Received: 25 June 2021 | Accepted: 12 September 2021

DOI: 10.1002/jmv.27337

REVIEW

JOURNAL OF MEDICAL VIROLOGY WILEY

Genetic drift in the genome of SARS COV-2 and its global health concern

Iqra Bano[1] | Mehmoona Sharif[2] | Sadia Alam[1]

[1]Department of Microbiology, The University of Haripur, Haripur, Pakistan

[2]Department of Microbiology, Quaid I Azam University, Islamabad, Pakistan

Correspondence
Sadia Alam, Department of Microbiology, The University of Haripur, Haripur 22620, Pakistan.
Email: sadia.alam2004@gmail.com and sadia.alam1@uoh.edu.pk

Abstract

The outbreak of the current coronavirus disease (COVID-19) occurred in late 2019 and quickly spread all over the world. The severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) belongs to a genetically diverse group that mutates continuously leading to the emergence of multiple variants. Although a few antiviral agents and anti-inflammatory medicines are available, thousands of individuals have passed away due to emergence of new viral variants. Thus, proper surveillance of the SARS-CoV-2 genome is needed for the rapid identification of developing mutations over time, which are of the major concern if they occur specifically in the surface spike proteins of the virus (neutralizing analyte). This article reviews the potential mutations acquired by the SARS-CoV2 since the pandemic began and their xqsignificant impact on the neutralizing efficiency of vaccines and validity of the diagnostic assays.

KEYWORDS

COVID-19, genetics, mutation, SARS-CoV-2, serodiagnosis, vaccination, variants

1 | AN OVERVIEW OF COVID-19 PANDEMIC

Coronaviruses belong to a group of viruses that infect many organisms. They are responsible for mild to serious respiratory diseases. During the period of 2002–2012, two highly infectious coronaviruses of zoonotic origin, Middle East Respiratory Syndrome Coronavirus and severe acute respiratory syndrome coronavirus (SARS-CoV) emerged in humans and became a major problem of 21st century.[1] At the end of 2019, new deadly coronavirus emerged in Chinese city of Wuhan that causes unusual episodes of viral pneumonia and quickly spread globally.[2] On 30th December 2019, World Health Organization (WHO) declared this viral infection as the sixth Public Health Emergency of International Concern. This outbreak of COVID-19 has posed a remarkable threat to public health around the world.[3] On 11th February 2020, the new virus was declared to be severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by the International Committee on Taxonomy of Viruses. On the same day, this disease was titled COVID-19 by WHO.[4] As of May 6, 2021, over 162 million people from more than 210 countries have confirmed SARS-CoV-2 infection, and >3.3 million people have died due to COVID-19. An acceptable result to viral clearance was shown by a few antiviral drugs like remdesivir as well as anti-inflammatory drugs like tocilizumab.[5] Multiple vaccines such as messenger RNA (mRNA), adenovirus-vectored, protein subunit and inactivated SARS-CoV-2 vaccine are in clinical trials in several countries.[6] Pfizer-BioNTech, Moderna, AstraZeneca and Janssen (Johnson & Johnson) COVID-19 vaccines have received temporary authorization from different countries and WHO.

2 | THE EMERGENCE AND GENOMIC CHARACTERISTICS OF SARS-COV-2

A group of Chinese scientists isolated a bronchoalveolar lavage fluid sample of severe pneumonia patients and through meta-genomic sequencing of RNA, they discovered that Betacoronavirus is the cause of this new infection.[7] Initially, the sequence of the SARS-CoV-2 genome was revealed on January 10, 2020, in the Gene Bank, and the whole genome sequences were printed on January 12, 2020.[8] Based on

sequence alignment with phylogenetic investigation, SARS-CoV-2 is presently reported as the most up to date member of genus *Betacoronavirus* (β-CoV) within *Coronaviridae* family and *Nidovirales* order. The *Coronaviridae* family contain an enveloped virus having a nonsegmented genome of positive single strand RNA (ssRNA) with cap at the 5′ end and poly-A tail at the 3′ end, which itself act directly as mRNA for the formation of poly-proteins. Based on the analysis of the complete genome sequence, the genome of Beta-CoVs contains few nonstructural and four structural proteins such as spike, membrane, envelope, and nucleocapsid protein.[1] The genome of coronavirus is reported as the largest genome among the other known coronaviruses having 32%–43% GC content. The genomic sequence of SARS-CoV-2 shows different lengths that range from 29.8 to 29.9 kilo-base having 12 open reading frames (ORFs) encoding 27 different proteins.[9] More than 90% amino acids within the four structural genes of SARS-CoV-2 are identical with that of SARS-CoV, except for the S-gene which diverges.[10] The genome of SARS-CoV-2 does not contain the gene for hemagglutinin-esterase that is recognized in a few Beta-CoVs.[11] Approximately 2/3rd RNA of SARS-CoV-2 contains the region ORF1a/b having 16 nonstructural protein (nsp1-16) for the transcription and replication of virus and is considered as largest ORF (pp1ab). The remaining 1/3rd of the genome contains ORF that encodes structural and accessory proteins[12] (Figure 1).

3 | PHYLOGENETIC ANALYSIS AND TAXONOMY

The evolutionary tree analysis of complete genome showed correlation among SARS-CoV-2 and other coronaviruses that originate from bats and are grouped within the subgenus named Sarbecovirus

