International Journal for Research in Applied Sciences and Biotechnology

ISSN: 2349-8889
Volume-9, Issue-1 (January 2022)

www.ijrasb.com

https://doi.org/10.31033/ijrasb.9.1.4

Article Review: Molecular Study for Mutation of N-gene and S-gene COVID-19 Virus

Zina F. H. AL-Obaidi[1] and Abeer Ali Marhoon[2]
[1]College of Biotechnology, AL-Nahrian University, Baghdad, IRAQ.
[2]AL-Rafidian University College, Baghdad. IRAQ.

[2]Corresponding Author: abeer.marhoon@ruc.edu.iq

ABSTRACT

Thousands of individuals are affected every day by the current covid-19 pandemic, which is caused by a novel coronavirus called SARSCoV2. As a result, medicines and vaccines that are effective against all SARSCoV2 subtypes are critical today. Viral genome mutations are prevalent, and they can affect the encoded proteins, resulting in varying levels of detection and illness treatment effectiveness. Despite its clinical relevance, the SARS-CoV-2 gene set remains uncertain, making COVID-19 biology difficult to understand. A single type of mutation in the S gene that was changed the anticodon 614 from aspartic acid to glycine (D614G) consequence in increased virus infection. Herein, we report the gene mutation of structural proteins particularly spike and nucleocapsid proteins in viral genome. The overall prevalence of S and N gene mutations in SARS-CoV-2 were investigated. Among the structural proteins, our findings suggest that nucleocapsid had the highest mutation density, whereas Spike D614G was the most prevalent 93.9 %, found largely in genomes worldwide. These findings indicate that while designing diagnostics tools and therapeutic alternatives, the virus genotype in a certain community should be taken into account.

Keywords- COVID-19 mutation, SARS-CoV, Spike glycoprotein, S-gene mutation, N-gene Mutation, COVID-19 receptor binding domain, COVID-19 antigens

I. INTRODUCTION

The first time COVID-19 infection was originated in China then subsequently spread to approximately every country across the world. As the more patients were identified, various states around the world implemented restrictions on travel and quarantines [1]. The medical, scientific, and global health sectors have had to respond to emerging viral gene mutations during the severe acute respiratory syndrome coronavirus 2 known as SARS-CoV-2 pandemic [2]. Coronaviruses are distinguished by the crown shape spikes across the viral genome surface, which give them their name. Coronaviruses are divided into four sub-groups: alpha, beta, gamma, and delta. The alpha-beta sub-groups are considered to have begun in mammals and bats, while the gamma-delta sub-groups attack mostly by birds. In addition, NL63 (Alpha-coronavirus), 229E (Alpha-coronavirus), HKU1 (Beta-coronavirus) and OC43 (Beta-coronavirus), are the four most prevalent human coronaviruses. Although HKU1&NL63 were just formerly found during the SARS pandemic, they are suspected to possess disseminating in the human being species for interval of time. The structures are encapsulated viruses' molecule that cause 15 to 30% of severe pneumonia symptoms, in population around the world. The civet or raccoon is the intermediate host of SARS-CoV, and the camel is the intermediate host of MERS-CoV,4 however, the host of the pandemic strain of SARS CoV-2 is unknown, but it is thought to be closely linked to Bat-Cov- RaTG13. Despite the RaTG13 bat virus's genome's resemblance to Malayan pangolins coronaviruses SARS-CoV-2, are the most comparable to SARS CoV-2. In general, SARS CoV-2 has six (same) major receptor binding domain segments [3-7]. As of April 2021, it was perhaps lethal, known as COVID-19, and infecting above than 150-billion people around the world. Other nations where this form of CoV is expanding include South Korea, Italy and Iran [8]. Patients with symptoms are ranges from moderate to severe, regardless of the fact that the great majority cases are asymptomatic. With moderate or no pneumonia, minor symptoms include headache, fever, fatigue and cough; acute respiratory disease, significant clinical symptoms, dyspnoea and hypoxia might occur. Respiratory failure, shock and multi-organ dysfunction, are common in catastrophic instances. Comorbidities in older people include high blood pressure, diabetes, cardiovascular disease, and obesity, and respiratory problems have greater disease mortality rates. With the increasing number of CoVs, researchers are focusing on finding what is causing the virus to spread so quickly and broadly. Factors include age, demographic characteristics, clinical patient safety, or other possibly undiscovered elements have a substantial impact on clinical outcomes which ranging from asymptotic to fatality according to epidemiological studies [9]. The key characteristic of the virus has been determined by genetic and structural analyses, which could improve the effectiveness and frequency of virus transmission among human cells [10]. The genomic sequence of this recently discovered virus is 79.6 percent identical to the severe acute respiratory syndrome coronavirus (SARS-CoV) [11].

