# **REVIEW / DERLEME**

DOI: 10.4274/mjima.galenos.2021.2021.15 Mediterr J Infect Microb Antimicrob 2021;10:15

Erişim: http://dx.doi.org/10.4274/mjima.galenos.2021.2021.15



# **Predominant Mutations of SARS-CoV-2: Their Geographical Distribution and Potential Consequences**

SARS-CoV-2'nin Yaygın Mutasyonlarının Coğrafi Dağılımı ve Olası Sonuçları

Sezin ÜNLÜ, 🖻 Aylın ÜSKÜDAR-GÜÇLÜ, 🗗 Ahmet BAŞUSTAOĞLU

Başkent University Faculty of Medicine, Department of Medical Microbiology, Ankara, Turkey

### Abstract

Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) emerged in late December 2019 in Wuhan, China. More than 83 million people have been infected, and more than 1.8 million people have died, as reported to the World Health Organization on the 3<sup>rd</sup> of January, 2021. Analysis of genetic variations is critical for understanding the spreading pattern of SARS-CoV-2 across several countries. This review aimed to gather information about the prominent mutations of SARS-CoV-2 by analyzing the origin, viral pathogenesis, and mutation rate. Moreover, we concluded their potential impacts on SARS-CoV-2 therapeutics. Mutations in the spike protein (D614G, N501Y, E484K, A222V, S477N, and G485R), ORF1ab (P323L, N628N, Y455Y, A97V, and F106F), nucleocapsid protein (R203K and G204R), ORF8 (L84S), and ORF3a (Q57H and G251V) were examined in this review by analyzing relevant articles from the beginning of the current pandemic to the most recent date. A detailed analysis of articles demonstrates that D614G is the major variation distributed globally, and its frequency increased rapidly from early in March, followed by several other variations in either spike or different proteins. In addition, it was seen that the currently circulating N501Y and E484K variants revealed a public concern regarding vaccines' efficacy. Investigation of variations of SARS-CoV-2 would lead to understanding their potential mechanism of action against SARS-CoV-2, thereby suggesting suitable therapeutics. Several mechanisms were suggested to have a role in SARS-CoV-2 mutation rate and evolution. Possible therapeutics and vaccines against SARS-CoV-2 were proposed.

Keywords: SARS-CoV-2, genomic variations, impacts of SARS-CoV-2 mutations, geographical distribution

# Öz

Siddetli akut solunum sendromu Koronavirüs-2 (SARS-CoV-2), 2019 yılının aralık ayının sonunda Cin'in Wuhan kentinde ortaya çıkmıştır. Dünya Sağlık Örgütü'nce 3 Ocak 2021'de bildirildiği üzere SARS-CoV-2, 83 milyondan fazla insanı enfekte ederken 1,8 milyondan fazla insanın da hayatını kaybetmesine sebep olmuştur. Genetik varyasyonların analizi, SARS-CoV-2 yayılma modelini anlamak için kritik bir öneme sahiptir. Bu derleme, bu mutasyonların kökenini, viral patogenezini ve mutasyon oranını analiz ederek SARS-CoV-2'nin yaygın mutasyonları hakkında bilgi toplamayı amaçlamıştır. Aynı zamanda, bu mutasyonların SARS-CoV-2 terapötiklerine karşı potansiyel etkilerini de incelemek amaçlanmıştır. Bu derlemede, Spike proteini (D614G, N501Y, E484K, A222V, S477N, G485R), ORF1ab (P323L, N628N, Y455Y, A97V, F106F), nükleokapsid protein (R203K, G204R), ORF8 (L84S) ve ORF3a (Q57H, G251V) mutasyonları, mevcut pandeminin başlangıcından güncel zamana kadar olan ilgili makaleler analiz edilerek incelenmiştir. Makalelerin ayrıntılı analizi, D614G'nin küresel olarak dağılan başlıca varyasyon olduğunu ve sıklığının mart ayının başlarından itibaren hızla arttığını ve ardından hem spike hem de farklı proteinlerdeki diğer birkaç varyasyonun geldiğini göstermektedir. Buna ek olarak, yeni yayılan N501Y ve E484K varyantlarının aşı etkinliği açısından endişe oluşturduğu görülmüştür. Şiddetli akut solunum sendromu Koronavirüs-2 varyasyonlarının araştırılması, SARS-CoV-2'ye karşı potansiyel etki mekanizmalarının anlaşılmasına ve böylece ona karşı terapötikler önerilmesine

Anahtar Kelimeler: SARS-CoV-2, genomik varyasyonlar, SARS-CoV-2 mutasyonlarının etkileri, coğrafi dağılımı

Cite this article as: Ünlü S, Üsküdar Güçlü A, Başustaoğlu A. Predominant Mutations of SARS-CoV-2: Their Geographical Distribution and Potential Consequences. Mediterr J Infect Microb Antimicrob. 2021;10:15.



Address for Correspondence/Yazışma Adresi: Aylin Üsküdar-Güçlü PhD, Başkent University Faculty of Medicine, Department of Medical Microbiology, Ankara, Turkey

Phone: +90 533 447 20 98 E-mail: uskudaraylin@gmail.com

Received/Geliş Tarihi: 09.01.2021 Accepted/Kabul Tarihi: 03.03.2021 ORCID ID: orcid.org/0000-0002-1872-028X

©Copyright 2021 by the Infectious Diseases and Clinical Microbiology Specialty Society of Turkey Mediterranean Journal of Infection, Microbes and Antimicrobials published by Galenos Yayınevi.

Published: 24 February 2021

# Introduction

Two major pathogenic zoonotic outbreaks of beta coronaviruses have been seen in the past two decades. The first encounter, named Severe acute respiratory syndrome Coronavirus-2 (SARS-CoV-2), was in November 2002, which was identified for the first time in Guandong, China[1]. By July 2003, more than 8,000 people were infected, and 774 deaths were reported<sup>[2]</sup>. Ten years later, in 2012, a significantly more lethal virus, Middle East respiratory syndrome Coronavirus (MERS-CoV), emerged. It was identified in the Arabian Peninsula first and then transmitted through 27 countries[3]. The transmission rate of MERS-CoV was lower than SARS-CoV-2, but its mortality rate was higher. The mortality rate of MERS-CoV was 35.5% compared with 9.6% for SARS-CoV<sup>[4]</sup>. In late December 2019, novel severe acute respiratory syndrome coronavirus (SARS-CoV-2) was reported in Wuhan City, China. World Health Organization (WHO) declared a pandemic on March 11, 2020<sup>[5]</sup>. This new disease is named Coronavirus disease-2019 (COVID-19). Globally, it infected more than 83 million people and resulted in the death of >1.8 million people, as reported by the WHO on the 3th of January, 2021.

The identification of genomic variations for SARS-CoV-2 is a global interest due to differences in mortality rate and incidence among different countries. Wu and colleagues studied a seafood market worker who was hospitalized on 26th of December, 2019. This was the first genomic sequence of SARS-CoV-2, which provides a significant opportunity to compare genomic studies with this reference sequence (GenBank accession number NC\_045512, Global Initiative for Sharing All Influenza Data (GISAID) accession ID: EPI ISL 402124)[6]. According to the GISAID, currently, there are six major clades of SARS-CoV-2 genomes sequenced. SARS-CoV-2 clades are afterward named S, G, V, GR, GH, and L. In addition, clade O represents mutations that do not belong to any other clades. Clade L includes the reference sequence. Six major clades are diverged from each other by small changes according to the reference sequence. Clade S contains L84S mutation in nonstructural protein 8 (NSP8). For clade V, the coexisting mutation of L37F and G251V in nonstructural proteins 6 and 3, respectively (NSP6 and NSP3), is characterized. Clade G contains the spike protein (S) mutation of D614G. For clades GH and GR, in addition to D614G mutation, NS3-Q57H and N-G204R are found, respectively<sup>[7]</sup>.

SARS-CoV-2 is a single-stranded RNA virus with a positive polarity and a genome of 26–32 kb in size that encodes 27 proteins from 14 ORFs<sup>[8]</sup>. The first ORF (ORF1ab) covers two-thirds of the viral genome, which includes nonstructural protein NSP1– NSP16. One-third of the SARS-CoV-2 genome consists of major structural proteins, such as spike glycoprotein (S), membrane (M), envelope (E), and nucleocapsid (N) protein. In addition to these proteins, SARS-CoV-2 has six accessory proteins<sup>[9]</sup>. Spike

protein has two subunits, S1 and S2, which are known as crucial factors for the entry of coronavirus into the host cells. While S1 is responsible for recognizing and binding to the host cells through angiotensin-converting enzyme 2 (ACE2) receptor, S2 has a role in host cell membrane fusion<sup>[10]</sup>. Therefore, genomic variations, especially in spike mutations, are a global interest for new vaccines and drug designing. Analyzing genomic variations is also crucial for understanding the mechanism of pathogenesis and viral drug resistance<sup>[11]</sup>.