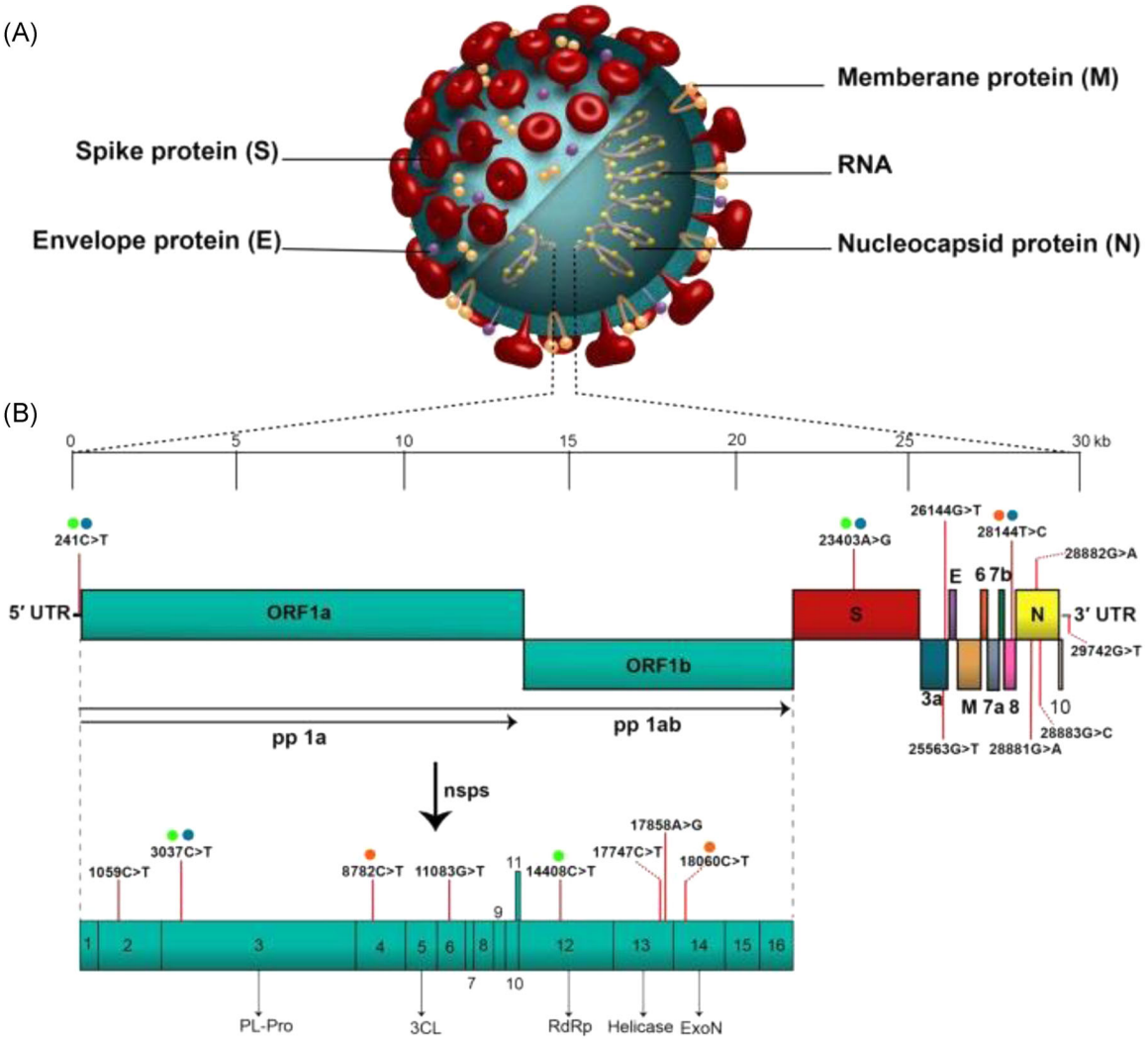

FIGURE 1 Schematic description of morphology and genome of SARS-CoV-2.(A) Virus is covered with S, M, and E protein. Inside phospholipid bilayers, the RNA is encompassed by the N-protein that is phosphorylated. (B) There are 29903 nucleotide bases and they contain 5′-UTR, ORF1a, and b that encodes 16 nonstructural proteins, 4 structural genes encoding S, M, N, and E proteins, 6 genes that code for ORF3a, 6, 7a, 7b, 8, and 10 accessory proteins, along with the 3′-UTR. The vertical red lines with circles having the same color on the genome indicate the position of 17 high-frequency mutations and co-mutations.[12] ORF, open reading frame; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; UTR, untranslated region

and genus Betacoronavirus. The matrix representation with the parsimony (MRP) pseudo-sequence supertree identified that RaTG13 (MN996532), bat-SL-CoVZC45 (MG772933), bat-SL-CoVZXC21 (MG772934), and SARS-CoV-2s constituted one major clade[13] (Figure 2). Particularly, the closest relative of SARS-CoV-2 is RaTG13 (MN996532) originated from bat *Rhinolophus affinis*, which has been previously reported by phylogenetic analysis of SARS-CoV-2 constructed with the genomic sequence.[14] MRP pseudo-sequence supertree also exhibited civet-sampled coronavirus (AY572035) as the closest relative of the SARS-CoVs.

SARS-CoV-2 has a 79% similar genome sequence with SARS and 50% with Middle East Respiratory Syndrome (MERS).[15] The spike proteins of SARS-CoV-2 have 1273 amino acids which are larger than that of SARS-CoV (1255) and bat SARSr-CoVs (1245–1269). It is different from other members within subgenus Sarbecovirus due to the S protein and 76.7%–77.0% sequence of amino acids are similar with SARS-CoVs from civets as well as humans, 75%–97.7% are similar with coronavirus found in bats within the same subgenus and 90.7%–92.6% showed similarity with coronavirus found in pangolins.[10] Another unique feature within the genome of SARS-CoV-2 is that it contains four amino acid residues (PRRA) within the intersection of S1 along with S2 subunits of spike protein. Polybasic cleavage site (RRAR) is produced due to these amino acid residues that permit efficacious cleavage by furin along with many proteases. It is confirmed from structural study that the furin cleavage site decreases the stability of spike protein within SARS-CoV-2 and encourage its receptor binding. As compared to SARS-CoV, SARS-CoV-2 is also highly transmissible due to the presence of the furin cleavage site.[16]

4 | GENETIC DIVERSITY AND PATHOGENICITY OF SARS-COV-2

The genetic diversity of SARS-CoV-2 is critical for its competency, durability as well as pathogenesis. One of the studies on SARS-CoV-2 origin showed that the major reason for the genetic diversity of the virus is random mutation and recombination.[17] The rate of mutation in SARS-CoV-2 is around 8×10^{-4} nucleotides/genome annually, which is very high for RNA viruses.[18] From the analysis of 220 genome sequences within the database, it has been revealed that as compared to Asia, the rate of mutation is high in Europe and North America. The genome of SARS-CoV-2 has nine putative recombinant patterns, containing six recombinant regions within S-protein and one in every RNA-dependent RNA polymerase, *nsp* 13 and *ORF* 3a.[19] Furthermore, the genome analysis recommended that the element for receptor binding within SARS-CoV-2 might conceivably emerge due to recombination between the coronavirus that was found in the pangolin along with RaTG13.[20] Mutation in the S-protein is a major issue of concern as it might alter tropism and pathogenicity of the virus. It has been predicted that mutation might enhance ACE-2 binding affinity, which is a key determinant of SARS-CoV-2 infectivity.[21]