Moreover, the SARS CoV-2 genome contains numerous genes that encode structural proteins including spike, nucleocapsid, envelope, and membrane protein [12]. The virus uses host machinery to create its lipid envelope (S proteins), providing it a crown shape. The S protein is a trimetric protein with two distinct domains that helps the virus connect to the host cell and cleave it. The three key domains of the S1 protein are the receptor binding domain (RBD), the C-terminal domain (CTD), that interacts to biological ACE2, as well as the N-terminal domains [12].

The globular region of the S1 subunit mediates CoV attachment to the receptor, whereas membrane fusion is caused by the RBD at the amino terminus of the S1-component and the carboxyl terminus domain of the S2- component. S2, also called the bio membrane anchored stalk domain, is responsible for CoV–human cell fusion [13]. With 79.5 percent genomic sequences homology, the S- protein RBD of both SARS CoV and SARS CoV-2 is substantially conserved. SARS CoV-2 structural sequencing was discovered in the S2 subunit CTD including all linked coronaviruses using cryo electron imaging. The S1 subunit RBD protein sequence, on the other hand, comprises unique sequences, which are typical in coronaviruses in general. Due of frequent contacts with the host immune system, this is most likely the product of significant evolutionary pressure [14]. Antigenic evolution of SARS-CoV-2 is still being detected, especially in the spike protein RBD. This has become the prevalent variant over time, subsequently replacing the original SARS-CoV-2 genotype [15]. The virus's cellular and protein receptors are both promising targets for therapeutic research against pathogens [16]. Several factors, including differing national quarantine policies and different races or genetic origins, have been confirmed to alter COVID-19 mortality and infection rates in different geographical regions. In particular with respect to a few more predictable and viral-mutations, host-genetic susceptibility, and genetic-variability are all unpredictable elements, have a significant impact on COVID-19 clinical characteristics [1].

Viral replication produces mutations as a natural by-product. The mutation rate of RNA viruses is higher than that of DNA viruses. Rotaviruses, on either side, produce fewer mutation than all the other RNA viruses due to the presence of an enzyme that corrects some replication errors [17]. Regardless of their consequences on viral fitness, the population will be dominated by modifications identified in the genome sequences of these viral progenitors. When studying SARS-CoV-2 pathogenesis, the terms mutation, variant, and strain are closely interrelated, although there are crucial differences to be recognized. The actual alteration in sequence is referred to be a mutation: D614G is an aspartic acid to glycine substitution at position 614 of the spike glycoprotein. It's worth noting that certain mutations are founder mutations, exhibiting distinct patterns of geography. These mutations could indeed definitely drive viral mutation rate, tend to result in a viral infection with a little more reduction in organic disease - causing fitness due to genotype-phenotype relationship, as well as, across the other side, enabling sudden antigenic realignment to exit host immune response and trying to invent a drug - resistant virus, thereby transforming to a more infective or lethal virus. Despite the fact that SARS CoV-2 has proofread genomic systems for repairing RNA reproduction defects, mutations in its genomes occur, resulting in an increase in viral surviving characteristics [18]. Overall, due to asymptomatic people, It's nearly impossible to find all SARS CoV-2 mutations and their links to clinical changes. The SARS CoV-2 virus's known mutations, as well as changes in virus structural system and pathogenicity in geographic regions and genotype, are discussed in this review work. Exploring the virus's transmission mechanism is critical for preventing future outbreaks.

SARS-CoV-2 genome structure, origin and genetic patterns

The first patients of the viral (SARS-CoV-2) were discovered in Wuhan, China, in December 2019. It has since spread around the globe, with additional variations being recorded [19]. The new coronavirus type SARS-CoV-2, a positive-sense, single stranded wrapped RNA belongs to the Coronaviridae family, is responsible for the current COVID-19 epidemic. The nucleic acid pattern of SARS CoV-2 has been identified as those of coronavirus [20]. The SARS CoV-2 virus contains a positive-stranded RNA genome comprising 29891 nucleotides and 9860 amino acids [21]. This genome was contained inside spherical nucleocapsid polypeptides, which are then wrapped in an envelope [22].