The mortality rate of SARS-CoV-2 significantly differs from one country to another. For instance, while Turkey's mortality rate is 2.5%, the mortality rate is 0.9% in Singapore. Compared with Turkey and Singapore, Belgium has a significantly high mortality rate, which is 15.4%[12]. The differences in mortality rates are considered an important pattern to better understand the virus's mutation rate and its ability to evolve. Therefore, this perspective review is conducted to gather information about predominant mutations of SARS-CoV-2 in one place by examining the origin of these mutations and their spreading pattern. Geographic distributions of predominant SARS-CoV-2 mutations and population dynamics in a time-specific manner can be the keystone to detect the pattern of viral evolution because variations may differ markedly across various regions over time. Therefore, we finally discussed the implications of these mutations in SARS-CoV-2 on the virus's virulence diversity.

# **Predominant Spike Protein Variations**

Due to spike protein's role in mediating viral entry into the host cell, most therapeutic agents based on antibodies targeted spike protein. Every genome sequenced so far (26<sup>th</sup> of June, 2020) demonstrates that D614G is the most prevalent transversion mutation, which changes nucleotide adenosine into guanosine at position 23403<sup>[13]</sup>.

# 1. D614G Variant

In the Wuhan reference sequence, a missense mutation at position 23403 in the spike protein-encoding gene was identified, which leads to an amino acid change from aspartate (D) to glycine (G) [13]. This variant was first reported on the 24<sup>th</sup> of January, 2020, in a sample from China (hCoV-19/Zhejiang/HZ103/2020; 24<sup>th</sup> of January, 2020), and four days later, the earliest D614G variant in Europe was reported in Germany (hCoV-19/Germany/BavPat1-ChVir929/2020; 28<sup>th</sup> of January, 2020)<sup>[14]</sup>.

To understand the origin of this variant, the first reported spike mutations until early March were analyzed. Among 183 samples available at that time, seven D614G mutations were found. Five of the D614G variants also had two other coexisting mutations: in the NSP3 gene, a silent C>T mutation at positions 3037 and 14409 existed. The earliest variant from Germany had a C>T mutation at position 3037 but not at 14409. In a very close time

frame, four D614G variants were sampled from China. On 7<sup>th</sup> of February, a Wuhan sequence was sampled, and it was found that neither of these two coexisting mutations were present. However, the other three sequences were closely related to a German sequence. In the sequence from Zhejiang sampled on 24<sup>th</sup> of January, all three mutations were identified. The other two sequences were sampled on 28<sup>th</sup> of January, and those sampled on 6<sup>th</sup> of February did not have a mutation at position 14409 like the German sequence. Therefore, it is not possible to conclude the origin of the D614 variant. It may have arisen from either China or Germany<sup>[14]</sup>.

G614 genotype was identified at a very low prevalence in early March. On 3<sup>rd</sup> of March, among the sequenced SARS-CoV-2 genomes, the number of genomes having D614 was found to be seven, while the number of genomes having G614 was 166. At the end of March (25th), the number of genomes having D614 was 1017, and the number of genomes having G614 rose to 429. By April, it sharply dominated and continued its expansion throughout Europe. The G614 variant form was introduced in America and Canada early in March and became dominant shortly after, as expected[14]. By the 25th of June, 2020, approximately 74% of all published sequences had the D614G variant<sup>[15]</sup>. According to the most recent updates on GISAID, D614G mutation occurred more than 144409 times, which accounts for 87.1% of sequences uploaded in GISAID (Figure 1). This significant increase in a concise time demonstrates a transmission advantage of G614 over D614. It is safe to conclude that D614G is globally distributed quickly, as seen in Figure 1.

From late December 2019 to mid-March 2020, of 1449 European SARS-CoV-2 genomes, 954 (66%) had D614G, and in 2795 worldwide genomes, D614G variation was found in 1237 (44%) isolates. After the submission of samples to GISAID between 17<sup>th</sup> and 30<sup>th</sup> March, it was seen that the G614 variant markedly became predominant worldwide<sup>[16]</sup>. Of 48635 SARS-CoV-2 samples as of 26<sup>th</sup> of June, 36500 had the D614G mutation in the spike protein that belongs to clade G, particularly predominant in Europe, Oceania, South America, and Africa<sup>[13]</sup>. According to data from the 15<sup>th</sup> of July, the D614G mutation was detected in 68 countries<sup>[17]</sup>.

Spike protein variations can be analyzed with three distinct clades; G, GR, and GH. Most recent data[18] provided by GISAID demonstrate that the contribution of G clade among other clades, which consists of the D614G mutation, is 32.6% in Africa, which is followed by 32.3% in Europe, 21.3% in South America, 16% in North America, and 11% in Asia, and finally the lowest distribution is detected in Australia and Oceania at 6.1%. Alternatively, the GR clade consisting of G204R-nucleocapsid mutation with D614G shows a higher frequency. For instance, GR clade makes up 60.6% of the clade distribution in South America, 41.6% in Europe, followed by 39.6% in Africa, 27.7% in Asia, 14.9% in Australia and Oceania, and 12.3% in North America. The third clade that consists of the NS3 Q57H mutation with D614G is the GH clade, which is also distributed frequently. 59.9% of the GH clade has been found in North America, followed by 21.9% in Asia. While, in Europe, GH consistutes 9.1% of the sequences, in South America, it constitutes 9%. GH clade variations are seen less frequently in Africa and Australia, and it is detected in Oceania at 6.6%. Clades having D614G mutation with and without coexisting mutations are shown in Figure 2 with distribution in six continents.

Due to this marked increase in samples harboring D614G mutation worldwide, understanding the potential resulting significance became the mutual aim of many scientists. It is suspected that nonsynonymous mutations of spike may have a markedly functional outcome, especially those in position 614, due to mutation's location in carboxy (C)-terminal of subdomain 1, S1<sup>[19]</sup>. This part of subdomain 1 precisely associates with subdomain 2, S2.

To test whether D614G has a functional outcome regarding transmission or replication, Zhang and colleagues experimented. They used pseudoviruses (PVs) pseudotyped with S-protein with and without 614 missense mutation (SD614 and SG614, respectively) from transfected human embryonic kidney 293T cells (HEK293T), which was then induced to express human angiotensin-converting enzyme-2 (hACE2-293T). It was found that there were nine-fold differences in the efficiency of infection between PV with G614 genotype and PV with D614 genotype. Therefore, mutant variant G614 showed higher infectivity than D614. The



Figure 1. Changes in the percentage of D614G mutation uploaded by months[13,14,18]

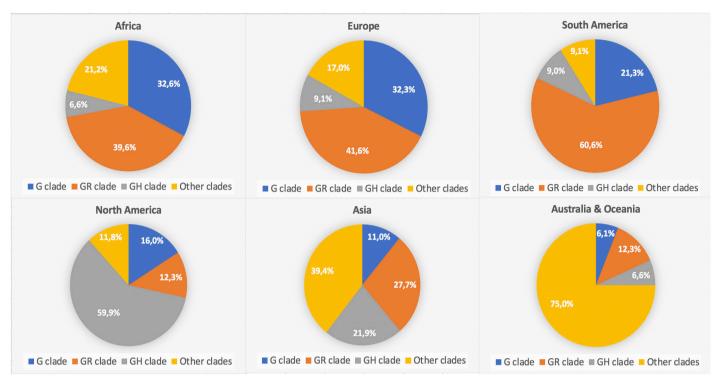


Figure 2. Distribution of clades (G, GR, and GH) across six continents<sup>[18]</sup>

same group also investigated the G164 variant mechanism by comparing S1 to S2 ratio between the variant and wild-type form. S1 to S2 ratio for PV<sup>G614</sup> was found 4.7 times higher than PV<sup>D614</sup>. Thus, changing aspartate to glycine leads to a stronger interaction between S1 and S2, thereby limiting S1 shedding. Another mechanism that enhances the total S-protein level compared with the wild-type was found<sup>[20]</sup>.

#### 2. A222V Variant

In spike protein at position 222, a novel mutation was found in summer 2020, and it spread very frequently: S: A222V. This is an amino acid substitution found in the N-terminal domain (NTD) of the spike protein. Even though NTD of spike protein is responsible for neither receptor binding nor membrane fusion, this A222V variant quickly rose and predominantly spread in Europe. The A222V variant was first detected in Spain (seven sequences) and Netherlands (one sequence) on 20<sup>th</sup> of June. Twenty-eight days later, the first sequence outside of Spain that harbored A222V was detected in England. Later, it was seen in Switzerland and Ireland and all-around Europe. Mid-August through September, A222V was found outside Europe for the first time and New Zealand and Hong-Kong, probably in patients with travel history from Europe<sup>[21]</sup>.

Analyzing cellular immune response and CD4+ and CD8+ T-cell responses is of importance for designing vaccines and developing therapeutics for the SARS-CoV-2<sup>[22]</sup>. Following the T-cell response is key to analyzing the cellular immune

response, and it was found that virus-specific T-cell long-lasting immunity was induced by SARS-CoV-2<sup>[23]</sup>. CD4+ T-cell responses are revealed mostly at site A222V, which is placed in B-cell epitopes. Therefore, this A222V mutation may affect the structure of B-cell epitopes<sup>[24]</sup>.