5 | MUTATION AND GENETIC VARIATION

Mutation is one of the most important mechanisms that is responsible for the evolution of RNA viruses.[22] Different studies have been conducted for the recognition of genomic variation of SARS-CoV-2, and revealed different types of genetic variations including missense, insertion, noncoding, synonymous as well as deletion mutation.[23] According to the WHO, among 5775 distinct variants, the most frequent type of mutations were missense mutation (2969 variants) and synonymous mutations (1965) in SARS-CoV-2.[24]

In different studies, genetic analysis has reported mutations in a few genes which include ORFs like *ORF1ab*, *3a*, *6*, *7*, *8*, *10*, *S*, *N*, *E*, as well as *M*. However, *nsp1*, *nsp2 nsp3*, *nsp12*, and *nsp15* of ORF1ab, *ORF8* and *S* genes have also a large number of mutations among the other genes.[25] In addition, two insertion mutations with known effects were identified on ORF1ab.[15]

Among the other known mutations, the most common mutations are 241C>T placed on 5′-untranslated region (UTR), 14408C>T placed on nsp12, 3037C>T placed on nsp3, and 23403A>G.[26] In addition, 5′-UTR and 3′-UTR have noncoding mutations and may affect the packaging and titers of SARS-CoV-2.[27] Based on various studies, it has been found that frame-shift mutation also occurs in different regions of the genome, except M gene. These deletions alter the 3D structure of the virus which affects its virulency, pathogenesis, and host innate immune responses.[28]

6 | EFFECT OF MUTATION IN OUTBREAK OF SARS-COV-2

The sequence of SARS-CoV-2 genome showed more spot mutations on nsp12 as compared to Asian viral genome.[29] Reportedly, co-mutations were also found such as 241C>T (in 5′-UTR) with 3037C>T (F105F), 28144T>C (L84S), and 23403A>G (D614G) along with 8782C>T (S75S) with 28144T>C (L84S) and 18060C>T>C (L6L). In addition, 241C>T leader mutation in the European viral genome coexisted with three mutations such as 3037C>T (F105F), 14408C>T (P323L), and 23403A>G (D614G) that led to high COVID-19 infection rate, which showed that these four mutations play a key role in increasing viral transmission.[30] Similarly, in March 2020, another study showed that variants of SARS-CoV-2 having G614 within the spike protein replaced the original D614 form and became world dominant form. According to WHO, the largest clade was D614G, which had five subclades correlated with it. Moreover, almost every strain having D614G mutation altered the proteins for viral replication. As this protein is a target for antiviral drugs such as remdesivir and favipiravir, it might be possible that strains of SARS-CoV-2 become resistant to these drugs and multiply quickly.

7 | CLADES OF SARS-COV-2

Based on the Global initiative on sharing all influenza data (GISAID) nomenclature system, the genomes o SARS-CoV-2 were separated into seven major clades such as L to which the reference strain of

