Notable Variants of SARS-CoV-2

Hassan Naji

ABSTRACT

The current coronavirus pandemic has brought many deaths and affected economies around the world. The causative agent of COVID-19 is identified as SARS-CoV-2, a close relative to SARS-CoV that caused an outbreak of severe acute respiratory distress in 2002-2003. During the course of the pandemic, there has been many reports of mutations that occurred to SARS-CoV-2 with several lineages that has been identified around the globe. In this paper, a summary of the notable variants of SARS-CoV-2 is discussed.

Keywords: SARS-CoV-2, COVID-19, Variants, Lineage, B.1.1.7, Lineage B.1.351, Lineage B.1.1.207, Cluster 5, 501.V2 Variant, P.1 Variant, SARS-CoV-2 Genome, non-structural proteins, NSPs, Open reading frame 1ab, ORF1ab, Furin.

I. INTRODUCTION

SARS-CoV-2 is the cause of recent outbreak of COVID-19 around the globe. Since the emergence of the virus from Wuhan, China, and it is spread to many other countries, the scientific community engaged in the development of vaccines to save individuals from this virus. According to WHO [1] and as of February 12, 2021, the number of confirmed COVID-19 cases are 107, 423,526 including 2,360,280 deaths. The virus is constantly mutating, and several lineages of SARS-CoV-2 have been identified in various part of the globe.

SARS-CoV-2 is an enveloped virus with single-strand positive RNA genome of 26-32 kb in length. It belongs to the Coronaviridae family, within the genus beta-coronavirus. This genus is divided in four lineages (subgenus): lineage A, lineage B, lineage C and lineage D. Similar to SARS-CoV that was responsible for severe acute respiratory syndrome outbreak in 2002-03 in Asia, SARS-CoV-2 belongs to lineage B, also known as subgenus Sarbecovirus. The MERS-CoV (Middle East Respiratory Syndrome- Coronavirus), responsible for MERS outbreaks since 2012, is a member of lineage C. Bats are considered the natural reservoir of SARS-CoV-2. The phylogenetic analysis of SARS-CoV-2 shows 96% similarity of SARS-CoV-2 to RaTG13 isolate of bat coronavirus [2].

The positive sense RNA genome of SARS-CoV-2 contains 2 overlapping open-reading frame (ORFs) genes namely, ORF1a and ORF1b (Fig. 1). These genes comprise of two-third of the RNA genome of virus and are translated into various proteins from different ORFs. The remaining one-third genome codes 4 structural proteins, S, E, M, and N, and accessory gene termed ORF8. The accessory gene encodes non-structural proteins including, 3a, 6, 7a, 7b, 8, and 10. The accessory gene, ORF8, is only present in lineage B of β-coronaviruses and shows wide similarities in SARS-CoV-2 and SARS-CoV. The ORF8 gene contains 366 nucleotides and codes a protein with 121 amino acids. This ORF8 genome has been a source of mutations in the previous SARS-CoV outbreak [3].

II. MUTATIONS IN SARS-CoV-2

Since the emergence of SARS-CoV-2 in late 2019 in China, several novel mutations have risen in the viral genome, leading to many new variants of the SARS-CoV-2 in different countries. The neutralizing immunoglobulins produced after infection or vaccination should be highly potent and clinically effective for the treatment and prevention of COVID-19. Studies from SARS-CoV-2 convalescent plasma showed that some individuals elicited low levels of plasma neutralizing antibodies with decreased effectivity. Such low-level neutralizing antibodies can induce mutations for resistance in COVID-19 patients on re-infection. However, the pace and degree of mutations in SARS-CoV-2 resistance against neutralizing antibodies is still unclear [5].