# Predominant Receptor-Binding Domain of Spike Variations

As previously described, the mechanism of spike protein-mediated virus entry in the host<sup>[25]</sup> is that S1 subunit of spike binds ACE-2 receptor by receptor binding domain (RBD)<sup>[26]</sup>. Two beta-sheets and three loops were found to have a role in the binding. In comparison with SARS-CoV, eight amino acids of 16 were conserved. The remaining eight may affect the binding affinity of RBD to ACE2<sup>[27]</sup>. Likewise, recombinant RBD of SARS-CoV was detected to enhance its binding to hACE2. Compared with SARS-CoV, SARS-CoV-2 was found to bind hACE2 significantly stronger, emphasizing the reason why SARS-CoV-2 is more transmissible<sup>[28]</sup>.

### 1. N501Y

During late autumn 2020, a new SARS-CoV-2 lineage VUI 202012/01 (Variant under Investigation, the year 2020, month 12, variant 01) containing N501Y variation co-occurring with  $\Delta H69/\Delta V70$  that causes the loss of two amino acids and several other spike protein variations has emerged and spread quickly in

the UK<sup>[29]</sup>. However, sequence harboring solely N501Y mutation was firstly detected in April 2020 in Brazil<sup>[30]</sup>.

N501Y variation is located in receptor-binding motif (RBM) of spike glycoprotein at position 501 and causes an amino acid change from asparagine to tyrosine<sup>[31]</sup>. This variation modifies the location where the virus contacts hACE2 and increases the virus's infectivity because it is located in one of the six major sites in RBD known as having a role in altering antigenicity<sup>[29,31]</sup>. Therefore, tracking the evolution and behavior of this variation helps understand the interaction between SARS-CoV-2 and hACE-2<sup>[29]</sup>.

### 2. E484K

Currently, three lineages become a major concern. Among them, B.1.351 lineage, which has K417N, E484K, and N501Y mutations, emerged in South Africa in August 2020<sup>[32]</sup>. Even though it was first identified in South Africa, it was detected to emerge separately in Brazil in early October and spread rapidly<sup>[33]</sup>.

E484K variation is located in RBM, main motif providing site for interaction with hACE2<sup>[34]</sup>, of spike glycoprotein at position 484 and causes an amino acid change from glutamic acid to lysine<sup>[32]</sup>. Co-occurrence of E484K and N501Y variations is suggested to be in accordance with the evolution of SARS-CoV-2. It was found that co-occurrence of E484K and N501Y causes more conformational changes in the protein structure than solely N501Y can do. Structural analysis demonstrates that E484K variation provides a new site for hACE2 binding. This binding is found to rigidify binding affinity with hACE2<sup>[35]</sup>.

Messenger RNA (mRNA) vaccines provide protection against infectious diseases by triggering an immune response inside the body by producing antibodies. The activity of mRNA vaccines against sites encoding E484K, N501Y alone, and their combination with K417N variation was tested. It was found that vaccine-elicited monoclonal antibodies (mAbs) target different sites of RBD similar to what happens in individuals who recovered from natural infection. However, the activity of these mAbs was reduced or abolished in the presence of these variants<sup>[36]</sup>.

#### 3. S477N-G485R

The receptor-binding domain may also be crucial due to its ability to shift the virus interaction with ACE2. Understanding the pattern of RBD mutations and their geographical differences is of importance in developing vaccines that target RBD. S477N is the major mutation found in RBD that caused an amino acid change to asparagine (N), which was first detected on 26<sup>th</sup> of January, 2020, in Victoria, Australia, and a while later, it dominated the region. Co-occurring mutations of S477N seen in Australia differ from those in the United Kingdom, demonstrating that the S477N variant is harbored separately in these lineages. S477N

variant was detected in solely 1% of SARS-CoV-2 sequences from Australia that were uploaded until June. Through August, this percentage rose to 90%. Even though the S477N variant made up solely 4.3% of all SARS-CoV-2 sequences worldwide and had never been seen in Africa, Asia, and South America, it dominated Oceania with 57.5%<sup>[37]</sup>. Another group analyzed 11571 SARS-CoV-2 sequences from various regions: 8564 from North America, 1426 from Oceania, 1017 from Asia, 441 from Europe, 103 from Africa, and 20 sequences from South America. One thousand and fifteen sequences of 1017 (99.8%) in Asia, all 103 sequences from Africa (100%), 422 sequences of 441 from Europe (95.7%), 8493 sequences of 8564 from North America (99.2%), all 20 sequences from South America, and finally 1087 sequences of 1426 from Oceania (76.2%) harbored the S477N variation according to sequences collected in July 2020<sup>[38]</sup>.

G485R is the third major mutation seen in spike protein and second major mutation of RBD after S477N. It was first seen in China on 6<sup>th</sup> of February, 2020<sup>[38]</sup>. One thousand and sixteen sequences of 1017 from Asia (99.9%), all 103 sequences from Africa (100%), 423 sequences of 441 from Europe (95.9%), 8514 sequences of 8564 from North America (99.6%), all 20 sequences from South America (100%), and 1086 sequences of 1426 from Oceania (76.2%) harbored the G485R mutation according to sequences collected in July 2020<sup>[39]</sup>.

S477N variation caused an amino acid change to asparagine (N), which may cause a shift in glycosylation sites of RBD<sup>[40]</sup>. Asparagine in RBD, in addition to S1 and S2 domains, can impact viral glycosylation. For instance, it is responsible for shielding distinct epitopes caused by antibody neutralization, spike protein folding, host–virus interaction, viral entry into a host, and immune evasion. Therefore, this change to asparagine affects the neutralization of antibody and viral pathogenicity<sup>[41]</sup>. There is a growing interest in vaccine candidates, depending on the glycoprotein. This is why S477N is significant for SARS–CoV-2 variations. These receptor-binding variations can change the relationship between virus and ACE2; they may cause the involvement of different receptors<sup>[38]</sup>.

## Predominant ORF1AB Variations

As mentioned before, ORF1ab constitutes two-thirds of the viral genome, which is translated by ORF1 as ORF1a and ORF1b. Nonstructural proteins 1–16 are the product of the cleavage of ORF1. Among these nonstructural proteins, the first 10 belong to ORF1a. By contrast, ORF1b comprises RdRp (NSP-12) through NSP-16<sup>[42]</sup>.

ORF1a is a component that can deal with cellular stresses and keep functioning. Its major role is mediating replication of the virus. In addition, ORF1b functions as helicase, exonuclease, and endonuclease<sup>[43]</sup>.

# 1. RNA-dependent RNA Polymerase (RdRp) Encoding Region of ORF1b Polyprotein

RdRp, also named NSP12, is a keystone of replication and transcription of RNA viruses, highly conserved due to the significant homology between SARS-CoV-2 and SARS-CoV-2<sup>[44]</sup>. Drugs that target and tightly bind to RdRp have been used against various viruses as well as SARS-CoV-2. Such drugs that inhibited RdRp are favipiravir, galidesivir, remdesivir, and ribavirin<sup>[44]</sup>. Therefore, mutations in RdRp can reduce drug efficacy by diminishing the binding affinity of drugs to RdRp. It is the main antiviral drug target<sup>[45,46]</sup>.

### 1.1. P323L

An RdRp missense mutation at position 14408 (C14408T) that leads to an amino acid change from proline to leucine (P323L in RdRp) was first detected in China on 24th of January, whereas it was uploaded on GISAID on 10th of April. The second case was reported in Italy on 20th of February. Two days later, a third case was detected in Australia. In Europe and North and South America, dominating P323L was accompanied by two other mutations, A23403G in spike (D614G) and C3037T in NSP3 (F106F), while P323L with these comutations remained minor in Asia. Shortly, after P323L dominance in Europe, P323L became prominent in North and South America and Africa, respectively. Among the sampled sequences (5th of May, 2020) from North and South America and Africa, viral genomes that had this variant with two other comutations became prominent as 81.3%, 59.4%, and 80.3%, respectively<sup>[47]</sup>. As of 1st of June, 2020, P323L was detected in 72.1% of SARS-CoV-2 isolates<sup>[45]</sup>.

Quick increment and dominance of P323L variation globally arouse curiosity regarding its characterization due to the potential of presenting better vaccines and treatments. P323L variation where proline changes to leucine leads to a much more stable structure due to leucine's stabilizing effects in the alpha helix. However, proline is a helix breaker. Therefore, a change from proline to leucine may rigidify structure and reduce molecular flexibility of RdRp<sup>[38]</sup>. These changes may lead to this quick increment and dominance of P323L by affecting RNA genome replication.

# 1.2. N628N-Y455Y-A97V

The top three single nucleotide polymorphism genotypes in RdRp (other than P323L) are N628N mutation, Y455Y mutation, A97V mutation. The first identification of N628N variation was in an Asian sample sequenced on January 22<sup>th</sup>, while Y455Y variation was first seen in the UK on 9<sup>th</sup> of February. Then, A97V variation was first seen in an Asian sample on 4<sup>th</sup> of March<sup>[47]</sup>.