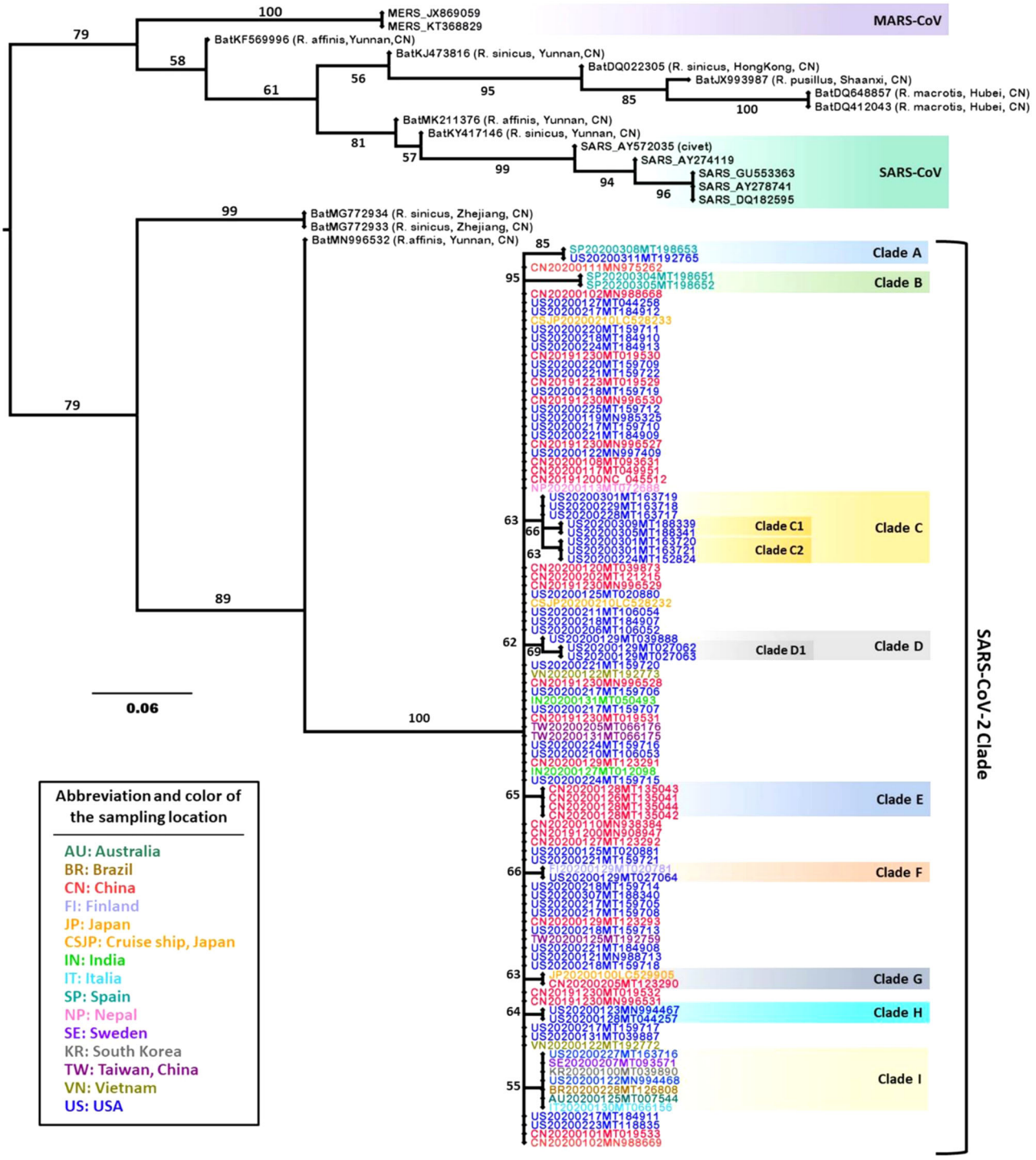

FIGURE 2 Phylogenetic supertree illustrated the evolution of SARS-CoV-2 by using a protein source.[14] MRP (Matrix representation with parsimony) pseudo-sequence supertree is constructed by using source phylogenetic trees for phylogenetic analysis of nine SARS-CoV-2 along with 5 SARS-CoV, 2 MERS-CoV, and 11 bat coronaviruses as outgroups. MAFFT (Multiple *Alignment* using Fast Fourier Transform) is used for the alignment of amino acid sequences and phylip file was formed by Clustal W. MRP supertree is constructed by using published supertree software Clann (version 4.2.4). By using PhyML program, ML (Maximum likelihood) phylogenies were utilized to construct source phylogenetic trees based with 100 bootstrap replications. FigTree v1.4.4 software is used for visualization of the phylogenetic tree. In the MRP pseudo-sequence supertree, SARS-CoV-2 is placed on one main branch while SARS-CoV and MERS-CoV belonged to another main branch. Particularly, MRP supertree analysis disputed RaTG13 bat coronavirus as the last common ancestor of SARS-CoV-2. MERS-CoV, Middle East Respiratory Syndrome Coronavirus; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2

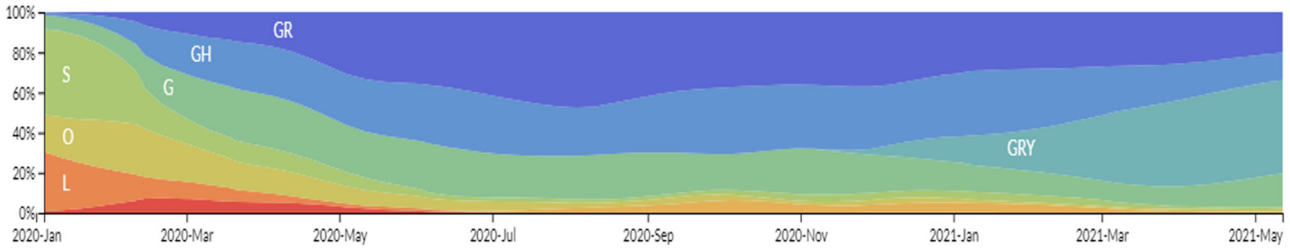

FIGURE 3 Frequencies of seven clades by Global initiative on sharing all influenza data (GISAID) nomenclature.[31] ACE-2, angiotensin-converting enzyme-2; β-CoV, betacoronavirus; COVID-19, coronavirus disease 2019; GISAID, Global initiative on sharing all influenza data; MAFFT, Multiple Alignment using Fast Fourier Transform; MERS-CoV, Middle East Respiratory Syndrome Coronavirus; ML, maximum likelihood; MRP, matrix representation with parsimony; nsp: nonstructural proteins; ORFs, open reading frames; RBD, receptor binding domain; RdRp, RNA-dependent RNA polymerase; SARS-Cov-2, severe acute respiratory syndrome coronavirus 2; S protein, spike protein; UTR, untranslated region

SARS-CoV-2 belongs, S, G, V, GH, GV, and GR. At the beginning of pandemic, in January 2020, the major clades were S, L, and O. S continued to be predominant at first while the L clade split into G and V. Furthermore, G split into GR and GH and after that into GV. After June 2020, GR split into GRY. As of March 2021, the GRY clade is taking up the greatest proportion of G clade (Figure 3). GRY clade represents the UK B.1.1.7 strain that has spread to over 90 countries. These clades come from mutations within the reference strain. Such mutations include L84S mutation in NS8 for clade S, L37F and G251V mutations for clade V, D614G mutation in S protein for clade G. Moreover, GH, GR, and GV clades are characterized by NS3-Q57H, N-G204R, and S-A222V mutations along with D614G mutation. However, the O clade stands for others that do not match any of the seven main clades.[31] The S clade is equal to PANGO A lineage (original virus). The G clade represents PANGO B.1 lineage with the GR clade equivalent to the PANGO B.1.1 lineage. The V clade is equal to PANGO B.2 lineage while L clade also represents another early lineage.

8 | EFFECT OF THE D614G VARIANT ON VACCINE EFFICACY

Efforts to synthesize an effective vaccine started after the release of primary viral sequence in January (2020).[32] However, the coronavirus since its first infection continuously developed new mutations that need to be investigated to ensure protection by the serum neutralization activity following natural infection or vaccination. Currently, about 30 vaccines are under clinical trials against SARS-CoV-2 and some of them have entered phase-3 clinical testing.[33] Data after vaccine trials on humans and in animal models suggested that disease caused by novel coronavirus can be prevented with neutralizing antibodies however, consistent mutations particularly in the spike-protein resulted in emerging new strains of SARS-CoV-2 which are of major concern. These repeated mutations raised a critical question of whether these new variants of the virus can be neutralized by the serum responses generated against the parental or early circulating strains. Among these spike mutations, D614G mutation was acquired early in the pandemic and has now become the world dominant form.[34] D614G mutation is a non-synonymous mutation occurred by the replacement of aspartic acid with glycine at 614 position of the viral S-protein. Most of the vaccines against SARS-CoV-2 were primarily developed from the D614 form of the virus that was found in China at the beginning of the pandemic.[32] Weissman et al.,[33] investigated that G614 mutation neither enhanced virus resistance against vaccines nor mediated in escape neutralization. However, it neutralized at a greater level by serum with the D614 form of the virus. This study also revealed that the G614 variants of SARS-CoV-2 are even more susceptible to the neutralizing antibodies induced against either strain of the virus. The serum response against the mRNA-LNP vaccine (nucleoside modified) not only appeared to recognize the G614 variant but also triggered robust immune response.[35,36]

