Journal of Translational Science

Review Article

ISSN: 2059-268X

Implications of the novel mutations in the SARS-CoV-2 genome for transmission, disease severity and the vaccine development

Hikmet Akkız*

Department of Gastroenterology and Hepatology, The University of Çukurova, Adana, Turkey

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of the coronavirus disease 2019 (COVID-19), is steadily mutating during its continuous transmission among humans. During the COVID-19 pandemic, SARS-CoV-2 has accumulated mutations throughout viral genes encoding the ORF1a, ORF1b, ORF3, ORF8, nucleocapsid (N) and Spike (S) proteins. Mutations in the S protein are especially crucial because the S protein is key for the first step of viral transmission. A novel isolate of SARS-CoV-2 virus carrying the spike protein amino acid change D614G recently emerged and rapidly became dominant around the world. The consistent increase of the G614 variant at regional levels shows that the variant may have a fitness advantage. The spike protein D614G mutation increases SARS-CoV-2 infection in multiple human cell types. In addition, clinical samples obtained from COVID-19 patients carrying G614 variant have a higher levels of viral RNA. Although clinical and in vitro data suggest that the spike protein D614G mutation changes the virus phenotypically, the impact of the mutation on the rate of transmission between people, disease severity and the vaccine and therapeutic development is controversial. The review discusses the effects of novel mutations in the SARS-CoV-2 genome on the transmission, the clinical outcomes and vaccine and drug development.

Abbreviations: SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; COVID-19: Coronavirus disease 2019; S protein: Spike protein; N protein: Nucleocapsid protein; M protein: Membrane protein; SARS-CoV: Severe acute respiratory syndrome coronavirus; MERS-CoV: Middle East respiratory syndrome; CFR: Case fatality rate; WHO: The world health organization; ACE2: Angiotensin converting enzyme 2; TMPRSS2: Transmembrane protease serin 2; NTD: N-terminal domain; RBD: Receptor binding domain; GISAID: Global ınitiative for sharing all ınfluenza data.

Introduction

On december 31, 2020, China notified the World Health Organization (WHO) about a cluster of pneumonia cases of unknown etiology in Wuhan, Hubai Province [1,2]. The aetiological agent has been characterized as a SARS-Like betacoronavirus, named SARS-CoV-2 [2]. SARS-CoV-2 is the seventh coronavirus known to infect human. Severe Acute Respiratory Syndrome coronavirus (SARS-CoV), Middle East Respiraratory Syndrome (MERS-CoV) and SARS-CoV-2 can cause severe disease, whereas HKU1, NL63, OC43 and 229E are associated with mild symptoms [3-5]. Three major pathogenic zoonotic disease outbreaks by betacoronaviruses have been seen during the past two decades [3,5]. SARS-CoV caused a global pandemic in 2003 with an approximately 10% case fatality rate (CFR). SARS-CoV has not circulated in humans since 2004. MERS-CoV was first reported from Saudi Arabia in 2012 and has continued to infect humans with limited human-to-human transmission leading to a CFR of approximately 34.4% [3,5]. SARS-CoV-2 is characterized by its rapid spread and virulent human – to – human transmission [6]. On March 11, 2020, The World Health Organization (WHO) has declared COVID-19 to be a pandemic. Cellular entry of coronaviruses depends on the binding of spike protein (S) to a specific cellular receptor, angiotensin converting enzyme 2 (ACE2) receptor, and followed by cleavage with transmembrane protease serine 2 (TMPRSS2), both of which are abundantly expressed in not only the airways, lungs and nasal/oral mucosa and also the intestine. The binding affinity of the S protein and ACE2 has been found to be major determinant of SARS-CoV-2 replication and disease severity [6,7].

Some viruses have remarkable capacity to adapt to new hosts and environments which is highly dependent on their ability to generate de novo diversity in a short period [8]. The mutation rate of an organism is defined as the probability that change in genetic information is passed to the next generation. Rates of spontaneous mutations vary greatly among viruses [8,9]. RNA viruses mutate faster than DNA viruses, single – stranded viruses mutate faster than double-strand viruses and genome size appears to correlate negatively with the mutation rate [8-10]. Mutations originate from replication errors, nucleic acid damage and editing of genetic materials by host encoded proteins or by specialized molecular systems such as diversity - generating retro – elements (DGRs) [8]. If these changes are not corrected, they will be passed to the viral progency and hence will contribute to elevating the viral mutation rate [8].

Viruses mutate in the course of an epidemic as a result of natural selection, random genetic drift or features of recent epidemiology [11].