Wang and colleagues demonstrated total frequencies of these variants. They genotyped 15140 SARS-CoV-2 isolates, 8309 of them had single mutations, uploaded until the 1st of June, 2020.

Among 15140 isolates, frequencies of Y455Y, N628N, and A97V were 8.2%, 2.7%, and 1.7%, respectively<sup>[48]</sup>.

Phylogenomic analysis from Western India also elicited that A97V was detected in more than 6% frequency. It is a nonsynonymous mutation at position 4489 in ORF1ab, which results in an amino acid change from alanine to valine at position 97 in RdRp<sup>[49]</sup>, where norovirus RdRp-associated nucleotidyltransferase (NiRAN) domain is located<sup>[44]</sup>. The secondary structure of RdRp was affected by the A97V variant by changing three alphahelixes with beta-sheets. Shifting the tertiary structure of RdRp may disrupt its RNA replication and maintain genomic fidelity function. These changes in RdRp may cause an increment in the rate of mutagenesis in the SARS-CoV-2 genome<sup>[50]</sup>.

A study found that while P323L raised the mutation rate, N628N had the opposite impact. Moreover, they found that mutation at N628N lowered the possibility of mutation in M and E protein by tenfold<sup>[47]</sup>.

Finally, Y455Y variation was found to have a strong relationship with mutations in M and E protein<sup>[47]</sup>.

#### 2. ORF1ab NSP3 Variations

ORF1a comprises NSP3, which is the greatest replicase subunit<sup>[51,52]</sup>. It is a multifunctional protein that can repress interferon responses by functioning as a viral protease<sup>[53]</sup>. NSP3 is suggested to provide a scaffold for association of replication and transcription complex<sup>[54]</sup>. Inhibitors of NSP3 may have the potential for preventing SARS-CoV-2 replication in host cells<sup>[55]</sup>.

# 2.1. F106F

NSP3 is a significant complex of replication and transcription<sup>[56]</sup>. The ORF1ab 3037C>T in NSP3 is a synonymous mutation<sup>[57]</sup> that causes an amino acid change to F106F. Among 10022 SARS-CoV-2 sequences uploaded from 1<sup>st</sup> of February to 1<sup>st</sup> of May, F106F is the most prevalent variation found in 60.2% of SARS-CoV-2 isolates<sup>[17]</sup>. On 4<sup>th</sup> of May, an analysis of 80 genomes of SARS-CoV-2 showed 57 isolates from Turkey harboring a high frequency of 3037C > T variation (71%), and 13 of them had a travel history from Saudi Arabia, while six of them had a travel history from Iran. Therefore, 3037C > T variation most probably spread to Turkey from these countries<sup>[58]</sup>.

# **Predominant Nucleocapsid Protein Variations**

Nucleocapsid protein has a markedly active role in the virus life cycle processes like building a ribonucleoprotein helical structure, while the genome is packaging, regulating RNA synthesis and transcription<sup>[59]</sup>. The nucleocapsid protein of SARS-CoV-2 is also responsible for the modulation of host cell metabolism. N protein creates an immune response during SARS-CoV-2 infection due to its extremely high immunogenic property<sup>[60]</sup>.

61485 nucleocapsid protein genes from the beginning of SARS-CoV-2 pandemic to 17<sup>th</sup> of July demonstrated 30221 variations in the nucleotide level and 28327 variations in amino acid level. Geographically distinct amino acid variations were mostly seen in Wales, India, England, Scotland, and Spain at 68.6%, 68.2%, 67.7%, 51.2%, and 42.9%, respectively[61].

### R203K-G204R

At amino acid position 203, five mutation types (K, M, S, I, and G) were found. R203K was detected as the most frequent variation, followed by G204R, and they are coexisting mutations<sup>[62]</sup>. R203K-G204R variations caused by change at position 28881-3 GGG/AAC in the SR-rich (serine-arginine) domain of nucleocapsid protein can affect encoded proteins by shifting their structure<sup>[63]</sup>. These variants were found in 68.1% and 67.9% of mutated strains of SARS-CoV-2, respectively[61]. Co-occurring variations of R203K and G204R were detected in 48 countries during the second wave of COVID-19 (data retrieved from 5th of August, 2020). R203K-G204R variation is markedly prominent in the UK, where their frequency was 15.4%, followed by Wales (4.3%)[63].

Analysis of 2492 SARS-CoV-2 genomes uploaded until 30<sup>th</sup> of March shows that solely R203K-G204R are two of the six other variants found in six regions, South and North America, Asia, Europe, Africa, and Australia<sup>[64]</sup>. Besides, R203K-G204R are also detected in China, the origin of SARS-CoV-2, which supports the idea that R203K-G204R variations might have emerged at the very beginning of the pandemic. 28881-3:AAC genotype was first detected on 24th of February in Europe (hCoV-19/ Netherlands/Berlicum 1363564/2020, EPI ISL 413565). However, it was identified in patients who had traveled to Italy. Besides, on 28th of February, the second occurrence of 28881-3 GGG/AAC was again collected from patients in Rome<sup>[65]</sup>. SARS-CoV-2 sequences uploaded till 9th of April was investigated, and it was found that Portugal had the highest frequency of R203K - G204R, 60%, followed by Netherlands, 50%, Belgium, 31%, and England, 26%. Alternatively, Spain remained at approximately 4%, while France remained at approximately 3%. In North America, the dominance of the 28881-3GGG genotype was observed in New York, which constitutes 95% of the SARS-CoV-2 genome having 28881-3 GGG instead of 28881-3 AAC genotype<sup>[66]</sup>. As of 30<sup>th</sup> of July, 2020, R203K-G204R variations continue to increase. In South America, their frequency was high, even though, in May, June, and July, very few sequences were uploaded. Data from 30<sup>th</sup> of July, 2020, signify that these variations are now prominent in Europe, Africa, and Asia [67].

R203K-G204R variation effects on pathogenicity are investigated by various scientists. These two amino acid changes added lysine, a polar hydrophilic positively charged amino

acid, between serine and arginine, thereby discontinuing SR dipeptide. SR dipeptide interruption blocks phosphorylation of the SR-rich domain, critical for basic N protein functions[66].

# **Predominant ORF8 Variation**

Among other accessory proteins (ORF3a, ORF6, ORF7a, ORF7b, and ORF10), ORF8 is significantly variable and susceptible to variations regarding the virus's capability of spreading [68]. It was detected that a large deletion of the ORF8 gene in patients with SARS-CoV-2 from Singapore and Taiwan also had a few other mutations in the viral genome<sup>[69,70]</sup>. It was concluded that a large deletion in the ORF8 gene leads to milder clinical symptoms<sup>[71,72]</sup>.

ORF8 protein functions as an interferon (IFN) antagonist<sup>[73]</sup> due to IFN inhibition activity by degrading interferon regulatory factor 3 (IRF3), which results in marked replication of the virus within the cells<sup>[74]</sup>. Besides, ORF8 in SARS-CoV-2 was found to prevent interferon-beta activity (IFN- $\beta$ )<sup>[75]</sup>. ORF8 is an important site that needs to be investigated deeply from both molecular and evolutionary aspects to broaden the significance of its role in SARS-CoV-2.

### L84S

In the Wuhan reference sequence, a missense mutation at position 28144 in ORF8, which causes an amino acid change from leucine to serine (L84S), was identified<sup>[76]</sup>. L84S is a nonsynonymous mutation that was identified on 30th December in China (MT291826). It emerged in China followed by other European countries, especially Spain, where it was dominant. Shortly after, in mid-January, it emerged in the USA<sup>[17]</sup>.

During the six months of its emergence, L84S (clade S84) did not spread globally, whereas it was found as the second major clade, which represented 10.8% of all sequenced SARS-CoV-2, after clade G614, which represented 71.1% of sequences. L84S spread mainly in North America (17%) and Oceania (16%)[77].

Clade S84 has one subclade, L84S/P5828L, which was first identified on 20th of February, 2020, in the USA (EPI ISL\_413456). The 10022 genomes of the virus were analyzed from 68 countries. The highest percentage of the sample was obtained from America (3543 samples), followed by England and Ireland. Among 10022 SARS-CoV-2 genomes, 5775 unique mutations were identified and 2969 of them were missense mutations. Among these missense mutations, 1669 were detected in L84S<sup>[17]</sup>.

The amino acid alteration caused by the L84S variant was thought to discard four hydrophobic bonds, destabilizing ORF8<sup>[78]</sup>. Basu and colleagues in their study demonstrated that L84S alteration is the most disruptive variation on protein stability among others, followed by the NSP4 mutation<sup>[79]</sup>.

# **Predominant ORF3a Variations**

ORF3a gene is the largest ORF in SARS-CoV-2 genome, consisting of 274 amino acids and in charge of forming an ion channel. Its localization on the cell surface makes it highly potential for virus-host interaction; it releases the virus from the host<sup>[80]</sup>. ORF3a works with several signal transduction pathways such as c-Jun N-terminal kinase and nuclear factor kappa B, thereby, producing proinflammatory cytokine and chemokine responses, which results in an aggravation of the host response in patients with SARS-CoV-2 if their production is unregulated<sup>[81]</sup>.