The underlying mechanism seems to be the result of mutations in RBD (receptor binding domain) of the spike protein resulting in enhanced exposure of neutralization epitopes to antibodies. Even though G614 has substituted the unique D614 sequence in the novel coronavirus throughout the world, studies demonstrated that this is not an escape variation but rather more vulnerable to be neutralized by the sera of mice, nonhuman primates, and humans immunized with vaccines developed from the D614 form of the virus. So, the hurdles in the synthesis of an effective vaccine against SARS-CoV-2 are getting reduced.[33]

9 | OTHER SPIKE MUTATIONS: EFFECT ON NEUTRALIZATION ACTIVITY

Serum neutralization activity following natural infection or vaccination prevent viral infection, but an effective protection requires serum neutralization instead potency alone. This is due to the increased level of variation detected in some viral populations in major viral antigens.[37] Since the COVID-19 pandemic began, different SARS-CoV-2 population has been sequenced to assist detection of either single mutation in novel coronavirus. Currently, a new strain of the virus designated B.1.1.7 has appeared in United Kingdom (also called 20I/501Y.V1) that has multiple mutations in the RBD (receptor binding domain) and N-terminal domain of spike (target sites for neutralizing antibodies). Likewise, other variants B.1.351 have appeared in South Africa[38] and P.1 in Brazil.[39] There is the deletion mutation in B.1.351 and P.1 variants that include removal of

TABLE 1 Mutations adapted by the different variants of SARS-CoV-2 since December 2019 and their consequences

Variants	First discovery	Mutations identified	Consequence of mutations	References
1. Wild type	China in December 2019.	The original parental strain of virus without any mutations	–	43
2. B.1 (D614G)	China in early February 2020.	Replacement of aspartic acid with glycine at 614 position of the viral spike protein	• No increased viral resistance against vaccines but instead neutralized at greater level by the antibodies induced against D614 form of the virus. • Increased viral transmission and infectivity.	33,43
3. B.1.1.7 (N501Y)	Detected initially in UK in September 2020.	17 Mutations including 4 deletions and 13 nonsynonymous mutation in ORF1ab, ORF8 and N has been identified.	• Increased transmissibility. • Does not resist neutralization with postvaccine and convalescent serum however, moderately at reduced level.	40,43,44
4. B.1.1.298	Denmark	Y453F mutation in RBD	• Exhibited neutralization like parental type (D614G).	43
5. B.1.427 B.1.429	United states	L452R mutation in RBD	• Exhibited neutralization like D614G. • Seems to spread more easily. • L452R mutation enhanced attachment to ACE2.	43
6. P.2	Brazil (April 2020)	• Three spikes missense mutation • E484K • D614G • V1176F • ORF1a • L3468V, • L3930F • 5'UTR • R18C • Mutation in N- protein include: • A119S • R203K • G204R • M234I	• Potential reduction in neutralization by mAb treatments, convalescent, and postvaccination sera.	39,43
7. P.1	Primarily detected in the United States in January 2021 and was initially identified in travelers of Brazil in japan.	P.1 lineage contains three mutations in RBD of the spike protein including, i. K417T ii. E484K iii. N501Y	• Increased transmissibility and tendency for viral re-infection.	39,43,45
8. B.1.351	Initially detected in South Africa in December 2020 and was first identified in the United States at the end of January 2021.	This lineage emerged by substitution in spike protein like, RBD: • K417N • E484K	• Vaccine and convalescent serum have reduced cross neutralization of B.1.351 lineage. • Increased transmission.	38,43

(Continues)

TABLE 1 (Continued)

Variants	First discovery	Mutations identified	Consequence of mutations	References
		• N501Y Non RBD: • D614G • D215G • D80A • A701V • L18F		
9. B.1.525	First detected in United Kingdom/Nigeria in December 2020.	This lineage harbors following spike mutations, • 69del • A67V • 70del • 144del • D614G • E484k • F888L • Q677H	• Reduced neutralization by convalescent and postvaccination sera.	46
10. B.1.526	First detected in New York in November 2020.	Spike mutations include, • L5F • D253G • T951 • E484K • S477N • A701V • D614G	• Reduced neutralization by convalescent and postvaccination sera.	46,47
11. A.23.1	Uganda	Have 12–17 amino acid mutations (7 in spike protein).	• Data is scarce but presence of E484K can be associated with major concern of immune escape.	48
12. B.1.617	Most prevalent and common variant in India emerged in late 2020.	It has two prominent mutations in the critical receptor binding domain i.e., E484Q and L452R.	• Increased transmission possibly due to enhanced binding efficiency between viral • spike proteins and human Angiotensin Converting Enzmye-2 (hACE2). • Reduced sensitivity to vaccine (BNTI62b2 mRNA) elicited antibodies. • Significant reduction in neutralization by postvaccination sera.	49–51
13. B.1.617.1	India in December 2020.	Spike mutations include, • G142D • T951 • L452R	• Significant reduction in neutralization by postvaccination sera and EUA monoclonal antibody treatments.	51

TABLE 1 (Continued)

Variants	First discovery	Mutations identified	Consequence of mutations	References
		• E154K • D614G • P681R • E484Q • Q1071H		
14. B.1.617.2	India in December 2020.	Spike mutations include, • G142D • T19R • 156del • 157del • L452R • R158G • DG14G • D950N • P681R • T478K	• Significant reduction in neutralization by post vaccination sera and EUA monoclonal antibody treatments.	48
15. B.1.617.3	India in December 2020.	Spike mutations include, • G142D • E484Q • D614G • T19R • L452R • D950N • P681R	• Significant reduction in neutralization by post vaccination sera and EUA monoclonal antibody treatments.	48,51

Abbreviations: mAb, monoclonal antibodies; mRNA, messenger RNA; ORF, open reading frame; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

three amino acids in Orf1ab and mutations in the RBD (E484K and N501Y). Early reports demonstrated that despite RBD mutation (N501Y) in the B.1.1.7, it does not escape postvaccine neutralization (Moderna, mRNA-1273, and NVX-CoV2373, Novavax) and serum samples from convalescent individuals, though moderately at a reduced level.[40,41] Moreover, further variation in the B.1.351 variant may lead to escaping neutralization (Table 1).[42]

Rees-Spear et al.[52] evaluated the significant impact of the individual amino acid mutation on SARS-CoV-2 neutralization by creating a pseudotype of the virus using the spike sequence of B.1.1.7 variant. Their results revealed that repeated alterations in the RBD of spike can result in escaping neutralization by some of the monoclonal antibodies (mAbs). However, these mutations are not enough to cancel the effect of serum responses that are more resistant to these mutations especially after severe infection but not after a mild illness. Neutralization efficiency with mAbs specific to spike-proteins reduced drastically following successive mutation. However, in contrast, polyclonal antibodies obtained from early infected individuals are still active against a range of spike mutated pseudo-types but with reduced potency in few samples.