The high mutation rate of RNA viruses (106 to 10–4 substitutions/nucleotide/cell infection) is an essential element in their evolution. This behaviour can be partially explained by the fact that the RNA polymerase is unable to fix errors during genomic replication [23]. The SARS-CoV-2 genome is more than 30 kb long and contains 14 open reading frames (ORFs) that encode 27 proteins [24, 25].

ORF1(a/b) codes for a protein complex that is post-transnationally split into Sixteen non-structural protein (ranges from nsp1 to 16), that constituted the SARS CoV-2 genome's replicase-transcriptase complex. The complex consists replication-related enzymes such as major protease (nsp5), papa in shape protease (nsp3), and primary RNA based RNA polymerase enzyme (RdRp; nsp12), nsp7-nsp8 primase complex, helicase/tri-phosphatase (nsp13) and exoribonuclease (nsp14), endonuclease (nsp 15) [26]. Furthermore, ORF1a/b, spike (Surface glycoprotein), nucleocapsid (N), membrane (M), envelope (E), hema-gglutinin esterase (HE) or hema-gglutinin (H) and distinct variable accessory proteins are among the ORFs (open reading frames) encoded by coronaviruses. The Spike (S), and

nucleocapsid (N) proteins are the vital structural protein-encoding genes found at the 3′ end of the SARS-CoV-2 genome. Various viral functions are based on these proteins, from virus entry through virus particle production [27]. The S and N proteins are responsible for the formation of surface glycoproteins for association, suppression of IFN induction in hosts, and reduction of interferon regulatory protein-3 activation.

S-protein

The spike, also known as S glycoprotein, is a membrane protein with a molecular mass of 150 kDa that is found on the virus's outermost layer. It aids viral entry by promoting viral particle attachment to the target cell's plasma membrane, which determines the virus's infectivity and pathogenicity. The S proteins of SARS-CoV-1 and 2 form homo-trimers that protrude from surface of the viral and encourage viral entrance into the cells of the host Angiotensin converting enzyme-2 is a protein that interacts with angiotensin (ACE2), The virus's primary receptor is produced in cells of the lower respiratory tract. In other situations, however, they might use CD209L as a substitute receptor. On the other hand, uses the dipeptidyl-peptidase- 4 receptor as its primary target. This strategy, which reduces virus-neutralizing autoimmune response, allows these viruses to propagate throughout diverse cell types [28-30] In the endoplasmatic reticulum, Spike protein is transcribed and translated before being transported to the Golgi apparatus [31].

SARS CoV-2 structural spike protein (S) is classified into two subtypes: namely S1 and S2 protein [32]. SARS CoV-2 structural spike protein (S) is an important resource for specific antibodies. The receptor binding domain (RBD) is found in the S1 subunit, and it is because of this domain that the spike protein identifies and attaches to its target cell's receptor, angiotensin converting enzyme two (ACE2). The S2 component is responsible for the junction of the virus and cell membranes [33] [34]. Three variables have been found as potential contributors to SARS-CoV-2 pathogenicity, and they may have an influence on virulence. The following are some of them: 1) variations in the RBD of the S protein, 2) alterations in accessory protein characteristics, and 3) the insertion of a polybasic-cleavage-site in the S protein Surprisingly, SARS CoV and SARS CoV-2 differed in S protein RBD contains five of the six essential amino acids necessary for S to bind to the ACE2 receptor. The SARS CoV-2 Spike protein binds to ACE2 with a higher affinity than the SARS-CoV Spike protein, that can explain why SARS-CoV-2 is more transmissible and contagious during recent pandemics [14]. Furin, as well as several other proteases, can readily cleave the protein due to the presence of a polybasic cleavage site (RRAR) at the junction of S1 and S2 and is a factor in virus pathogenicity and target species [35].