Mutations exist in the S protein of the virus to evade the monoclonal or polyclonal effect produced by the host immune response through vaccination or infection. Koyama, Platt, and Parida [6] analyzed 10,022 genomes of SARS-CoV-2 and reported 65,776 variants. Of these mutations, there are 2 in-frame insertions, 11 frame-shift deletions, 36 stop gained variations, 66 non-coding insertions, 100 in-frame deletions, 142 non-coding deletions, 484 mutations in non-coding regions, 1,965 synonymous mutations and 2,969
missense mutations, comprising a total of 5,775 distinct mutations. In the 2,969 missense mutations, there are 1,905 variants in the open reading frame lab (ORF1ab). ORF1ab is the longest ORF of the viral genome. It transcribes into a multiprotein which is further cleaved into 16 non-structural proteins (NSPs). Out of these proteins, the maximum number of missense mutations were found in NSP3, within which, the most common type of mutations was observed in A58T followed by mutations in P153L [6].

Voloch et. al [7] reported the evolution of SARS-CoV-2 in 180 participants during increase in COVID-19 infection in Rio de Janeiro, Brazil in December 2020. Three main clades of SARS-CoV-2 were associated with COVID-19 epidemic in Rio de Janeiro, Brazil. They identified a new novel strain originated from lineage B.1.1.28. The new novel strain is distinguished from other strains by 5 single nucleotide variants (SNVs). These variants include C29754U, G28975U, G28628U, C28253U, and C100U. The mutation E484K was widespread in all samples [7]. Three Spike protein mutations: N501Y, 69-70 deletion and P681H has shown up many times in SARS-CoV-2 variants and are associated with increased antigenicity and viral transmissibility. Many more mutations have been reported in SARS-CoV-2 making the transmission of the virus to minks possible. Efforts are being made to collect and sequence new SARS-CoV-2 specimens. These new sequences are analyzed to determine and characterize emerging variants of SARS-CoV-2. This practice has led to the identification of another variant of SARS-CoV-2 in Nigeria. Detailed analysis is being done to determine the specific mutations and their impact on transmissibility and severity of infection in COVID-19 patients.

Characterization of the new variant of SARS-CoV-2 is important for various reasons. Some of these include, increased transmission rates, increased disease severity and mortality rates and less susceptibility to therapeutic agents such as monoclonal antibodies. New variants of SARS-CoV-2 with D614G mutation allow the virus to replicate and to be transmitted quicker in the epithelial cells of human respiratory tract as compared to wild type SARS-CoV-2, leading to greater spread of infection among individuals. Moreover, mutations in the emerging variants enable the virus to evade host immunity, induced either naturally or by vaccination. The immune response in both cases, produce polyclonal antibody response against several parts of the viral S protein. To evade the effect of these antibodies, the emerging variants accumulate mutations in the S protein. The new SARS-CoV-2 variants also have the ability to evade specific diagnostic tests for viral detection. For these reasons, strain surveillance of SARS-CoV-2 is being carried out on the base of new identified viral sequences [4].

III. LINEAGE B.1.1.7

The lineage B.1.1.7 or 20I/501Y.V1, VOC 202012/01, of SARS-CoV-2 is considered variant 01 of concern that emerged in South England in September 2020, and later reported in several other countries including the United States. According to WHO [1] and as of December 2020, lineage B.1.1.7 is reported in 31 other countries and territories. The viral genome of this lineage contains 23 mutations, out of which 13 mutations are non-synonymous, 6 are synonymous, and 4 are deletions. The non-synonymous mutations in S protein include Y144 del, 69-70 del, D1118H, P681H, A570D, N501Y, T7161 and S982A and 2 mutations in ORF8 stop codon. The synonymous mutations include 5 in ORF1ab namely, T1001I, A1708D, I2230T and GSF3675/3677 deletion and 1 in the M gene [8]. The N501Y mutation and the 69-70 deletion in the spike protein of lineage B.1.1.7 of SARS-CoV-2 are particularly associated with increased rate of transmission and increased severity of infection leading to higher death rate among COVID-19 patients [8]. The spontaneous transmissibility of SARS-CoV-2 due to N501Y mutation has been studied in mouse models and provided evidence for high transmission rate of the virus [9]. In the N501Y mutation, an arginine (N) at position 501 of the Receptor binding domain (RBD) of the S protein is replaced with tyrosine (Y). Tyrosine substitution in N501Y mutation increases the binding affinity of spike protein to ACE2 receptors because N501 is a key amino acid to binding with ACE2 receptor. The 69/70 deletion leads to conformational change in the S protein and decreases the sensitivity to neutralizing antibodies in human convalescent SARS-CoV-2 serum [10]. The P681H occurs close to the S1/S2 furin cleavage site and leads to high variability in SARS-CoV-2. The 69-70 deletion and the P681H mutation have spontaneously emerged several times. However, a study by Wu et al. [11] reported that the neutralizing ability of the non-human primate (NHP) sera and phase I participants vaccinated with mRNA-1273 was not reduced by the mutations in lineage B.1.1.7 [11].