10 | EFFECT OF VIRUS VARIANTS ON DIAGNOSTIC CAPACITY

The observed mutations in the novel coronavirus have not been reported to affect the efficacy of the presently developed vaccine.[53] However, a mutation in the viral protein sequences and nucleic acid has placed currently in vitro diagnostic tests at risk if the mutations occur at the site critical for binding of primer or antibody in the RT-PCR and other immunoassays. It is especially of concern if antibody-based SARS-CoV-2 diagnostic assays are used to test the presence and concentration of viral proteins in oropharyngeal, nasopharyngeal or saliva fluids of infected individuals. The most common immunoassays used for the diagnosis of novel corona viral proteins are enzyme-linked immuno-sorbent assay and

lateral flow assays which target mostly the immunogenic viral proteins like nucleocapsid proteins (N) and spike-proteins (S). S-proteins are highly immunogenic having unique sequence to novel coronavirus[54] thus reducing the risk of cross-reactivity with other coronaviruses like MERS, SARS, and human coronaviruses such as OC43, 229E, NL639, and HKU-1. Although, targeting S-protein in immuno diagnostic assays is significant in minimizing the risk of cross-reaction and false-positive results, it is not without risk as mutations are more likely to occur in S-protein which could affect the validity of diagnostic assays along with the functioning of virus in a number of ways like increased transmission and infection rate.[55] The efficacy of diagnostic assays which mainly rely upon SARS-CoV-2 S-protein is highly vulnerable, as mutation at this site escapes successful detection that leada to an increased rate of false-negative results. In contrast, point mutations are not more likely to occur in the N-protein of the virus and are less likely to affect its function. Thus, diagnostic tests targeting N- protein of the virus are highly efficient than those targeting S-protein due to its conserved sequence (limited mutations in N-protein) and strong immunogenicity.[56] Although, the N-protein is less likely to mutate but not rigidly invulnerable to mutations hence, in vitro diagnosis and vaccine development must consider the potential N-protein mutations. Moreover, diagnostic assays that rely upon polyclonal antibodies have a significant advantage over tests that assess the single epitope by using mAb as polyclonal antibodies are more likely to report accurate results despite of mutation in any epitope by recognizing multiple analytes simultaneously.[57] None of the novel SARS-COV-2 variants including 501Y in South Africa, D796H, H69/V70, and D614G represented the escape variant while detecting with polyclonal antibodies directed against N-protein.[58] Even the recent strain B.1.1.7 that has 17 mutations could be detected by using these antibodies and does not seem to impact drastically on the Berlin–Charité protocol (98% sequence can be detected with present primers and probe) but may challenge the commercially available kits directed against spike-proteins.[59] Recently, Vogels et al.,[60] studied how the frequency of variation affects the efficiency of qRT-PCR assay and indicated GGG → AAC mutation at position 28881–28883 along viral genome that overlaps the CCDC-N forward primer. Similarly, another study revealed the transition mutation (C → T) positioned at 26340 of the viral genome, which was found to impair the Cobas E-gene qRT-PCR assay.[61] Conclusively, all the available data claimed that consistent mutations and variation can eventually lead to the impairment of diagnostic assays.

11 | CONCLUSION

COVID-19 is the third life-threatening pandemic that has challenged not only global health but also psycho-social and economic health worldwide. A novel coronavirus in late 2019 emerged in China called SARS-CoV-2 has caused unusual episodes of pneumonia that quickly spread across the world. Rapid genome sequencing of SARS-CoV-2 during the current pandemic revealed antigenic drift in the viral genome due to presence of several mutations, especially in the viral spike-protein. This antigenic drift resulted in better survival of the virus because of natural selection, as neutralizing antibodies raised upon either natural infection or vaccinations act against surface proteins particularly against Spike proteins and alteration in this protein might lead to escape variants. Therefore, prompt identification of the developing mutations over time is needed for monitoring the accurate treatment processes, vaccination, and well-validated diagnostic assays.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

DATA AVAILABILITY STATEMENT

All data presented in the article are included in the manuscript.