Apoptosis is one of the major defense mechanisms of the host against viral infections<sup>[82,83]</sup>. ORF3a of SARS-CoV has a significant regulation in apoptosis<sup>[84]</sup>. In SARS-CoV-2, a study demonstrated that cells had a markedly increased apoptosis marker caspase-3 when ORF3a was in the environment, thereby inducing apoptosis<sup>[85]</sup>.

### 1. Q57H

25563G > T is a missense mutation that causes an amino acid change from glutamine to histidine at 57 positions (Q57H) of ORF3a. This variation was first found in Australia on 5<sup>th</sup> of February, 2020 (EPI\_ISL\_480608). In contrast, it was claimed that this variation was first seen in Singapore on 16<sup>th</sup> of February<sup>[86]</sup>. Q57H accounted for 28.1% of sequences uploaded to GISAID globally until 6<sup>th</sup> of July, 2020. It was distributed mainly in South and North America with a frequency of 59.5%<sup>[87]</sup>.

The genotyping analysis of 28726 SARS-CoV-2 genomic sequences from America up to the 14<sup>th</sup> of July, 2020, demonstrated that Q57H is the top third variation in the United States, which was detected in America for the first time on 28<sup>th</sup> of February<sup>[86]</sup>.

Q57H variation may affect apoptosis activity of ORF3a and leads to increment in the viral load in the host<sup>[79]</sup>. The location of Q57H variation may cause some potential harm. It is located near TNF receptor-associated factor and caveolin binding domain<sup>[88]</sup>. The variation in this location can cause functional deficits in innate immune signaling receptor, nod-like receptor pyrin domain-containing 3 (NLRP3) inflammasome activity<sup>[89]</sup>. Because of NLRP3's ability to prompt an inflammatory response, it may be a potential target for therapeutics of SARS-CoV-2.

#### 2. G251V

26144G > T is a missense mutation that causes an amino acid change from glycine to valine at 251 positions (G251V) of ORF3a. G251V variation is lately defined as marker mutation of clade  $V^{[90]}$ . As of  $2^{nd}$  of April, 2020, G251V was detected in all regions except Africa and Austria<sup>[10]</sup>. A highly detailed study analyzed 48635 SARS-CoV-2 genomic sequences on  $26^{th}$  of June, and it was found that 3792 sequences harbored the G251V variation

 $(7.7\%)^{[13]}$ . A study that analyzed 537 SARS-CoV-2 genomes found that G251V was the most common variation seen in 53 samples  $(9.9\%)^{[91]}$ .

After investigating this variation, it was found that G251V caused the loss of phosphatidylinositol-specific phospholipase X-box domain (PIPI\_CX\_DOMAIN) B-cell like epitope, thereby affecting signal transduction pathways<sup>[92]</sup>.

# Conclusion

The ability of RNA viruses to mutate rapidly makes them prone to gain resistance to drugs and evade immunological surveillance<sup>[93]</sup>. Mutations and genomic diversity of SARS-CoV-2 may be the keystone to a better understanding of its pathogenesis and transmission. In this review, we gathered significant variations in one place by time-dependently tracking their spreading pattern. The predominant variations and their first occurrence place and date were summarized in Table 1. Implications of these variations may demonstrate their significance regarding SARS-CoV-2 transmission and pathogenesis.

The most prominent variation is found in the spike protein, which is D614G. While its frequency was low at the beginning of March, it continued to increase at a high frequency and became dominant worldwide before March ended. Another spike variation was A222V, which originated in Europe in the summer and spread from there rapidly. In addition, S477N-G485R comutations have also been detected globally. Currently,

Table 1. Summary of the variations and date and location of first occurrence

Variation	First occurrence date	First identified location
L84S	30.12.2019	China
N628N	22.01.2020	Asia
D614G	24.01.2020/28.01.2020	China/Germany
P323L	24.01.2020	China
F106F	24.01.2020	China
S477N	26.01.2020	Australia
Q57H	5.02.2020	Australia
R203K	24.02.2020	Netherlands
A97V	3.04.2020	Asia
G485R	2.06.2020	China
A222V	20.07.2020	Spain/Netherlands
Y455Y	2.09.2020	England
N501Y	December 2020	England
E484K	August 2020	South Africa

circulating N501Y and E484K variants having mutations in RBD of spike protein became a major concern due to their effects on the binding affinity of SARS-CoV-2 with hACE2. Surveillance of these new variations is vital due to the possibility that they may escape from immunity activated by vaccination[94]. Public concern regarding vaccines efficacy against these new strains has emerged. Current vaccines were tested against them and a sixfold decrement was observed in neutralizing titer levels against a B.1.351 strain, containing N501Y, K417N, and E484K variations, which has evolved in South Africa. Redesigning the current vaccines may become an issue, as reported by Tanne<sup>[95]</sup>. Because spike variation is responsible for mediating viral entrance into a host, most therapeutics target spike protein.

ORF1 variations are another class that needs attention due to their roles in virus replication. P323L is the major variation in that class because it is distributed globally in a concise time. N628N-Y455Y-A97V were found to be in the top four variations in RdRp after P323L. F106F was found in NSP3 of ORF1a that comprises NSP3, which is the largest replicase subunit. Tracking the variations in NSP3 may also have the potential to prevent SARS-CoV-2 replication in host cells[55].

Nucleocapsid protein plays an active role in the regulation of RNA synthesis and transcription. R203K is one of the most prominent variations of the nucleocapsid protein, which cooccurs with G204R mostly. These variations change the SR region. In the second wave, these variations are found in 48 countries, which show their global distribution<sup>[58]</sup>. Due to their effect on the SR region, these variations may interfere with the function of the nucleocapsid protein.

L84S variation is in ORF8 accessory protein. Even though it is one of the earliest variations, detected on 30th of December in China, it spread very slowly. After six months of its emergence, it became a variant seen in six continents. It is the second major clade after the D614G clade. This variation may cause destabilization of the protein structure.

Q57H and G251V were found in ORF3a accessory protein. Until midsummer, Q57H was the third major variation in America. It may affect the apoptosis activity of ORF3a. G251V variation is defined as a marker mutation of clade V. It was detected in all continents except Africa and Australia until 2<sup>nd</sup> of April, 2020. This variation can affect the signal transduction pathway.

The mortality rate of SARS-CoV-2 significantly differs from one country to another. The differences in mortality rates are considered an important pattern to better understand the virus's mutation rate and its ability to evolve. Therefore, gathering information about predominant mutations of SARS-CoV-2 in one place by examining the origin of these mutations and their spreading pattern can illuminate research in the future. Tracking the geographic distributions of dominant SARS-CoV-2 mutations and population dynamics in a time-specific manner

can be the keystone to detect the pattern of viral evolution, thereby providing significant information regarding new drug and vaccine development.

#### **Fthics**

Peer-review: Externally and internally peer-reviewed.

# **Authorship Contributions**

Concept and Design: S.Ü., A.Ü.G., A.B., Data Collection or Processing: S.Ü., Analysis or Interpretation: S.Ü., A.Ü.G., A.B., Literature Search: S.Ü., A.B., Writing: S.Ü., A.Ü.G.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

#### References

- Ashour HM, Elkhatib WF, Rahman MM, Elshabrawy HA. Insights into the Recent 2019 Novel Coronavirus (SARS-CoV-2) in Light of Past Human Coronavirus Outbreaks. Pathogens. 2020;9:186.
- Lam WK, Zhong NS, Tan WC. Overview on SARS in Asia and the world. Respirology. 2003;8:2-5.
- Baharoon S, Memish ZA. MERS-CoV as an emerging respiratory illness: A review of prevention methods. Travel Med Infect Dis. 2019;32:101520.
- Lu L, Zhong W, Bian Z, Li Z, Zhang K, Liang B, Zhong Y, Hu M, Lin L, Liu J, Lin X, Huang Y, Jiang J, Yang X, Zhang X, Huang Z. A comparison of mortalityrelated risk factors of COVID-19, SARS, and MERS: A systematic review and meta-analysis. J Infect. 2020:81:18-25.
- Cucinotta D, Vanelli M. WHO Declares COVID-19 a Pandemic. Acta Biomed. 2020;91:157-60.
- Wu Z, Harrich D, Li Z, Hu D, Li D. The unique features of SARS-CoV-2 transmission: Comparison with SARS-CoV, MERS-CoV and 2009 H1N1 pandemic influenza virus. Rev Med Virol. 2020:2171.
- Yap PSX, Tan TS, Chan YF, Tee KK, Kamarulzaman A, Teh CSJ. An Overview of the Genetic Variations of the SARS-CoV-2 Genomes Isolated in Southeast Asian Countries. J Microbiol Biotechnol. 2020;30:962-6.
- Wang H, Li X, Li T, Zhang S, Wang L, Wu X, Liu J. The genetic sequence, origin, and diagnosis of SARS-CoV-2. Eur J Clin Microbiol Infect Dis. 2020;39:1629-35.
- Khan MI, Khan ZA, Baig MH, Ahmad I, Farouk AE, Song YG, Dong JJ. Comparative genome analysis of novel coronavirus (SARS-CoV-2) from different geographical locations and the effect of mutations on major target proteins: An in silico insight. PLoS One. 2020;15:0238344.
- 10. Lokman SM, Rasheduzzaman M, Salauddin A, Barua R, Tanzina AY, Rumi MH, Hossain MI, Siddiki AMAMZ, Mannan A, Hasan MM. Exploring the genomic and proteomic variations of SARS-CoV-2 spike glycoprotein: A computational biology approach. Infect Genet Evol. 2020;84:104389.
- 11. Laamarti M, Alouane T, Kartti S, Chemao-Elfihri MW, Hakmi M, Essabbar A, Laamarti M, Hlali H, Bendani H, Boumajdi N, Benhrif O, Allam L, El Hafidi N, El Jaoudi R, Allali I, Marchoudi N, Fekkak J, Benrahma H, Nejjari C, Amzazi S, Belyamani L, Ibrahimi A. Large scale genomic analysis of 3067 SARS-CoV-2 genomes reveals a clonal geo-distribution and a rich genetic variations of hotspots mutations. PLoS One. 2020;15:0240345.
- 12. Durmaz B, Abdulmajed O, Durmaz R. Mutations Observed in the SARS-CoV-2 Spike Glycoprotein and Their Effects in the Interaction of Virus with ACE-2 Receptor. Medeni Med J. 2020;35:253-60.