Correspondence to: Hikmet Akkız, Department of Gastroenterology and Hepatology, The University of Çukurova, Adana, Turkey, E-mail: hakkiz@superonline.com

Key words: *SARS-CoV-2, COVID-19, mutation, vaccines*

Received: September 05, 2020; **Accepted:** September 15, 2020; **Published:** September 17, 2020

Because these factors can work together it can be difficult to differentiate when the virus mutation becomes common through fitness or by chance [11]. These mutations may result in enhanced fitness or better efficiencies in evading the host immune system. Antigenic drift results in gradual accumulation of mutations in the viral genome over time which likely will not alter the virus drastically, and it will be recognizable to antibodies [11,12]. However, crosstalk between resistance mutations and the fitness landscape enables antibody resistance to develop across populations. To give an example, longer flu seasons have been shown to be associated with increased selection mutations in the influenza virus genome. The persistence of the COVID-19 pandemic may enable accumulation of immunologically significant mutations in the population even as vaccines are developed [11,12]. Antigenic drift is seen among the common coronaviruses OC43 and 229E and in SARS-CoV-1 (12). A single SARS-CoV amino acid change, spike D480A/G in RBD emerged in patients and became the dominant variant among 2003/2004 viruses [12]. D480A/G variant escapes neutralizing antibody and immune pressure. Although, there is no evidence yet of antigenic drift for SARS-CoV-2 with extended human-to-human transmission, SARS-CoV-2 could also acquire mutations with fitness advantages and immunological resistance [12].

The SARS-CoV-2 RNA – dependent – RNA polymerase (RdRp) is a key component of the replication / transcription machinery [9]. RdRps are multi-domain proteins able to catalyze RNA-template dependent formation of phosphodiester bonds between ribonucleotides in the presence of divalent metal ion [13]. SARS-CoV-2 shares a high homology for RdRp compored to SARS-CoV, suggesting that its function and mechanism of action might be well preserved. This has been confirmed by a recent cryo-EM structural study obtained for SARS-CoV-2 RdRp. In most viruses, RNA polymerase lacks proofreading capability [13]. Although coronaviruses are RNA viruses, the viruses have genetic proofreading mechanisms, therefore SARS-CoV-2 sequence diversity is very low [12]. However, natural selection can act upon rare but favorable mutation [12]. Studies show that coronaviruses have lower sequence diversity and mutation rates in comparison to more common viruses such as influenza and HIV [11,12]. Interestingly, despite these charactersitics, SARS-CoV-2 has already accumulated significant genetic diversity since the begining of the pandemic.

Recently, researchers tracking the spread of COVID-19 around the world have discovered a novel SARS-CoV-2 variant carrying the spike protein amino acid change D614G and demonsrated that G614 variant has become the most prevalant strain in the global pandemic. G614 variant has also been found to be associated with greater infectivity and higher viral load. Although, clinical and in vitro data showed that SARS-CoV-2 spike protein D614G mutation increases infectivity, the effects of the mutation on the rate of transmission between people, disease severity, vaccine and drug development are controversial.

Genomic landscape of SARS-CoV-2

To understand the clinical implication of SARS-CoV-2 mutations and to develop vaccines and neutralizing antibodies against the virus, we need to know the genomic landscape and biological behavior of key proteins of SARS-CoV-2. Coronaviruses belong to the coronaviridea family [1,2]. Similar to SARS-CoV and MERS-CoV, SARS-CoV-2 is an envelope, single-stranded and positive-sense RNA virus which genome encodes four major structural proteins ; spike (S), envelope (E), membrane (M), and nucleocapcid (N), about 16 nonstructural proteins (nsp 1-16), and five to eight accessory proteins [3-5]. Among them, the S protein plays a pivotal role in viral attachment, fusion, entry and transmission [3-5]. The coronaviruses S protein is divided into two domains: The S1 domain mediates receptor binding and the S2 mediates downstream membrane fusion. S protein includes an N-terminal S1 subunit responsible for virus receptor binding and a C terminal S2 subunit responsible for virus cell fusion. S1 is further divided into an N-terminal domain (NTD) and receptor binding domain (RBD). SARS-CoV-2 binds ACE2 receptor [3-5,14].

Two notable genomic features of SARS-CoV-2 have been reported:

a) On the basis of structural studies and biochemical experiments, SARS-CoV-2 appears to be optimized for binding to ACE2 receptor

b) The spike protein of SARS-CoV-2 has a functional polybasic (furin) cleavage site at the S1 – S2 boundary through the insertion of 12 nucleotides [3].

RBD in the spike protein is the most variable part of the coronavirus genome. Six RBD amino acids have been shown to be critical for binding to ACE2 receptors and for determining the host range of SARS-CoV-like viruses. They are Y442, L472, N479, D480, T487, and Y4911. Five of six residues differ between SARS-CoV-2 and SARS-CoV. On the basis of structural studies and biochemical experiments, SARS-CoV-2 seems to have an RBD that binds with high affinity to ACE2 [3,5,14].