ORCID

Sadia Alam http://orcid.org/0000-0003-3102-8252

REFERENCES

1. Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol*. 2019;17(3):181-192. https://doi.org/10.1038/s41579-018-0118-9
2. Hui DS, Azhar EI, Madani TA, et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—The latest 2019 novel coronavirus outbreak in Wuhan, China. *Int J Infect Dis*. 2020;91:264-266. https://doi.org/10.1016/j.ijid.2020.01.009
3. Deng SQ, Peng HJ. Characteristics of and public health responses to the coronavirus disease 2019 outbreak in China. *J Clin Med*. 2020;9(2):575. https://doi.org/10.3390/jcm9020575
4. Wang C, Wang Z, Wang G, Lau JY, Zhang K, Li W. COVID-19 in early 2021: current status and looking forward. *Signal Transduct Target Ther*. 2021;6:114. https://doi.org/10.1038/s41392-021-00527-1
5. Xu X, Han M, Li T, et al. Effective treatment of severe COVID-19 patients with tocilizumab. *Proc Natl Acad Sci*. 2020;117(20):10970-10975. https://doi.org/10.1073/pnas.2005615117
6. Forni G, Mantovani A. COVID-19 vaccines: where we stand and challenges ahead. *Cell Death & Differ*. 2021;28:626-639. https://doi.org/10.1038/s41418-020-00720-9
7. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. 2020;382:727-733. https://doi.org/10.1056/nejmoa2001017
8. Gralinski LE, Menachery VD. Return of the coronavirus: 2019-nCoV. *Viruses*. 2020;12(2):135. https://doi.org/10.3390/v12020135
9. Chan JF, Kok KH, Zhu Z, et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infec*. 2020;9(1):221-236. https://doi.org/10.1080/22221751.2020.1719902
10. Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579:270-273. https://doi.org/10.1038/s41586-020-2012-7
11. Wu F, Zhao S, Yu B, et al. Complete genome characterization of a novel coronavirus associated with severe human respiratory disease in Wuhan. *bioRxiv (Preprint)*. 2020;36:224-233. https://doi.org/10.1101/2020.01.24.919183
12. Rahimi A, Mirzazadeh A, Tavakolpour S. Genetics and genomics of SARS-CoV-2: a review of the literature with the special focus on genetic diversity and SARS-CoV-2 genome detection. *Genomics*. 2020;113(1):1221-1232. https://doi.org/10.1016/j.ygeno.2020.09.059
13. Li T, Liu D, Yang Y, et al. Phylogenetic supertree reveals detailed evolution of SARS-CoV-2. *Sci Rep*. 2020;10:22366. https://doi.org/10.1038/s41598-020-79484-8
14. Hu B, Guo H, Zhou P, Shi ZL. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol*. 2020;19:141-154. https://doi.org/10.1038/s41579-020-00459-7

15. Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The lancet*. 2020;395(10224):565-574. https://doi.org/10.1016/s0140-6736(20)30251-8
16. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. *Nat Med*. 2020;26:450-452. https://doi.org/10.1038/s41591-020-0820-9
17. Domingo E. Mechanisms of viral emergence. *Vet Res*. 2020;41(6):38. https://doi.org/10.1051/vetres/2010010
18. Jia Q, Bielefeldt-Ohmann H, Maison R, Masleša-Galić S, Bowen R, Horwitz MA. Analysis of the mutation dynamics of SARS-CoV-2 reveals the spread history and emergence of RBD mutant with lower ACE2 binding affinity. *bioRxiv (Preprint)*. 2020
19. Shafique L, Ihsan A, Liu Q. Evolutionary trajectory for the emergence of novel coronavirus SARS-CoV-2. *Pathogens*. 2020;9(3):240. https://doi.org/10.3390/pathogens9030240
20. Flores-Alanis A, Sandner-Miranda L, Delgado G, Cravioto A, Morales -Espinosa R. The receptor binding domain of SARS-CoV-2 spike protein is the result of an ancestral recombination between the bat-CoV RaTG13 and the pangolin-CoV MP789. *BMC Res Notes*. 2020;13(1):1-6. https://doi.org/10.1186/s13104-020-05242-8
21. Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *J Virol*. 2020;94(7).
22. Zhao Z, Li H, Wu X, et al. Moderate mutation rate in the SARS coronavirus genome and its implications. *BMC Evol Biol*. 2004;4(21). https://doi.org/10.1186/1471-2148-4-21
23. Wang C, Liu Z, Chen Z, et al. The establishment of reference sequence for SARS-CoV-2 and variation analysis. *J Med Virol*. 2020;92(6):667-674. https://doi.org/10.1002/jmv.25762
24. Laamarti M, Alouane T, Kartti S, et al. 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(11):e0240345. https://doi.org/10.1371/journal.pone.0240345
25. Wen F, Yu H, Guo J, Li Y, Luo K, Huang S. Identification of the hypervariable genomic hotspot for the novel coronavirus SARS-CoV-2. *J Infect*. 2020;80:671-693. https://doi.org/10.1016/j.jinf.2020.02.027
26. Koyama T, Platt D, Parida L. Variant analysis of SARS-CoV-2 genomes. *Bull World Health Organ*. 2020;98(7):495-504. https://www.who.int/bulletin/volumes/98/7/20-253591.pdf
27. Mishra A, Pandey AK, Gupta P, et al. Mutation landscape of SARS-CoV-2 reveals five mutually exclusive clusters of leading and trailing single nucleotide substitutions. *bioRxiv (Preprint)*. 2020
28. Addetia A, Xie H, Roychoudhury P, et al. Identification of multiple large deletions in ORF7a resulting in in-frame gene fusions in clinical SARS-CoV-2 isolates. *J Clin Virol*. 2020;129:104523. https://doi.org/10.1016/j.jcv.2020.104523
29. Pachetti M, Marini B, Benedetti F, et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J Transl Med*. 2020;18:1-9. https://doi.org/10.1186/s12967-020-02344-6
30. Yin C. Genotyping coronavirus SARS-CoV-2: methods and implications. *Genomics*. 2020;112(5):3588-3596. https://doi.org/10.1016/j.ygeno.2020.04.016
31. Hamed SM, Elkhatib WF, Khairalla AS, Noreddin AM. Global dynamics of SARS-CoV-2 clades and their relation to COVID-19 epidemiology. *Sci Rep*. 2021;11(1):1-8. https://doi.org/10.1038/s41598-021-87713-x
32. Lurie N, Saville M, Hatchett R, Halton J. Developing Covid-19 vaccines at pandemic speed. *New Eng J Med*. 2020;382(21):1969-1973. https://doi.org/10.1056/nejmp2005630
33. Weissman D, Alameh MG, de Silva T, et al. D614G spike mutation increases SARS CoV-2 susceptibility to neutralization. *Cell Host & Microbe*. 2021;29(1):23-31. https://doi.org/10.1016/j.chom.2020.11.012
34. Isabel S, Graña-Miraglia L, Gutierrez JM, et al. Evolutionary and structural analyses of SARS-CoV-2 D614G spike protein mutation now documented worldwide. *Sci Rep*. 2020;10(1):1-9.
35. Korber B, Fischer WM, Gnanakaran S, et al. Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. *BioRxiv*. 2020;182:812-827. https://doi.org/10.1016/j.cell.2020.06.043
36. Zhang L, Jackson CB, Mou H, et al. The D614G mutation in the SARS-CoV-2 spike protein reduces S1shedding and increases infectivity. *BioRxiv*. 2020;11:6013. https://doi.org/10.1038/s41467-020-19808-4
37. Burton DR, Poignard P, Stanfield RL, Wilson IA. Broadly neutralizing antibodies present new prospects to counter highly antigenically diverse viruses. *Science*. 2012;337(6091):183-186. https://doi.org/10.1126/science.1225416
38. Tegally H, Wilkinson E, Lessells RJ, et al. Sixteen novel lineages of SARS-CoV-2 in South Africa. *Nat Med*. 2021;27(3):440-446.
39. Voloch CM, da Silva Francisco R, de Almeida LGP, et al. Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil. *J Virol*. 2021;95(10):1-6.
40. Xie X, Liu Y, Liu J, et al. Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. *Nat Med*. 2021;27(4):620-621. https://doi.org/10.1038/s41591-021-01270-4
41. Shen X, Tang H, McDanal C, et al. SARS-CoV-2 variant B. 1.1. 7 is susceptible to neutralizing antibodies elicited by ancestral Spike vaccines. *Cell Host & Microbe*. 2021;29(4):529-539.e3. https://doi.org/10.1016/j.chom.2021.03.002
42. Wibmer CK, Ayres F, Hermanus T, et al. SARS-CoV-2 501Y. V2 escapes neutralization by South African COVID-19 donor plasma. *Nat Med*. 2021;27:1-4. https://doi.org/10.1038/s41591-021-01285-x
43. Garcia-Beltran WF, Lam EC, St. Denis K, et al. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell*. 2021;184(9):2372-2383. https://doi.org/10.1016/j.cell.2021.03.013
44. Davies NG, Abbott S, Barnard RC, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B. 1.1. 7 in England. *Science*. 2021;372(6538). https://doi.org/10.1126/science.abg3055
45. Maggi F, Novazzi F, Genoni A, et al. Imported SARS-CoV-2 Variant P. 1 in Traveler Returning from Brazil to Italy. *Emerg Infect Dis*. 2021;27(4):1249-1251. https://doi.org/10.3201/eid2704.210183
46. Jangra S, Ye C, Rathnasinghe R, et al. SARS-CoV-2 spike E484K mutation reduces antibody neutralization. *The Lancet Microbe*. 2021;2(7):283-284. https://doi.org/10.1016/S2666-5247(21)00068-9
47. Annavajhala MK, Mohri H, Wang P, et al. A novel and expanding SARS-CoV-2 variant, B. 1.526, identified in New York. *medRxiv*. 2021:1-27. https://doi.org/10.1101/2021.02.23.21252259
48. Boehm E, Kronig I, Neher RA, et al. Novel SARS-CoV-2 variants: the pandemics within the pandemic. *Clin Microbiol Infect*. 2021;27:1-24. https://doi.org/10.1016/j.cmi.2021.05.022
49. Kumar V, Singh J, Hasnain SE, Sundar D. Possible link between higher transmissibility of B.1.617 and B.1.1.7 variants of SARS-CoV-2 and increased structural stability of its spike protein and hACE2 affinity. *bioRxiv*. 2021:1-12. https://doi.org/10.1101/2021.04.29.441933
50. Yadav PD, Sapkal GN, Abraham P, et al. Neutralization of variant under investigation B. 1.617 with sera of BBV152 vaccinees. *bioRxiv*. 2021:1-11. https://doi.org/10.1101/2021.04.23.441101
51. Greaney AJ, Loes AN, Crawford KHD, et al. Comprehensive mapping of mutations to the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human serum antibodies. *bioRxiv*. 2021;4:2020-12. https://doi.org/10.1101/2020.12.31.425021
52. Rees-Spear C, Muir L, Griffith SA, et al. The effect of spike mutations on SARS-CoV-2 neutralization. *Cell Rep*. 2021;34(12):108890. https://doi.org/10.1016/j.celrep.2021.108890
53. Dearlove B, Lewitus E, Bai H, et al. A SARS-CoV-2 vaccine candidate would likely match all currently circulating variants. *Proc Natl*