- Mercatelli D, Giorgi FM. Geographic and Genomic Distribution of SARS-CoV-2 Mutations. Front Microbiol. 2020;11:1800.
- Korber B, Fischer WB, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Foley B, Giorgi EE, Bhattacharya T, Parker MD, Partridge DG, Evans CM, Silva TI. Spike Mutation Pipeline Reveals the Emergence of a More Transmissible Form of SARS-CoV-2. BioRxiv. 2020;1-33.
- Yurkovetskiy L, Wang X, Pascal KE, Tomkins-Tinch C, Nyalile TP, Wang Y, Baum A, Diehl WE, Dauphin A, Carbone C, Veinotte K, Egri SB, Schaffner SF, Lemieux JE, Munro JB, Rafique A, Barve A, Sabeti PC, Kyratsous CA, Dudkina NV, Shen K, Luban J. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. Cell. 2020;183:739-51.
- Isabel S, Graña-Miraglia L, Gutierrez JM, Bundalovic-Torma C, Groves HE, Isabel MR, Eshaghi A, Patel SN, Gubbay JB, Poutanen T, Guttman DS, Poutanen SM. Evolutionary and structural analyses of SARS-CoV-2 D614G spike protein mutation now documented worldwide. Sci Rep. 2020;10:14031.
- 17. Koyama T, Platt D, Parida L. Variant analysis of SARS-CoV-2 genomes. Bull World Health Organ. 2020;98:495-504.
- COVID-19 VirusMutation Tracker. Available from: www.cbrc.kaust.edu.sa/ covmt/index.php?p=vis-clade-continent
- Laha S, Chakraborty J, Das S, Manna SK, Biswas S, Chatterjee R. Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission. Infect Genet Evol. 2020;85:104445.
- Zhang L, Jackson CB, Mou H, Ojha A, Rangarajan ES, Izard T, Farzan M, Choe H. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. bioRxiv [Preprint]. 2020:2020.06.12.148726.
- Hodcroft EB, Zuber M, Nadeau S, Crawford KHD, Bloom JD, Veesler D, Vaughan TG, Comas I, Candelas FG, Stadler T, Neher RA. Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. medRxiv [Preprint]. 2020:2020.10.25.20219063.
- Grifoni A, Weiskopf D, Ramirez SI, Mateus J, Dan JM, Moderbacher CR, Rawlings SA, Sutherland A, Premkumar L, Jadi RS, Marrama D, de Silva AM, Frazier A, Carlin AF, Greenbaum JA, Peters B, Krammer F, Smith DM, Crotty S, Sette A. Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. Cell. 2020;181:1489-501.
- 23. Snyder TM, Gittelman RM, Klinger M, May DH, Osborne EJ, Taniguchi R, Zahid HJ, Kaplan IM, Dines JN, Noakes MN, Pandya R, Chen X, Elasady S, Svejnoha E, Ebert P, Pesesky MW, De Almeida P, O'Donnell H, DeGottardi Q, Keitany G, Lu J, Vong A, Elyanow R, Fields P, Greissl J, Baldo L, Semprini S, Cerchione C, Mazza M, Delmonte OM, Dobbs K, Carreño-Tarragona G, Barrio S, Imberti L, Sottini A, Quiros-Roldan E, Rossi C, Biondi A, Bettini LR, D'Angio M, Bonfanti P, Tompkins MF, Alba C, Dalgard C, Sambri V, Martinelli G, Goldman JD, Heath JR, Su HC, Notarangelo LD, Martinez-Lopez J, Carlson JM, Robins HS. Magnitude and Dynamics of the T-Cell Response to SARS-CoV-2 Infection at Both Individual and Population Levels. medRxiv. 2020;1-33
- 24. To KK, Hung IF, Ip JD, Chu AW, Chan WM, Tam AR, Fong CH, Yuan S, Tsoi HW, Ng AC, Lee LL, Wan P, Tso E, To WK, Tsang D, Chan KH, Huang JD, Kok KH, Cheng VC, Yuen KY. COVID-19 re-infection by a phylogenetically distinct SARS-coronavirus-2 strain confirmed by whole genome sequencing. Clin Infect Dis. 2020:1275.
- 25. Gallagher TM, Buchmeier MJ. Coronavirus spike proteins in viral entry and pathogenesis. Virology. 2001;279:371-4.
- 26. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science. 2020;367:1444-8.
- 27. Chen Y, Guo Y, Pan Y, Zhao ZJ. Structure analysis of the receptor binding of 2019-nCoV. Biochem Biophys Res Commun. 2020;525:135-40.
- 28. Tai W, He L, Zhang X, Pu J, Voronin D, Jiang S, Zhou Y, Du L. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. Cell Mol Immunol. 2020;17:613–20.

- Kemp SA, Harwey WT, Datir RP, Collier DA, Ferreira IATM, Carabelli AM, Robertson DL, Gupta RK. Recurrent Emergence and Transmission of a SARS-CoV-2 Spike Deletion H69/V70. 202 Preprint. bioRxiv. 2020;1-23.
- 30. Voloch CM, da Silva Francisco R Jr, de Almeida LGP, Cardoso CC, Brustolini OJ, Gerber AL, Guimarães APC, Mariani D, da Costa RM, Ferreira OC Jr; Covid19-UFRJ Workgroup, LNCC Workgroup, Adriana Cony Cavalcanti, Frauches TS, de Mello CMB, Leitão IC, Galliez RM, Faffe DS, Castiñeiras TMPP, Tanuri A, de Vasconcelos ATR. Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil. J Virol. 2021:119-21.
- Sharma T, Baig MH., Rahim M, Dong JJ, Cho J. Unbuttoning the impact of n501y mutant rbd on viral entry mechanism: A computational insight. Preprint. bioRxiv. 2021;1-15.
- 32. Gröhs Ferrareze PA, Franceschi VB, de Menezes Mayer A, Caldana GD, Zimerman RA, Thompson CE. E484K as an innovative phylogenetic event for viral evolution: Genomic analysis of the E484K spike mutation in SARS-CoV-2 lineages from Brazil. Preprint. bioRxiv. 2021;1-30.
- 33. Toovey OTR, Harvey KN, Bird PW, Tang JWW. Introduction of Brazilian SARS-CoV-2 484K.V2 related variants into the UK. J Infect. 2021:S0163-4453(21)00047-5.
- 34. Tegally H, Wilkinson E, Lessells RJ, Giandhari J, Pillay S, Msomi N, Mlisana K, Bhiman JN, von Gottberg A, Walaza S, Fonseca V, Allam M, Ismail A, Glass AJ, Engelbrecht S, Van Zyl G, Preiser W, Williamson C, Petruccione F, Sigal A, Gazy I, Hardie D, Hsiao NY, Martin D, York D, Goedhals D, San EJ, Giovanetti M, Lourenço J, Alcantara LCJ, de Oliveira T. Sixteen novel lineages of SARS-CoV-2 in South Africa. Nat Med. 2021 Feb 2.
- 35. Nelson G, Buzko O, Spilman P, Niazi K, Rabizadeh S, Soon-Shiong P. Preprint. bioRxiv. Molecular dynamic simulation reveals E484K mutation enhances spike RBD-ACE2 affinity and the combination of E484K, K417N and N501Y mutations (501Y.V2 variant) induces conformational change greater than N501Y mutant alone, potentially resulting in an escape mutant. Preprint. bioRxiv. 2021;1-13.
- 36. Wang Z, Schmidt F, Weisblum Y, Muecksch F, Barnes CO, Finkin S, Schaefer-Babajew D, Cipolla M, Gaebler C, Lieberman JA, Oliveira TY, Yang Z, Abernathy ME, Huey-Tubman KE, Hurley A, Turroja M, West KA, Gordon K, Millard KG, Ramos V, Silva JD, Xu J, Colbert RA, Patel R, Dizon J, Unson-O'Brien C, Shimeliovich I, Gazumyan A, Caskey M, Bjorkman PJ, Casellas R, Hatziioannou T, Bieniasz PD, Nussenzweig MC. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. Nature. 2021 Feb 10.
- Chen AT, Altschuler K, Zhan SH, Chan YA, Deverman BE. COVID-19 CG enables SARS-CoV-2 mutation and lineage tracking by locations and dates of interest. Elife. 2021;10:63409.
- Begum F, Banerjee AK, Ray Ul. Mutation Hot Spots in Spike Protein of SARS-CoV-2 Virus. Preprints. 2020 16 April.
- Nguyen TT, Pathirana PN, Nguyen T, Nguyen QVH, Bhatti A, Nguyen DC, Nguyen DT, Nguyen ND, Creighton D, Abdelrazek M. Genomic mutations and changes in protein secondary structure and solvent accessibility of SARS-CoV-2 (COVID-19 virus). Sci Rep. 2021;11:3487.
- 40. Watanabe Y, Allen JD, Wrapp D, McLellan JS, Crispin M. Site-specific glycan analysis of the SARS-CoV-2 spike. Science. 2020;369:330-3.
- 41. Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, Guo L, Guo R, Chen T, Hu J, Xiang Z, Mu Z, Chen X, Chen J, Hu K, Jin Q, Wang J, Qian Z. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nat Commun. 2020;11:1620.
- 42. Banerjee S, Seal S, Dey R, Mondal KK, Bhattacharjee P. Mutational spectra of SARS-CoV-2 orf1ab polyprotein and signature mutations in the United States of America. J Med Virol. 2021;93:1428-35.
- Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, O'Meara MJ, Rezelj VV, Guo JZ, Swaney DL, Tummino TA, Hüttenhain R, Kaake RM, Richards AL, Tutuncuoglu B, Foussard H, Batra J, Haas K, Modak M, Kim M, Haas P, Polacco BJ, Braberg H, Fabius JM, Eckhardt M, Soucheray M, Bennett MJ, Cakir M, McGregor MJ, Li Q, Meyer B, Roesch F, Vallet T, Mac Kain A,