Recurrent mutations found in the genome of SARS-CoV-2

RNA viruses such as SARS-CoV-2, HIV and influenza accumulate mutations rapidly because enzymes that copy RNA are prone to making errors [15]. The mutation and evolution rate of RNA viruses is dramatically high, up to a million times higher than that of their hosts [13]. A high mutation rate is associated with virulance modulation and evolvability, factors that affect viral adaptation creating a balance between the integrity of genetic information and genome variability [13,14]. However, sequencing data suggest that coronaviruses change more slowly than most other RNA viruses. A slow change rate of coronaviruses might be due to a proofreading enzyme that corrects potentially fatal copying mistakes [12,15]. A typical SARS-CoV-2 virus accumulates only two single-letter mutations per month in its genome [15]. Mutations – most of them single amino acid changes – provide researchers to tract the spread and to estimate when SARS-CoV-2 started infecting humans [15]. New mutations will be observed as the virus spreads in humans. However, many mutations will have no affects on the viral spread or cause disease, because they don't alter the shape of protein [12,15]. The accumulation of mutations can be marker of viral fitness.

13 variation sites in SARS-CoV-2 genome were recently characterized including SARS-CoV-2 QRF1ab, ORF 3a, ORF8 and N regions among which positions 28144 in ORF8 and 878 in ORF1a showed a mutation rate of 30.53% and 29.47%, respectively [16]. Previously reported data demonsrate that SARS-CoV-2 is rapidly spreading across countries and genomes and new mutation hotspots are emerging [13]. Biological characterization of viral mutation can provide crucial knowledge on assessing viral drug resistance and immune escape. Additionally, viral mutation studies can be pivotal for designing new vaccines, antiviral drugs and diagnostic assays. Virus mutagenic capability depends upon several factors, including the fidelity of viral enzymes that replicate nucleic acids as SARS-CoV-2 RNA dependent RNA polymerase (RdRp) [13,14]. The mutation rate drives viral evoluation and genome variability, theraby enabling viruses to escape host immunity and to develop drug resistance [16].

The vast majority of mutations detected so far in SARS-CoV-2 circulating in humans are likely neutral or deleterious. Homoplasies, recurrent mutations, can arise by product of neutral evoluation or as a result ongoing selection. A study published in early May identified 198 recurrent mutations in the genome of SARS-CoV-2 nearly 80% of the recurrent mutations representing non-synonymous changes at the protein level [1]. The detected recurrent mutations may indicate ongoing adaptation of SARS-CoV-2 to its novel human host [1]. Most of these mutations have been detected in non-structural proteins Nsp6 (coronavirus replicase), Nsp11 (coronavirus guanin -N7 methyltransferase) and Nsp13 (zinc-binding domain) and in spike protein [1]. In SARS-CoV-2, all nonstructural protein are encoded by the ORF1ab polyprotein which constitutes about two-thirds of its genome and undergoes an auproteolytic process to form all 16 known nonstructurel proteins [1]. Mutations in this region are consistent with previous studies with SARS-CoV-1 and MERS-CoV, where homoplasies have been detected in proteins Nsp9 (SARS), Nsp13 (SARS), and Nsp6 (MERS) [1]. Mutations in the Nsp6 protein detected both in SARS-CoV-2 and MERS-CoV, suggest this adaptation may be beneficial. In fact, Nsp6 is likely involved in autophagy restriction in coronaviruses and mutation may favor infection by evading the delivery of viral components to lysosomes for degradation [17].

220 genomic sequences were recently analysed from the GISAID database obtained from COVID-19 patients worldwide from December to mid-March 2020 [13]. Researchers have characterized 8 novel recurrent mutations of SARS-CoV-2, located at positions 1397, 2891, 14408, 17746,17857,18060, 23403, and 28881 [13]. The study showed that mutations in 2891,3036,14408,23403 and 28881 positions are predominantly observed in Europe, but those located at positions 17746, 17857 and 18060 are exclusively present in North America [3]. The investigators have identified for the first time a silent mutation in RdRp gene in England on February 9th, 2020, while different mutation in RdRp gene changing its amino acid composition emerged on February 20th, 2020 in Italy [13]. Viruses with RdRp mutation have median of 3 point mutations. They concluded that the findings suggest that the virus is evolving and European, North American and Asian strains might coexist, each of them characterized by a different mutation pattern [13]. The contribution of RdRp mutations to this phenomenon should be investigated.