Acad Sci. 2020;117(38):23652-23662. https://doi.org/10.1073/pnas.2008281117

54. Jaimes JA, André NM, Chappie JS, Millet JK, Whittaker GR. Phylogenetic analysis and structural modeling of SARS-CoV-2 spike protein reveals an evolutionary distinct and proteolytically sensitive activation loop. *J Mol Biol*. 2020;432(10):3309-3325. https://doi.org/10.1016/j.jmb.2020.04.009
55. Chen J, Wang R, Wang M, Wei GW. Mutations strengthened SARS-CoV-2 infectivity. *J Mol Biol*. 2020;432(19):5212-5226. https://doi.org/10.1016/j.jmb.2020.07.009
56. Dutta NK, Mazumdar K, Gordy JT. The nucleocapsid protein of SARS–CoV-2: a target for vaccine development. *J Virol*. 2020;94(13-e00647-20. https://doi.org/10.1128/JVI.00647-20
57. Ascoli CA, Aggeler B. Overlooked benefits of using polyclonal antibodies. *Biotechniques*. 2018;65(3):127-136. https://doi.org/10.2144/btn-2018-0065
58. Ascoli CA. Could mutations of SARS-CoV-2 suppress diagnostic detection? *Nat Biotechnol*. 2021;39(3):274-275.
59. Ramírez JD, Muñoz M, Patiño LH, Ballesteros N, Paniz-Mondolfi A. Will the emergent SARS-CoV2 B.1.1.7 lineage affect molecular diagnosis of COVID-19? *J Med Virol*. 2021;93:2566-2568. https://doi.org/10.1002/jmv.26823
60. Vogels C, Brito AF, Wyllie AL, et al. Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT–qPCR primer–probe sets. *Nat Microbiol*. 2020;5:1299-1305. https://doi.org/10.1038/s41564-020-0761-6
61. Artesi M, Bontems S, Göbbels P, et al. A recurrent mutation at position 26340 of SARS-CoV-2 is associated with failure of the E gene quantitative reverse transcription-PCR utilized in a commercial dual-target diagnostic assay. *J Clin Microbiol*. 2020;58(10-e01598-20. https://doi.org/10.1128/JCM.01598-20

How to cite this article: Bano I, Sharif M, Alam S. Genetic drift in the genome of SARS COV-2 and its global health concern. *J Med Virol*. 2022;94:88-98. https://doi.org/10.1002/jmv.27337