Collier et al. [12] demonstrated the effect of B.1.1.7 S protein on the neutralization effect of antibodies present in individuals vaccinated with Pfizer BNT162b2 vaccine. They found that mutations in S protein in a pseudo-virus of the lineage B.1.1.7 cause decreased sensitivity in vaccinated individuals [12].

IV. LINEAGE B.1.351

The lineage B.1.351 or 20H/501Y.V2, of SARS-CoV-2 was first detected in Nelson Mandela Bay in South Africa, in October 2020. Since then, it has spread to many countries such as the U.S. and Zambia. This lineage is somehow similar to lineage B.1.17 but it emerged independently. The genome of this lineage contains 10 extensive mutations in the viral S protein, including D80A, L18F, D215G, D614G, L242-244del, E484K, K417N, A701V, R246I and N501Y. Three mutations, N501Y, E484K and K417N are located in the receptor-binding domain (RBD) of the S protein. RBD is the main target for the neutralizing antibodies. The key mutation E484K in B.1.351 lineage provides resistance to the neutralization effect of SARS-CoV-2 monoclonal antibody therapies and a panel of human convalescent sera [13]. Lineage B.1.351 also reduces neutralization effect of antibodies in non-human primate (NHP) sera and mRNA-1273 vaccinated individuals [11].

The D614G mutation in the S protein is due to a change from A-to-G mutation at position 23,403 in the reference strain of Wuhan, China. This mutation started to accumulate later in Europe and other countries. The D614G mutation enhances viral load in COVID-19 patients but does not
increase severity of the disease [14].

V. LINEAGE B.1.1.207

The lineage B.1.1.207 was identified by African Centre of Excellence for Genomics of Infectious Diseases (ACEGID), Redeemer’s University, Nigeria while analyzing SARS-CoV-2 genomic sequences. The lineage B.1.1.207 involves 2 SARS-CoV-2 sequences. The mutation P681H in the S protein of these sequences is similar to a mutation observed in lineage B.1.17 reported in the U.K. while the remaining mutations are different from those of lineage B.1.1.7. The P681H mutation has been noted directly adjacent to furin cleavage site; however, there is no evidence indicating the enhanced transmission of lineage B.1.1.207 variant with P681H mutation in Nigeria as compared to that reported in U.K. related to B.1.1.7 [15].

VI. CLUSTER 5

Cluster 5 was reported in mink and mink farm worker in Netherlands in April 2020, establishing that SARS-CoV-2 infection can spread from human to mink and vice versa. The combinations of mutations in cluster 5 variant of SARS-CoV-2 were not observed previously in any other variant. Since then, mink infection has been reported in Sweden, Spain, Italy, U.S. and Denmark. All the strains of mink-associated SARS-CoV-2 involved Y453F and D614G mutations in the S protein. Y453F mutation involves the substitution of tyrosine to phenylalanine at position no. 453 in the S protein. It was also observed during the outbreaks of Dutch mink farms. Cluster 5 variant contains 4 genetic changes in the S protein involving 1 deletion and 3 substitutions. Some strains of cluster 5 from Netherlands lacked this mutation, describing that this mutation is not necessary for infecting minks. In November 2020, Denmark officials reported 214 COVID-19 individuals infected with SARS-CoV-2 variants related to mink. These variants were characterized by specific variants related to mink that were identified in both animals and humans infected from mink farms [16].