Currently, researchers in the US and UK have queried the Global Initiative for Sharing All Influenza Data (GISAID) SARS-CoV-2 squence database looking for changes in the spike protein genetic sequence of more than 0.3% from the Wuhan reference [18]. Korber *et al.* started scouring thousands of coronavirus genetic sequences to detect mutations that may have changed the virus properties that will expedite the transmission of the virus [12,15]. The researchers detected a mutation at the 614th amino acid position of the spike protein that is caused by an aspartic acid (A) -to-glisin (G) nucleotid mutation at position 23.403 in the Wuhan reference strain [12]. They showed that D614G mutation emerged early during pandemic and viruses with a spike protein G614 variant spread rapidly worlwide and over the course of 1 month the variant became the globaly dominant form of SARS-CoV-2 [12]. D614G mutation was first detected in viruses collected in China and Germany in late January. The viruses with D614G mutation were detected in Europe in the early phase and have rapidly spread worldwide, especially to European and North American countiries [12]. The G614 variant has been shown to be almost always associated with three additional mutations in other parts of the SARS-CoV-2 genome: a C to T mutation in the 5' UTR (position 241 relative to the Wuhan reference sequence), a silent C-to-T mutation at position 3.037, and a C-to-T mutation at position 14.408 that results in an amino acid change in RdRp [12]. The most viruses with G614 variant may share a common ancestor. More recently, Zhang *et al.* showed a similar data ; The G614 genotype was not detected in February (among 33 sequences) and observed at a low frequency in March (26%), but rapidly increased by April (65%) and May (70%), indicating a transmission advantage over viruses with D614 [19].

Deletions within the viral genome are a natural phenomenon, almost inevitably related to the attenuation of the virus, however it can sometimes be associated with more severe infection. SARS-CoV-2 variants with a 382 – nucleotide deletion (D382) in the open reading frame 8 (ORF8) region of the genome have been detected in many countries. Su *et al.* reported a 382 – nucleotide (nt) deletion in SARS-CoV-2 that truncates open reading frame 7b (ORF7b) and ORF8, removing the ORF8 transcription regulatory sequence (TRS) and elimineting ORF transcription. Mutations and deletions in ORF8 of SARS-CoV-2 genome have been associated with reduced replicative fitness and virus attenuation [20]. In contrast, the SARS-CoV-2 382 nt deletion viruses showed significantly higher replicative fitness in vitro than the original virus, while no difference was observed in patient viral load, indicating that the deletion variant viruses retained their replicative fitness [20]. A robust antibody response to ORF8 has been observed in SARS-CoV-2 infection, suggesting that emergence of ORF8 deletions may be due to immune – driven selection and further deletion variants emerge during the sustained transmission of SARS-CoV-2 in humans. In another study published in Lancet, Young *et al.* investigated the effect of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response [21]. In the study, 278 patients with PCR-confirmed SARS-CoV-2 infection were screened for the D382 deletion variant and 131 were enrolled onto the study, of whom 92 (70%) were infected with original virus, ten (8%) had mixid type viruses and 29 (22%) had only D382 variant. Islam *et al.* have detected twelve deletion sites throughout the genome other than previously reported deletions at the coding sequence of the ORF8, spike, ORF7a proteins [22].

Can novel SARS-CoV-2 G614 variant be more infectious than the original D614 genotype?

Using phylogenomic data several groups have proposed that the G614 variant can confer increased transmissibility leading to positive selection. Examination of viral strains from Europe, North America, Australia, and Asia demonsrated that the G614 variant increased in frequency over a several month period [18]. The transition from the D614 to the G614 variant may have initially occured in China and spread quickly became dominant and continued spreading across North America, Oceania, and finally Asia. The study published by Korber *et al.* provided data that the G614 variant spread more rapidly than the D614 variant and became the globally dominant form of SARS-CoV-2 within a month [12]. The mutation that causes the D614G amino acid change is transmitted as a part of a conserved haplotype defined by 3 mutations that almost always track together [11,12]. They also showed that clinical samples obtained from patients infected with the G614 variant had a higher viral RNA levels compared to the original D614 variant. In addition, to examine whether D614G made the virus more transmissible, the researchers investigated its effect in vitro by using a genetically modified form of HIV that used the SARS-CoV-2 spike protein (pseudovirus) to infect cells [12,15]. And showed that the G614 variant displayed higher infectivity than D614 variant [12]. Taking into

account these data, they conclude that SARS-CoV-2 carrying spike protein G614 variant is more infectious than original D614 variant [12]. This rapid increase in frequency suggests improved infectivity in comparison to the Wuhan original variant.

