binding of spike protein to ACE2 receptor [12, 57]. This leads to the frequent circulation of the new variants of the virus in population. Moreover, it was experimentally shown that L452R and T478K substitutions at the RBD increase resistance against neutralizing antibodies [12, 59], suggesting that they are located in a strong antigenic epitope number 13 (Table 5).

Regarding mutations in the predicted epitopes outside the RBD, the E583D is located in epitope number 16 and it is only 40 residues away from the RBD (Figure 6 and table 5). So, E583D substitution may have effect on the antigenicity of the spike protein and makes it even partially resistant to neutralizing antibody. Furthermore, the substitution of D (Aspartic acid) residue by G (Glycine) residue at position 614 of the spike protein

is also detected in the Kurdistan isolates. Our data revealed that D614 residue is not located in the predicted epitopes, whereas it is directly at the upstream of the epitope 18 (Table 5). In contract, early studies revealed that D614 residue is located in the middle of a B-cell epitope, so the replacement of a large acidic aspartic acid by a small hydrophobic glycine in D164G variant would reduce the sensitivity of the spike protein to antibodies ^[60, 61].

Table 5: The amino acid residues/peptides in the predicted epitope regions of SARS-CoV-2 spike protein. These amino acid residues/peptides sequences represent the yellow peaks in the (figure 7). The residues of the peptides with scores above the default threshold of 0.5 are involved in epitope regions ^[26]. The point mutations are highlighted in yellow. The predicted epitopes located in the receptor binding domain (RBD) are bolded and underlined. The Epitope number 13 is located in receptor binding motif (RBM) of the spike protein ^[53].

No.	Start	End	Peptide	Length
1	13	37	SQCVNLTRTQLPPAYTNSFTRGVY	25
2	59	81	FSNVTWFHAIHVSGTNGTKRFDN	23
3	97	98	KS	2
4	138	154	DPFLGVYYHKNNKSWME	17
5	177	189	MDLEGKQGDFKLN	13
6	206	221	KHTPINLVRDLDPQGF	16
7	250	260	TPGDSSSGWTA	11
8	293	296	LDPL	4
9	304	322	KSFTVEKGIYQTSNFRVQP	19
10	329	363	Wuhan_YP_009724390.1 QX36542.1 QRW43499.1 QTQ61378.1 QTQ61379.1 FPNITNLCPFGEVFNATRFASVYAWNRKRISNCVA FPNITNLCPFGEVFNATRFSSVYAWNRKRISNCVA FPNITNLCPFGEVFNATRFASVYAWNRKRISNCVA FPNITNLCPFGEVFNATRFASVYAWNRKRISNCVA FPNITNLCPFGEVFNATRFASVYAWNRKRISNCVA	35
11	369	393	<u>YNSASFSTFKCYGVSPTKLNLCFT</u>	25
12	404	426	<u>GDEVROIAPGQTGKIADYNYKLP</u>	23
13	440	501	L452R NLDSKVGGNYNLYR ^L FRKSNLKP ^F ERDISTEIQAGS ^T PCNGVEGFNCYFPLQSYGFQPTN NLDSKVGGNYNLYR ^L FRKSNLKP ^F ERDISTEIQAGS ^T PCNGVEGFNCYFPLQSYGFQPTN T478K N501Y	62
14	516	536	<u>ELLHAPATVCGPKKSTNLVKN</u>	21
15	555	562	SNKKFLPF	8
16	580	583	E583D QTLE QTLE QTLD QTLE QTLE	4
17	602	606	TNTSN	5
18	615	632	D614G -VNCTEVPVAIHADQLTPT	18
19	634	644	RVYSTGNSNVFQ	11
20	656	666	VNNSYECDIPI	11
21	672	690	ASYQTQTNSPRRARSVASQ	19
22	695	710	YTMSLGAENSVAYSNN	16
23	748	748	E	1
24	773	779	EQDKNTQ	7
25	786	800	KQIYKTPPIKDFGGF	15
26	807	814	PDPSKPSK	8
27	828	842	LADAGFIKQYGDCLG	15
28	988	992	EAEVQ	5
29	1035	1043	GQSKRVDFC	9
30	1107	1118	RNFYEPQIITD	12
31	1133	1172	VNNTVYDPLQPELDSFKEELDKYFKNHTSPDVLDGDISGI	40
32	1203	1206	LGKY	4
33	1252	1267	SCCKFDEDDSEPVLKG	16
34	1269	1269	K	1