All the strains of cluster 5 involved Y453F mutation in the RBD of the S protein. Some strains of cluster 5 from Netherlands lacked this mutation leading to the conclusion that this mutation is not necessary for infecting minks. Nearly half of cluster 5 mink related strains contain a deletion of histidine and valine at positions 69 and 70 in the N-terminal domain of S protein. Out of the 3 substitutions, in the S protein, I692V and M1229I are conservative substitutions located seven amino acids downstream and within the transmembrane domain, respectively. The third, non-conservative substitution is S1147L at position 1147 in S2 subunit [17].

VII. 501.V2 VARIANT

The 501.V2 variant was reported in Columbus, Ohio and South Africa. During the second wave of COVID-19 infection in December 2020 in South Africa, the new 501.V2 variant of SARS-CoV-2 was widely spread and became predominant. 501.V2 variant contains 3 mutations in the RBD, which increases its binding affinity to ACE2 receptors and potential neutralizing antibody escape. The 3 mutations are N501Y, E484K and K417N. In addition to these mutations, a 9-nucleotide deletion in NSP6 is also observed in the 501.V2 variant [18]. The lineage B.1.1.7 of SARS-CoV-2 also contains the N501Y mutation, but both the variants are phylogenetically different. Preliminary studies from South Africa suggests that 501.V2 variant causes a higher viral load resulting in greater transmissibility. Further investigations are required to the severity and possible therapeutics of this variant [1].

VIII. P.1 VARIANT

The P.1 variant of SARS-CoV-2 was detected in 4 travelers from Brazil at Haneda Airport, Tokyo, Japan. This variant is a branch off the lineage B.1.1.28 but genetically and phylogenetically belongs to lineage P.1. The genome of P.1 variant contains 17 different mutations, with 3 mutations in the RBD of the S protein namely, N501Y, E484K and K417T. The P.1 variant is now present throughout Brazil and has also been detected in a re-infected person in Salvador, State of Bahia, Brazil. A study by Claro et al. (2020) determined the frequency of E484K and N501Y mutations in P.1 variant and lineage B.1.1.28 of SARS-CoV-2. The frequency of E484K mutation was 100% in P.1 variant, and 13% in its descendent lineage B.1.1.28. The E484K mutation serves as an escape mutation for the neutralizing monoclonal antibody against SARS-CoV-2. It alters the antigenicity of the viral particle. According to the National Institute of Infectious Diseases in Japan (2021) there is a 10-fold reduction in the neutralizing ability of convalescent plasma administered against P.1 variant of SARS-CoV-2. Similarly, the frequency of N501Y mutation was found in 100% P.1 variants. The N501Y mutation provides greater binding specificity and rapidly growing lineages [19]. Sabino et al. [19] reported the presence of P.1 variant in 42% of samples sequenced for SARS-CoV-2 cases in Manaus, Amazon. It is evaluated that almost 75% of population of Manaus had been infected with this SARS-CoV-2 variant as of October 2020. Evidence suggests change in antigenic profile and transmissibility of P.1 variant reduce the neutralizing effect of antibodies generated after a previous infection or through vaccination [20].

Three points mutation are shared between P.1 variant and lineage B.1.351 in South Africa, including K417N/T, N501Y, E484K in the S protein. Both variants also contain a deletion in the ORF1b gene, that is, del11288-11296. This similarity is thought to be associated with increase in COVID-19 patients in areas where previous attack rates are considered high. The mutation N501Y and deletion in ORF1b, are shared between P.1 variant and lineage B.1.1.7. These shared mutations between P.1 variant, lineage B.1.351 and lineage B.1.1.7, arise independently [21].

IX. SUMMARY

It is important to identify the differences in the genomic sequences of different SARS-CoV-2 strains for the development and success of vaccine candidates. It is
necessary to decrease the infectivity and harmful effects of the virus to live a better and healthy life.

REFERENCES