The several other groups investigating the effect of the G614 variant on the transmission capacity of the virus by used different pseudovirus systems, also demansrate that viruses carrying spike protein D614G mutation infected cells much more than D614 variant – up to ten times more efficiently in some cases [15]. An analysis by Zhang *et al.* of the S protein sequences available from the GenBank showed that the G614 genotype has not been detected in February (among 33 sequences) and observed at low frequency in March (26%), but increased rapidly by April (65%) and May (70%), indicating a transmission advantage over viruses with D614 [16]. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity [16]. The authors compared the functional properties of S proteins with aspartic acid (S^{D614}) and glycine (S^{G614}) at residue 614 and detected that retroviruses pseudotyped with S^{G614} infected ACE2 expressing cells markedly more efficiently than those with S^{D614} [16]. This greater infectivity has been found to be correlated with less S1 shedding and greater incorporation of S protein into the pseudovirion [16]. The researchers explained that the dominant G614 variant could effectively infect the four cell lines tested and could be 10-fold more infectious than the original D614 strain [16]. An effect that has been confirmed in vitro using HEK cells. Phenotypically, the D614G mutation is shown to be to have a greater number of the interaction between the S1 and S2 domains and limit S1 shedding, resulting in better overall infectivity [16].

In another study, Daniloski *et al.* have performed site-directed mutagenesis on a human codon-optimized spike protein to introduce the D614G variant and produce SARS-CoV-2 pseudotyped lentiviral particles with this variant and with D614 spike [23]. They show that in multiple cell lines including human lung epitelial cells, that the D614G mutation is up to 8-fold more effective at transducing cells than wild type [23]. The researchers have also demonsrated increased infection using both spike – pseudotyped lentivirus and intact SARS-CoV-2 virus [23]. Some studies showed that cleavage by the host protease furin at the spike S1/S2 site in SARS-CoV-2 is essential for cell-cell fusion and viral entry [24]. Daniloski *et al.* have performed in vitro digestion of both spike variants after pull-down to test for differences in furin-mediated cleavage and show that the G614 variant was more resistance to proteolitic cleavage in vitro and in human cells than the D614 variant, suggesting a possible mechanism for the increased transduction [23].

More recently, two relevant studies investigating the effect of G614 variant on transmission rate of the virus have been published from the United Kingdom (UK) (). The COVID-19 Genomics UK Consortium reported that G614 variant has an effect on the spread of SARS-CoV-2 in humans. The consortium has analysed genomes of around 25.000 viral samples. From these data researchers have identified more than 1300 instance in which a virus entered the United Kingdom and spread including examples of D and G type viruses [15,25]. In another study, Volz&Rambaut studied the UK spread of 62 COVID-19 clusters seed by D viruses and 245 by G viruses. The resarchers found that G viruses transmit slightly faster than original viruses carrying D614 variant and formed larger cluster of infections [15,26]. Their estimates of the difference in transmission rates hover around 20%. However, reserchers suggest that the true value could be a bit higher or lower [15,26].

But these studies come with many caveats and their relevance to human infections is unclear. Many researchers claim that there has not been solid proof that G614 variant has a significant effect on the spread of the virus or that a process of natural selection explains its rise [11,15]. Korber *et al.* have proposed that the rapid spread of SARS-CoV-2 with G614 variant was because it is more infectious than D614 variant [12]. They suggested that higher viral RNA in clinical samples obtained from G614 infections and higher RNA levels in pseudoviruses in vitro experiments may support this hypotesis. However, Grubaugh *et al.* suggest that although the data reported by current studies suggest increased viral transmissibility, they do not prove that G614 is more infectious or transmissible than viruses carrying D614 variant [11]. The researchers higlighted a significant limitation regarding the study published by Korber *et al.* that increased viral sheeding or RNA does not reflect viral transmission capacity [11]. Additionally, the researchers suggested that the increase in frequency of G614 variant could be explained by chance or epidemiology of the pandemic [11]. In February, most the area with the most COVID-19 cases shifted from China to Europe and then in March on to the United States [11]. The great majority of SARS-CoV- 2 lineages in the US arrived from Europe. They conclude that the current data show that G614 variant is less important for other risk factors, such as age or comorbidity [11].

Grubaugh *et al.* suggested that viral load and disease severity are not always correlated, particularly when viral RNA is used to estimate virus titre [11].

Another objection is whether the pseudovirus assays could demonstrate the ability of the virus to infect a cell in culture. Previous studies provide the data that D614G mutation is more infectious than D614 variant have been performed in a very artificial system [11,15]. Additionally, these assays don't account for the effect of other viral and host proteins and explain the biochemical host-pathogen crosstalk that must occur to support infection and transmission [11]. It is impossible to conclude that a single mutation alone would have a major impact in a large, diverse human population based on in vitro infectivity and fitness data [11]. Therefore, further studies both from an epidemiological and a laboratory perpective are needed to prove that the viruses carrying G614 variant is more replicative and more transmissible [11]. An analysis of more than 30 000 viral genomes in the United Kingdom showed that the SARS-CoV-2 spike G614 variant may increase transmission between people but researchers observed no diference in cell infectivity measured in laboratory [26].