4. Conclusion

The situation of COVID-19 in Kurdistan region seems to be at satisfactory levels of prevalence, recovery rate and mortality rate compared to the global figure. The border should be controlled with Turkey and Iran, where the prevalence of COVID-19 and mortality are high respectively. The main causes of death by COVID-19 should be further investigated to reduce the mortality rate of COVID-19 in Kurdistan region particularly in Sulaymaniyah and Halabja. Interestingly, the cumulative cases of COVID-19 suggests that the infectivity of SARS-CoV-2 might increase in moderate and hot months, whereas decrease in cold months.

In addition, the SARS-CoV-2 spike protein is continually evolved to emerge new variants of the virus. The investigation of spike protein sequences, from nine SARS-CoV-2 isolates in Kurdistan region, revealed single amino acid mutation of (D614G) and combined amino acid substitutions of (A348S, T478K and D614G), (L452R, E583D and D614G) and (N501Y, A570D and D614G), which are in or close to the RBD. Interestingly, all the isolates are D614G variants of SARS-CoV-2 that is already known for infectiousness and fast transmission. Moreover, the N501Y and A570D mutations indicate the endemic of the UK variant of concern in Kurdistan region. The detection of four infectivity-enhanced mutations (T478K, L452R, N501Y and D614G) in SARS-CoV-2 isolates suggests that COVID-19 must be paid great attention in Kurdistan region. The detected amino acid mutations (A348S, T478K, L452R, and N501Y), in the RBD of the spike proteins, are located in the predicted epitopes. This may make these variants resistant to some of the neutralizing antibodies and cause recurrent infection. More sequencing data and more investigations of SARS-CoV-2 variants are required to monitor, predict and control COVID-19 outbreak in Kurdistan region.

Acknowledgements

I would like to thank Assist. Prof. Dr. Dana Khdr Sabir at Charmo university for bioinformatics analysis.