- Miorin L, Moreno E, Naing ZZC, Zhou Y, Peng S, Shi Y, Zhang Z, Shen W, Kirby IT, Melnyk JE, Chorba JS, Lou K, Dai SA, Barrio-Hernandez I, Memon D, Hernandez-Armenta C, Lyu J, Mathy CJP, Perica T, Pilla KB, Ganesan SJ, Saltzberg DJ, Rakesh R, Liu X, Rosenthal SB, Calviello L, Venkataramanan S, Liboy-Lugo J, Lin Y, Huang XP, Liu Y, Wankowicz SA, Bohn M, Safari M, Ugur FS, Koh C, Savar NS, Tran QD, Shengjuler D, Fletcher SJ, O'Neal MC, Cai Y, Chang JCJ, Broadhurst DJ, Klippsten S, Sharp PP, Wenzell NA, Kuzuoglu-Ozturk D, Wang HY, Trenker R, Young JM, Cavero DA, Hiatt J, Roth TL, Rathore U, Subramanian A, Noack J, Hubert M, Stroud RM, Frankel AD, Rosenberg OS, Verba KA, Agard DA, Ott M, Emerman M, Jura N, von Zastrow M, Verdin E, Ashworth A, Schwartz O, d'Enfert C, Mukherjee S, Jacobson M, Malik HS, Fujimori DG, Ideker T, Craik CS, Floor SN, Fraser JS, Gross JD, Sali A, Roth BL, Ruggero D, Taunton J, Kortemme T, Beltrao P, Vignuzzi M, García-Sastre A, Shokat KM, Shoichet BK, Krogan NJ. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature. 2020;583:459-68.
- 44. Gao Y, Yan L, Huang Y, Liu F, Zhao Y, Cao L, Wang T, Sun Q, Ming Z, Zhang L, Ge J, Zheng L, Zhang Y, Wang H, Zhu Y, Zhu C, Hu T, Hua T, Zhang B, Yang X, Li J, Yang H, Liu Z, Xu W, Guddat LW, Wang Q, Lou Z, Rao Z. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. Science. 2020;368:779-82.
- Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, Masciovecchio C, Angeletti S, Ciccozzi M, Gallo RC, Zella D, Ippodrino R. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. J Transl Med. 2020;18:179.
- Velkov T, Carbone V, Akter J, Sivanesan S, Li J, Beddoe T, Marsh GA. The RNA-dependent-RNA polymerase, an emerging antiviral drug target for the Hendra virus. Curr Drug Targets. 2014;15:103-13.
- 47. Eskier D, Karakülah G, Suner A, Oktay Y. RdRp mutations are associated with SARS-CoV-2 genome evolution. PeerJ. 2020;8:9587.
- 48. Wang R, Hozumi Y, Yin C, Wei GW. Decoding SARS-CoV-2 Transmission and Evolution and Ramifications for COVID-19 Diagnosis, Vaccine, and Medicine. J Chem Inf Model. 2020;60:5853-65.
- 49. Paul D, Jani K, Kumar J, Chauhan R, Seshadri V, Lal G, Karyakarte R, Joshi S, Tambe M, Sen S, Karade S, Anand KB, Shergill SPS, Gupta RM, Bhat MK, Sahu A, Shouche YS. Phylogenomic analysis of SARS-CoV-2 genomes from western India reveals unique linked mutations. Preprint. BioRxiv. 2020;1-32.
- Banerjee A, Sarkar R, Mitra S, Lo M, Dutta S, Chawla-Sarkar M. The Novel Coronavirus Enigma: Phylogeny and Analyses of Coevolving Mutations Among the SARS-CoV-2 Viruses Circulating in India. JMIR Bioinformatics Biotechnol. 2020;1:20735.
- Snijder EJ, Bredenbeek PJ, Dobbe JC, Thiel V, Ziebuhr J, Poon LL, Guan Y, Rozanov M, Spaan WJ, Gorbalenya AE. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J Mol Biol. 2003;331:991-1004.
- Thiel V, Ivanov KA, Putics Á, Hertzig T, Schelle B, Bayer S, Weißbrich B, Snijder EJ, Rabenau H, Doerr HW, Gorbalenya AE, Ziebuhr J. Mechanisms and enzymes involved in SARS coronavirus genome expression. J Gen Virol. 2003;84:2305–15.
- 53. Forni D, Cagliani R, Clerici M, Sironi M. Molecular Evolution of Human Coronavirus Genomes. Trends Microbiol. 2017;25:35-48.
- Imbert I, Snijder EJ, Dimitrova M, Guillemot JC, Lécine P, Canard B. The SARS-Coronavirus PLnc domain of nsp3 as a replication/transcription scaffolding protein. Virus Res. 2008;133:136-48.
- Khan MT, Zeb MT, Ahsan H, Ahmed A, Ali A, Akhtar K, Malik SI, Cui Z, Ali S, Khan AS, Ahmad M, Wei DQ, Irfan M. SARS-CoV-2 nucleocapsid and Nsp3 binding: an in silico study. Arch Microbiol. 2021;203:59-66.
- 56. Ugurel OM, Ata O, Turgut-Balik D. An updated analysis of variations in SARS-CoV-2 genome. Turk J Biol. 2020;44:157-67.
- Yin C. Genotyping coronavirus SARS-CoV-2: methods and implications. Genomics. 2020;112:3588-96.
- Rehman S, Mahmood T, Aziz E, Batool R. Identification of novel mutations in SARS-COV-2 isolates from Turkey. Arch Virol. 2020;165:2937-44.