Can SARS-CoV-2 D614 variant affect clinical outcomes?

Although several studies showed that the G614 variant causes rapid viral spread than D614 variant, it is still unclear whether the G614 variant will have an impact on the clinical outcomes. Clinical and in vitro data suggest that D614G mutation changes the virus phenotype and may have an impact on transmission and disease severity [1,12,19]. Phenotypically, the G614 variant has been shown to have a greater number of functional spikes on its surface in comparison to the original variant. Additionally, the mutation has been demonstrated to stabilize the interaction between the S1 and S2 domains and limit S1 shedding, resulting in increased overall infectivity [19].

However, so far there is no evidence that infection with SARS-CoV-2 with the G614 variant will cause more severe disease than original Wuhan variant. Examination of global clinical data from 999 patients hospitalized with COVID-19 in the United Kingdom, showed that the cycle threshold was lower, indicating a higher viral load with the G614 variant than the original virus carrying D614 variant [12,18]. Although the researchers suggest that the virus with G614 is likely to be more infectious than original variant, they were not able to demonsrate an association between the G614 variant and disease

severity as measured by hospitalization outcome [12]. A comparison of G614 variant and hospitalization was not significant (p=0.66, Fischer exact test), although comparing ICU admission with IP and OP did have borderline significance (p=0.047). Regression analysis supported the result that G614 variant was not associated with greater levels of hospitalization but that higher age, male sex and lower viral loads were highly predictive of hospitalization [12]. Further analysis showed that viral load was not masking a potential D614G mutation effect on hospitalization. Univariate analysis also found higly significant associations between age and male gender and hospitalization [12]. Based on available data, Grubaugh *et al.* suggest that D614G is less important for COVID-19 than other risk factors, such as age and comorbidity [11].

These clinical observations are supported by two independent studies. Wagner *et al.* have used SARS-CoV-2 genomes from COVID-19 patients in Washington State to explore whether spike G614 variant has any impact on replication potential of the virus and clinical outcomes of the patients [27]. The researchers have not found convincing evidence for a difference in clinical outcomes among patients infected with SARS-CoV-2 carrying spike protein D614G mutation. Lorenzo – Redondo and coworkers reported the phylogenetic and phylodynamic analysis of 88 SARS-CoV-2 genomes from COVID-19 patients in Chicago and identified three distinct phylogenetic clades [28]. Clade 1 was most closely related to clades centered in New York and rapidly expanded across the USA, while clade 3 was most closely related to those in Washington. Clade 2 was localized primarly to to the Chicago area. Average viral loads in the airways of patients infected with clade 1 viruses have been sfound to be signifacantly higher than those clade 2 [28]. The reseachers showed that disease severity did not correlate significantly with clade or Ct value. Previous reports suggested that disease severity is correlated with viral load measured by Ct value during disease course, but their data suggest the same may not hold true for Ct values collected at diagnosis. In another study including more than 30 000 viral genomes, the impact of SARS-CoV-2 spike protein G614 variant on disease severity could not be shown [26]. Volz *et al.* also did not observe clinical differences COVID-19 patients with either variant.

The D382 variant causes clinically significant disease, but infection seems to be milder compared with disease by wild-type virus, with less pronounced cytokine release during the acute phase of infection. Young *et al*. observed that patients infected with the D382 variant had lower concentrations of proinflammatory cytokines, chemokines and growth factors that are strongly associated with severe COVID-19 [21].

The ımpact of the G614 variant on vaccine development

Developing a vaccine against SARS-CoV-2 is a high priority for preventing and mitigating future waves of the pandemic. Currently, eight companies are in the midst of phase 3 trials (large, prospective, placebo-controlled trials) to prove efficacy at least certain length of time and to prove safety at least in a certain number of people. Despite the significant diversity of vaccine formulations and delivery methods, many vaccine candidates focus on the spike sequences or peptides. The vaccine candidates have been developed prior to the emergence of the G614 variant. As mentioned before, the genetic sequence diversity of SARS-CoV-2 is low. However, because most vaccine and antibody therapies target the trimeric spike protein of SARS-CoV-2, any alterations in its genetic sequence could potentially reduce the efficacy of candidate vaccine or render the virus resistant to specific treatments [11,12,18, 29].

The crucial question is what is the implication of the G614 variant for vaccine development?