References

- Guan, W.-j., et al., *Clinical Characteristics of Coronavirus Disease 2019 in China*. New England Journal of Medicine, 2020. **382**(18): p. 1708-1720.
- Zhu, N., et al., *A Novel Coronavirus from Patients with Pneumonia in China, 2019*. N Engl J Med, 2020. **382**(8): p. 727-733.
- World Health Organization "Coronavirus Disease 2019 (COVID-19) Situation Report—51". https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57_10,2020.
- World Health Organization "Coronavirus (COVID-19) Dashboard". <https://covid19.who.int/>, 2021.
- Backer, J.A., D. Klinkenberg, and J. Wallinga, *Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28 January 2020*. Euro Surveill, 2020. **25**(5).
- Li, Q., et al., *Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia*. New England Journal of Medicine, 2020. **382**.
- Cao, X., *COVID-19: immunopathology and its implications for therapy*. Nat Rev Immunol, 2020. **20**(5): p. 269-270.
- Chan, J.F. and K.H. Kok, *Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan*. 2020. **9**(1): p. 221-236.
- Ahmed, S.F., A.A. Quadeer, and M.R. McKay, *Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies*. Viruses, 2020. **12**(3).
- Jeyanathan, M., et al., *Immunological considerations for COVID-19 vaccine strategies*. 2020. **20**(10): p. 615-632.
- Wrapp, D. and N. Wang, *Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation*. 2020. **367**(6483): p. 1260-1263.
- Li, Q., et al., *The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity*. Cell, 2020. **182**(5): p. 1284-1294.e9.
- Shin, M.D., et al., *COVID-19 vaccine development and a potential nanomaterial path forward*. Nature nanotechnology, 2020. **15**(8): p. 646-655.
- Wu, F., et al., *Neutralizing antibody responses to SARS-CoV-2 in a COVID-19 recovered patient cohort and their implications*. medRxiv, 2020: p. 2020.03.30.20047365.
- Duffy, S., *Why are RNA virus mutation rates so damn high?* PLoS Biol, 2018. **16**(8): p. e3000003.
- Lauring, A.S. and R. Andino, *Quasispecies theory and the behavior of RNA viruses*. PLoS Pathog, 2010. **6**(7): p. e1001005.
- Khailany, R.A., M. Safdar, and M. Ozaslan, *Genomic characterization of a novel SARS-CoV-2*. Gene Rep, 2020. **19**: p. 100682.
- Reid, A.H. and J.K. Taubenberger, *The origin of the 1918 pandemic influenza virus: a continuing enigma*. J Gen Virol, 2003. **84**(Pt 9): p. 2285-2292.
- Carrat, F. and A. Flahault, *Influenza vaccine: the challenge of antigenic drift*. Vaccine, 2007. **25**(39-40): p. 6852-62.
- Wise, J., *Covid-19: New coronavirus variant is identified in UK*. BMJ, 2020. **371**: p. m4857.
- Chand, M., et al., *Investigation of novel SARS-COV-2 variant: Variant of Concern 202012/01. Technical briefing 1*. Public Health England. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959438/Technical_Briefing_VOC_SH_NJL2_SH2.pdf,2020.
- World Health Organization "Novel Coronavirus (2019-nCoV) Situation Report—1". https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4,2020.
- Kurdistan Regional Government. *Situation Update on Coronavirus (COVID-19) Dashboard*. <https://gov.krd/coronavirus-en/situation-update/>, 2021.
- Current World Population, *Worldometers* <https://www.worldometers.info/world-population/>, 2021.
- National Centre for Biotechnology Information (NCBI). <https://www.ncbi.nlm.nih.gov/>, 2021.
- Jespersen, M.C., et al., *BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes*. Nucleic Acids Res, 2017. **45**(W1): p. W24-w29.
- Kurdistan Region Statistics Office, Ministry of Planning, *Kurdistan Regional Government-Iraq*. <http://www.krso.net/Default.aspx?page=article&id=12206&l=1,2020>.
- Carroll, L., *Prevalence*, in *Encyclopedia of Behavioral Medicine*, M.D. Gellman and J.R. Turner, Editors. 2013, Springer New York: New York, NY. p. 1530-1531.
- Dana, S., et al., *Factors Contributing to the Containment of the COVID-19 in Kurdistan Region of Iraq*. Frontiers in Emergency Medicine, 2020. **4**(2s).
- Aziz, P.Y., et al., *The strategy for controlling COVID-19 in Kurdistan Regional Government (KRG)/Iraq: Identification, epidemiology, transmission, treatment, and recovery*. International Journal of Surgery Open, 2020. **25**: p. 41-46.
- Becerra-Flores, M. and T. Cardozo, *SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate*. 2020. **74**(8): p. e13525.