- Liang Y, Wang ML, Chien CS, Yarmishyn AA, Yang YP, Lai WY, Luo YH, Lin YT, Chen YJ, Chang PC, Chiou SH. Highlight of Immune Pathogenic Response and Hematopathologic Effect in SARS-CoV, MERS-CoV, and SARS-Cov-2 Infection. Front Immunol. 2020;11:1022.
- 60. Kang S, Yang M, Hong Z, Zhang L, Huang Z, Chen X, He S, Zhou Z, Zhou Z, Chen Q, Yan Y, Zhang C, Shan H, Chen S. Crystal structure of SARS–CoV–2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. Acta Pharm Sin B. 2020;10:1228–38.
- Rahman MS, Islam MR, Alam ASMRU, Islam I, Hoque MN, Akter S, Rahaman MM, Sultana M, Hossain MA. Evolutionary dynamics of SARS-CoV-2 nucleocapsid protein and its consequences. J Med Virol. 2021;93:2177-95.
- 62. Maitra A, Sarkar MC, Raheja H, Biswas NK, Chakraborti S, Singh AK, Ghosh S, Sarkar S, Patra S, Mondal RK, Ghosh T, Chatterjee A, Banu H, Majumdar A, Chinnaswamy S, Srinivasan N, Dutta S, DAS S. Mutations in SARS-CoV-2 viral RNA identified in Eastern India: Possible implications for the ongoing outbreak in India and impact on viral structure and host susceptibility. J Biosci. 2020;45:76.
- 63. Patro LPP, Sathyaseelan C, Uttamrao PP, Rathinavelan T. Global variation in the SARS-CoV-2 proteome reveals the mutational hotspots in the drug and vaccine candidates. bioRxiv. 2020;1-25.
- Islam MR, Hoque MN, Rahman MS, Alam ASMRU, Akther M, Puspo JA, Akter S, Sultana M, Crandall KA, Hossain MA. Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity. Sci Rep. 2020:10:14004
- 65. Lorusso A, Calistri P, Mercante MT, Monaco F, Portanti O, Marcacci M, Cammà C, Rinaldi A, Mangone I, Di Pasquale A, Iommarini M, Mattucci M, Fazii P, Tarquini P, Mariani R, Grimaldi A, Morelli D, Migliorati G, Savini G, Borrello S, D'Alterio N. A "One-Health" approach for diagnosis and molecular characterization of SARS-CoV-2 in Italy. One Health. 2020;10:100135.
- Ayub MI. Reporting Two SARS-CoV-2 Strains Based on A Unique Trinucleotide-Bloc Mutation and Their Potential Pathogenic Difference. Preprints. 2020;1-14.
- 67. Arévalo SJ, Sifuentes DZ, Robles CH, Bianchi GL, Chávez AC, Casas RGS, Chavarría RP, Uceda-Campos G. Global Geographic and Temporal Analysis of SARS-CoV-2 Haplotypes Normalized by COVID-19 Cases during the First Seven Months of the Pandemic. Preprint. bioRxiv. 2020.
- 68. Pereira F. Evolutionary dynamics of the SARS-CoV-2 ORF8 accessory gene. Infect Genet Evol. 2020;85:104525.
- 69. Gong YN, Tsao KC, Hsiao MJ, Huang CG, Huang PN, Huang PW, Lee KM, Liu YC, Yang SL, Kuo RL, Chen KF, Liu YC, Huang SY, Huang HI, Liu MT, Yang JR, Chiu CH, Yang CT, Chen GW, Shih SR. SARS-CoV-2 genomic surveillance in Taiwan revealed novel ORF8-deletion mutant and clade possibly associated with infections in Middle East. Emerg Microbes Infect. 2020;9:1457-66.
- Su YCF, Anderson DE, Young BE, Linster M, Zhu F, Jayakumar J, Zhuang Y, Kalimuddin S, Low JGH, Tan CW, Chia WN, Mak TM, Octavia S, Chavatte JM, Lee RTC, Pada S, Tan SY, Sun L, Yan GZ, Maurer-Stroh S, Mendenhall IH, Leo YS, Lye DC, Wang LF, Smith GJD. Discovery and Genomic Characterization of a 382-Nucleotide Deletion in ORF7b and ORF8 during the Early Evolution of SARS-CoV-2. mBio. 2020;11:1610-20.
- 71. Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol. 2019;17:181–92.
- Muth D, Corman VM, Roth H, Binger T, Dijkman R, Gottula LT, Gloza-Rausch F, Balboni A, Battilani M, Rihtarič D, Toplak I, Ameneiros RS, Pfeifer A, Thiel V, Drexler JF, Müller MA, Drosten C. Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. Sci Rep. 2018;8:15177.
- Kopecky-Bromberg SA, Martínez-Sobrido L, Frieman M, Baric RA, Palese P. Severe acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and nucleocapsid proteins function as interferon antagonists. J Virol. 2007;81:548-57.
- 74. Wong HH, Fung TS, Fang S, Huang M, Le MT, Liu DX. Accessory proteins 8b and 8ab of severe acute respiratory syndrome coronavirus suppress

- the interferon signaling pathway by mediating ubiquitin-dependent rapid degradation of interferon regulatory factor 3. Virology. 2018;515:165-75.
- 75. Li JY, Liao CH, Wang Q, Tan YJ, Luo R, Qiu Y, Ge XY. The ORF6, ORF8 and nucleocapsid proteins of SARS-CoV-2 inhibit type I interferon signaling pathway. Virus Res. 2020;286:198074.
- Khailany RA, Safdar M, Ozaslan M. Genomic characterization of a novel SARS-CoV-2. Gene Rep. 2020;19:100682.
- Guan Q, Sadykov M, Mfarrej S, Hala S, Naeem R, Nugmanova R, Al-Omari A, Salih S, Mutair AA, Carr MJ, Hall WW, Arold ST, Pain A. A genetic barcode of SARS-CoV-2 for monitoring global distribution of different clades during the COVID-19 pandemic. Int J Infect Dis. 2020;100:216-23.
- Liu Q, Zhao S, Hou Y, Ye S, Sha T, Su Y, Zhao W, Bao Y, Xue Y, Chen H. Ongoing Natural Selection Drives the Evolution of SARS-CoV-2 Genomes. Preprint. MedRxiv. 2020;1-31.
- Basu S, Mukhopadhyay S, Das R, Mukhopadhyay S, Singh PK, Ganguli S. Impact of Clade Specific Mutations on Structural Fidelity of SARS-CoV-2 Proteins. Preprint. bioRxiv. 2020;1–50.
- Liu DX, Fung TS, Chong KK, Shukla A, Hilgenfeld R. Accessory proteins of SARS-CoV and other coronaviruses. Antiviral Res. 2014;109:97-109.
- Kanzawa N, Nishigaki K, Hayashi T, Ishii Y, Furukawa S, Niiro A, Yasui F, Kohara M, Morita K, Matsushima K, Le MQ, Masuda T, Kannagi M. Augmentation of chemokine production by severe acute respiratory syndrome coronavirus 3a/X1 and 7a/X4 proteins through NF-kappaB activation. FEBS Lett. 2006:580:6807-12.
- 82. Roulston A, Marcellus RC, Branton PE. Viruses and apoptosis. Annu Rev Microbiol. 1999;53:577-628.
- 83. O'Brien V. Viruses and apoptosis. J Gen Virol. 1998;79:1833-45.
- 84. Freundt EC, Yu L, Goldsmith CS, Welsh S, Cheng A, Yount B, Liu W, Frieman MB, Buchholz UJ, Screaton GR, Lippincott-Schwartz J, Zaki SR, Xu XN, Baric RS, Subbarao K, Lenardo MJ. The open reading frame 3a protein of severe acute respiratory syndrome-associated coronavirus promotes membrane rearrangement and cell death. J Virol. 2010;84:1097-109.
- 85. Ren Y, Shu T, Wu D, Mu J, Wang C, Huang M, Han Y, Zhang XY, Zhou W, Qiu Y, Zhou X. The ORF3a protein of SARS-CoV-2 induces apoptosis in cells. Cell Mol Immunol. 2020;17:881-3.

- Wang R, Chen J, Gao K, Hozumi Y, Yin C, Wei GW. Characterizing SARS-CoV-2 mutations in the United States. Preprint. Res Sq. 2020;3:49671.
- 87. Alam ASMRU, Islam OK, Hasan S, Al-Emran HM, Jahid IK, Hossain MA. Evolving Infection Paradox of SARS-CoV-2: Fitness Costs Virulence? Preprint. 2020;1-38.
- Hassan SS, Choudhury PP, Basu P, Jana SS. Molecular conservation and differential mutation on ORF3a gene in Indian SARS-CoV2 genomes. Genomics. 2020:112:3226-37.
- Shah A. Novel Coronavirus-Induced NLRP3 Inflammasome Activation: A Potential Drug Target in the Treatment of COVID-19. Front Immunol. 2020:11:1021.
- Capobianchi MR, Rueca M, Messina F, Giombini E, Carletti F, Colavita F, Castilletti C, Lalle E, Bordi L, Vairo F, Nicastri E, Ippolito G, Gruber CEM, Bartolini B. Molecular characterization of SARS-CoV-2 from the first case of COVID-19 in Italy. Clin Microbiol Infect. 2020;26:954-6.
- 91. Issa E, Merhi G, Panossian B, Salloum T, Tokajian S. SARS-CoV-2 and ORF3a: Nonsynonymous Mutations, Functional Domains, and Viral Pathogenesis. mSystems. 2020;5:00266-20.
- 92. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, Stockinger H. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. 2012;40:597-603.
- 93. Shen Z, Xiao Y, Kang L, Ma W, Shi L, Zhang L, Zhou Z, Yang J, Zhong J, Yang D, Guo L, Zhang G, Li H, Xu Y, Chen M, Gao Z, Wang J, Ren L, Li M. Genomic Diversity of Severe Acute Respiratory Syndrome-Coronavirus 2 in Patients With Coronavirus Disease 2019. Clin Infect Dis. 2020;71:713–20.
- 94. Emerging sars-cov-2 variants. (n.d.). Last access date: 2021, February 20. Available from https://www.cdc.gov/coronavirus/2019-ncov/more/science-and-research/scientific-brief-emerging-variants.html
- Tanne JH. Covid-19: Moderna plans booster doses to counter variants. BMJ. 2021;372:232.