Using epitope prediction for common HLA alleles, Daniloski *et al*. demonsrate that the G614 variant can alter predicted MHC binding [23]. For example, the predicted binding for one high – affinity epitope decreased by nearly 4 – fold. Although full-lenght spike protein likely can produce many immunogenic peptides, several vaccines use only portions of spike and thus it may be better to adapt vaccine efforts to the D614G variant given its global spread. Although there is minimal difference in ACE2 receptor binding between the spike variants, they showed that the G614 variant is more resistant to proteolitic cleavage in vitro and human cells, suggesting a possible mechanism for the increased transduction [23]. This result has important implications for the efficacy of spike – based vaccines currently under development in protecting against this recent and higly – prevalent SARS-CoV-2 isolate [23].

Typically, surface proteins outside of the viral virion are selected for antigens so that antibodies generated from a vaccine – trained B - cell can bind to the virus for neutralization [29]. Previous studies showed that D614G mutation is located in the external spike protein of SARS-CoV-2 that has strong immunogenicity therefore the mutation may have an impact on the ability of the virus to evade vaccine – induced immunity. However, Grubaugh *et al.* think that G614 mutation is unlikely to affect vaccine studies [11]. G614 variant is not located in the RBD of the spike protein but in the interface between the individual spike promoters that stabilize its trimeric form on the virion surface through hydrogen bonding [11]. This can lead to the loss of between promoter hydrogen bonds, modulate interactions between spike promoters or change glycosylation pattern [12]. Any of these changes can alter infectivity but it is less likely that it would drastically alter the immunogenecity of RBD epitopes through to be important for antibody neutralization [11]. Several research groups demonstrated that the antibodies generated from natural infection with carrying D614 or G614 variant could cross – neutralize, suggesting that the locus is not critical for antibody-mediated immunity [12,30,31]. *Grubaugh et al.* proposed that the D614G mutation is therefore unlikely to have a major impact on the effcacy of vaccines currently in the pipeline, some of which exclusively target the RBD [11]. Most available evidence suggests that D614G mutation does not stop the neutralizing antibodies generated from SARS-CoV-2 [15].

But evidence is emerging that other mutations could help the virus to avoid some antibodies. Koyama *et al*. suggest that the coronavirus genome higly prone to mutations that lead to genetic drift and escape from immune recognition [29]. Thus, it is imperetive that substrains with different mutations are also accounted for during vaccine development. The dominant G614 variant may cause random genetic drift resulting in vaccine mismatches that offer little protection to that group of patients [29]. Innovative vaccine development methods including conserved internal epitops, recombinant proteins, spanning epitopes or pooling multiple vaccines, will be required to combat the antigenic drift [29]. Recently, Dearlove *et al.* investigate the diversity seen in SARS-CoV-2 sequence and compare it to the sequence on which most vaccine candidates are based. They concluded that a SARS-CoV-2 vaccine candidate would likely match all currently circulating variants [32]. It is a possibility that the virus will acquire mutations that change its susceptibility to antibodies and immunity.

The specific effect of G614 variant on spike protein function in entry and fusion is not clear. Therefore, the impact of G614 variant on

antiviral drug candidates is unknown. We don't have data that the G614 mutation would interfere with therapeutic strategies such as monoclonal antibodies designed to disrupt spike binding with ACE2 or drugs that modulates downstream processes such as endosomal acidification [11]. We need to understand better the role of G614 mutation during the natural course of SARS-CoV-2 infection.

Conclusion

Although SARS-CoV-2 is not highly mutable, studies showed that variants in the SARS-CoV-2 genome may arise rapidly and may have effects on the COVID-19 pandemic. Global tracking data clearly indicate that the SARS-CoV-2 viruses carrying the G614 variant in spike protein spread faster than original D614 variant. This can be explained by the fact that the virus is more infectious than D614 variant. Cell studies using forms of viruses with either the original D614 or G614 variant of the spike protein showed that viruses carrying spike protein G variant were significantly more infectious. Currently, G614 variant is the pandemic. As a result, its characteristics are relevant. Considering in vitro and clinical data, it is clear that G614 variant has a distinct phenotype.

Although the G614 variant is associated with a higher viral load, the variant does not affect clinical outcomes. G614 variant is not located in the RBD of spike protein but in interface between the individuals Spike promoters. Therefore, the G614 variant is thought not to affect vaccine development against COVID-19 infection.

Financial support

The review article is not supported by any instituties.

Conflicts of interest

The author declare no conflicts of interest that pertain to this work.