32. Ramzi, Z.S., *Epidemiological and Clinical Characteristics of COVID-19 Patients in Kurdistan Region/Iraq*. Annals of the Romanian Society for Cell Biology, 2021. **25**(1): p. 2068-2075.
33. Goyal, P., et al., *Clinical Characteristics of Covid-19 in New York City*. 2020. **382**(24): p. 2372-2374.
34. Al-Jaf, S.M.A. and S.S. Niranjani, *Rapid detection of SARS CoV-2 N501Y mutation in clinical samples*. medRxiv, 2021: p. 2021.04.17.21255656.
35. Challen, R., et al., *Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: matched cohort study*. BMJ, 2021. **372**: p. n579.
36. Davies, N.G., et al., *Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7*. Nature, 2021. **593**(7858): p. 270-274.
37. *Demographic Survey: Kurdistan Region of Iraq*. https://reliefweb.int/sites/reliefweb.int/files/resources/KRSO_IOM_UNFPA_Demographic_Survey_Kurdistan_Region_of_Iraq.pdf, 2018.
38. Onder, G., G. Rezza, and S. Brusaferro, *Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy*. Jama, 2020. **323**(18): p. 1775-1776.
39. Shim, E., et al., *Transmission potential and severity of COVID-19 in South Korea*. Int J Infect Dis, 2020. **93**: p. 339-344.
40. Bwire, G.M., *Coronavirus: Why Men are More Vulnerable to Covid-19 Than Women?* SN comprehensive clinical medicine, 2020: p. 1-3.
41. Ali, K.M., H.M. Tawfeeq, and H.M. Rostam, *COVID-19 Second Spike as an Aftermath of the Sudden Restrictions Ease: Kurdistan Region of Iraq as an Example*. Passer Journal, 2020. **2**(2): p. 57-61.
42. Yurkovetskiy, L., et al., *Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant*. Cell, 2020. **183**(3): p. 739-751.e8.
43. Zhu, Z., et al., *Potent cross-reactive neutralization of SARS coronavirus isolates by human monoclonal antibodies*. Proceedings of the National Academy of Sciences, 2007. **104**(29): p. 12123-12128.
44. Lau, E.H.Y., et al., *Neutralizing antibody titres in SARS-CoV-2 infections*. Nature Communications, 2021. **12**(1): p. 63.
45. Bruni, M., et al., *Persistence of Anti-SARS-CoV-2 Antibodies in Non-Hospitalized COVID-19 Convalescent Health Care Workers*. 2020. **9**(10).
46. Al-Jaf, S.M.A., S.S. Niranjani, and Z.H. Mahmood, *Rapid, inexpensive methods for exploring SARS CoV-2 D614G mutation*. medRxiv, 2021: p. 2021.04.12.21255337.
47. Plante, J.A., et al., *Spike mutation D614G alters SARS-CoV-2 fitness*. Nature, 2020.
48. Aktas, E., *Bioinformatic Analysis Reveals That Some Mutations May Affect On Both Spike Structure Damage and Ligand Binding Site*. bioRxiv, 2020: p. 2020.08.10.244632.
49. Al-Rashedi, N.A.M., et al., *Genome Sequencing of a Novel Coronavirus SARS-CoV-2 Isolate from Iraq*. 2021. **10**(4).
50. Frampton, D., 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.
51. Peng, J., et al., *Estimation of secondary household attack rates for emergent SARS-CoV-2 variants detected by genomic surveillance at a community-based testing site in San Francisco*. medRxiv : the preprint server for health sciences, 2021: p. 2021.03.01.21252705.
52. Cherian, S., et al., *Convergent evolution of SARS-CoV-2 spike mutations, L452R, E484Q and P681R, in the second wave of COVID-19 in Maharashtra, India*. bioRxiv, 2021: p. 2021.04.22.440932.
53. Wang, R., et al., *Analysis of SARS-CoV-2 mutations in the United States suggests presence of four substrains and novel variants*. Communications Biology, 2021. **4**(1): p. 228.
54. Sharma, S., et al., *Emergence and expansion of highly infectious spike:D614G mutant SARS-CoV-2 in central India*. bioRxiv, 2020: p. 2020.09.15.297846.
55. Banerjee, R., et al., *Spike protein mutational landscape in India: Could Muller's ratchet be a future game-changer for COVID-19?* bioRxiv, 2020: p. 2020.08.18.255570.
56. Wang, R., et al., *Characterizing SARS-CoV-2 mutations in the United States*. Research square, 2020: p. rs.3.rs-49671.
57. Garcia-Beltran, W.F., et al., *Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity*. Cell, 2021. **184**(9): p. 2372-2383.e9.
58. El-Manzalawy, Y. and V. Honavar, *Recent advances in B-cell epitope prediction methods*. Immunome Res, 2010. **6 Suppl 2**(Suppl 2): p. S2.
59. Liu, Z., et al., *Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization*. Cell Host Microbe, 2021. **29**(3): p. 477-488.e4.
60. Koyama, T., et al., *Emergence of Drift Variants That May Affect COVID-19 Vaccine Development and Antibody Treatment*. Pathogens, 2020. **9**(5): p. 324.
61. Grifoni, A., et al., *A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2*. Cell Host Microbe, 2020. **27**(4): p. 671-680.e2.