References

1. Van Dorp L, Acman M, Richard D (2020) Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infection, Genetics, Evoluation* 83: 104351.
2. Wu F, Zhao B, Yu B (2020) A new Coronavirus associated with human respiratory disease in China. *Nature* 579: 265. [Crossref]
3. Anderson KG, Rombout A, Lipkin WI (2020) The proximal origin of SARS-CoV-2. *Nature Medicine* 26: 450-455. [Crossref]
4. Yang P, Wang X (2020) COVID-19: a new challenge for human beings. Nature Cellular & Molecular *Immunology*.
5. Jiang S, Hillyer C, Du L (2020) Neutralizing antibodies against SARS-CoV-2 and other human coronaviruses. *Cell Trends in Immunology* 41: 355-359.
6. Sungnak W, Huang N, Becavin C (2020) SARS-CoV-2 entry factors are highly expressed in nasal epitelial cells together with innate immune genes. *Nature Medicine* 2020. [Crossref]
7. Lamers MM, Beumer J, van der Vaart J (2020) SARS-CoV-2 productively infects human gut enterocytes. Science May.
8. Sanjuan R & Domingo – Calap P (2016) Mechanisms of viral mutation. *Cell Mol Life Sci* 73: 4433-4448.
9. Denison MR, Graham RL, Donaldson EF (2020) Coronaviruses, An RNA proofreading machine regulates replication fidelity and diversity. *RNA Biology* 8: 270-279. [Crossref]
10. Khailany RA, Safdar M, Ozaslan M (2020) Genomic characterization of a novel SARS-CoV-2. *Gene Reports* 19.
11. Grubaugh ND, Hanage WP, Rasmunsen Al (2020) Making sense of Mutation: What D614G Means forthe cOVID-19 Pandemic Remains Unclear. *Cell* 182.
12. Korber B, Fischer WM, Gnanakaran S (2020) Tracking changes in SARS-CoV-2 spike: Evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 182: 1-16.
13. Pachetti M, Matini B, Benedetti F (2020) Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J Transl Med* 18: 179.
14. Amanat F, Krammer F (2020) SARS-CoV-2 Vaccinies: Status Report. *Cell Immunity* 52: 583-590. [Crossref]
15. Callaway E (2020) Making sense of coronavirus mutations. *Nature* 585: 174-177. [Crossref]
16. Wang C, Liu Z, Chen Z (2020) Establishment of reference sequence for SARS-CoV-2 and variation analysis. *J Med Virol*.
17. Carmona – Gutierrez D, Bauer MA, Zimmermann A (2020) Digesting the crisis: autophagy and coronaviruses. *Microbiol Cell* 7 :119-128. [Crossref]
18. Ampel NM (2020) A mutation in SARS-CoV-2 spike protein is associated with increased infectivity. *N Eng J Med*.
19. Zhang L, Jackson CB, Mou H (2020) The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity.
20. Su YCF, Anderson DE, Young BE (2020) Discovery and genomic characterization of a 382 – nucleotide deletion in ORF-7b and ORF8 during the early evaluation of SARS-CoV-2. *Am S Microbiol* 11: e 01610- e 01620.
21. Young BE, Fong S-W, Chan Y-H (2020) Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *Lancet*.
22. Islam MR, Hoque M, Rahman MS (2020) Genome – wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity. *Nature Research* 10: 14004-14019
23. Daniloski Z, Jordan TX, Ilmain Jk (2020) The Spike D614G mutation increases SARS-CoV-2 infection of multiple human cell types. *BioRvix*.
24. Hoffman M, Kleine-Weber H, Pöhlmann S (2020) A Multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Mol Cell* 78: 779-784e5. [Crossref]
25. 25 – COVID-19 Genomics Consortium.
26. Volz E. Clarke S, Young L (2020) SARS-CoV-2 G614 variant, possible implications for infectiousness. SMC.
27. Wagner C, Roychouchury P, Hadfield J (2020) Comparing viral load and clinical outcomes in Washington State across D614G mutation in spike protein of SARS-CoV-2.
28. Lorenzo-Redondo R, Nam HH, Roberts SC (2020) A unique clade of SARS-CoV-2 viruses is associated with lower viral loads in patient upper airways. *MedRxiv* 19: 20107144. [Crossref]
29. Koyama T, Weerraratne D, Snowdon JL (2020) Emergence of drift variants that may affect COVID-19 vaccine development and antibody treatment. *Pathogens* 9: 324. [Crossref]
30. Ozono S, Zhang Y, Ode H (2020) Naturally mutated spike proteins of SARS-CoV-2 variants show differential levels of cell entry. *BioRxiv* 15: 151779.
31. Hu J, H C-L, Gao Q-Z (2020) D614G mutation of SARS-CoV-2 spike protein enhances viral infectivity and decreased neutralization sensitivity to individual convalescent sera. *BioRxiv* 20: 161323.
32. Dearlove B, Lewitus E, Bai H (2020) A SARS-CoV-2 vaccine candidates would likely match all currently circulating variants. *Proc Natl Acad Sci U S A* 117: 23652-23662. [Crossref]