N- Protein

SARS-nucleocapsid, CoV-2's commonly known as the N protein, is a part of the structure connected to the virus's RNA genome. This protein is involved in a wide range of processes of viral genome-related functions, notably virus replication, viral genome transmission, and the target cell's reaction to virus infection [29, 36]. Twelve nucleotide locations are considered centres in N, contributing for 48% of all mutations [37]. The nucleocapsid protein contains two primary domains known as N-terminal domain and C-terminal domain that bind with the M protein in viral assembly and can form the elongated, elastic helix viral-nucleocapsid by attaching to SARS CoV2-RNA. It's a versatile protein that aids virus transcription and assembly while also being necessary for SARS CoV-2 replication efficiency, and by increasing the host immunological response to viral infection, it plays a vital role in disease pathogenesis. It's worth noting that the virus nucleocapsid, which contains RNA and N-protein, is synthesised in the cytoplasm of the cells [31]. The RNA-binding domain is sited at the N-terminus, whereas the dimerization domain at the C-terminus and is split by phosphorylation is controlled by a core serine arginine rich linkage [38]. The nucleocapsid multimerization, translation inhibition action, and intracellular localisation of SARS-CoV are all affected by phosphorylation of the whole SR-link pattern [39].

Summary of included database studies

The database and Abstracts of Review papers published from 2020-2021 were surveyed on the impact of genome sequencing especially Spike and Nucleocapsid gene mutation of SARS CoV-2 in Covid-19 viral disease transferability and pathogenesis. The search of literature review was conducted by using global databases including science direct, PubMed, Google Scholar and Scopus.

Among theme, 1 study was from Spain [40] , 8 studies from Worldwide [31, 37, 41-46], 4 each study from Italy [47-49] , 1 study from Hong Kong [50, 51] 4 studies from India [52-55], 1 study from Korea [45] , 1 study from Russia [56], 1 study from UK [57] 1 study from Germany [58] 3 studies from China [59-61] 1 study from Israel [62, 63] 1 study from Romania [64] 1 study from Mexico [65] USA [66-70] 3 studies from Bangladesh [71-73], 1 study from Turkey [74], 1 study from Taiwan [75] 1 study from France [76], 1 study from Morocco [77] 1 study from Sweden [78] and 1 study from Philippines [79]. The illustration of characteristics of included studies are revealed in Table 1.

Table 1: Characteristics of "S and N" gene mutation based eligible studies

Author's Name	Location	Study period	Mutation Method	Participant	Muted cases	Region of Mutation	References

ISSN: 2349-8889
Volume-9, Issue-1 (January 2022)

https://doi.org/10.31033/ijrasb.9.1.4

Andres et al., 2020	Spain	March 2020	Deep sequencing of S-gene	18	18	S-gene	[40]
Becera et al., 2020	Worldwide	March-April 2020	NGS	NR	NR	S-gene	[41]
Benvenato et al., 2020	Italy	Jan-April 2020	NGS	79	79	S, N gene	[47]
Cusi et al., 2020	Italy	March 2020	Direct RNA sequencing	1	1	S-gene	[48]
Ip et al., 2020	Hong Kong	Jan-March 2020	Sanger sequencing	12	1	S-gene	[50]
Jacob et al., 2020	India	June 2020	NGS/WGS	600	NR	S-gene	[52]
Kim.et al., 2020	Korea	_	qRT-PCR	4	4	S-gene	[45]
Kozlovskay et al., 2020	Russia	March-April 2020	NGS	220	220	S, N gene	[56]
Micheli et al., 2020	Italy	Feb-April 2020	NGS	20	20	N-gene	[49]
Volz et al., 2021	UK	Jan-June 2020	qRT-PCR	986	21,231	S-gene	[57]
Ziegler et al., 2020	Germany	July 2020	NGS	1	1	S, N gene	[58]
Zhang et al., 2020	China	June-July 2020	NGS and qRT-PCR	6	6	S, N gene	[59]
Zuckerman et al., 2020	Israel	March 2020	NGS	8	8	S, N gene	[62]
Yap et al., 2020	Worldwide	Jan-April 2020	NGS	142	112	S, N gene	[42]
Mcnamra et al., 2020	USA	March-May 2020	NGS	175	175	S-gene	[66]
Surleac et al., 2020	Romania	Jan-Feb 2020	NGS	25	25	S, N gene	[64]
Toboada et al., 2020	Mexico	Feb-March 2020	Oxford Nanopore technology	17	17	S-gene	[65]
Jenjaroenpun et al., 2021	USA	July 2020	NGS	2	2	S-gene	[67]
Devendran et al., 2020	India	April 2020	RNA sequencing	10	10	S, N gene	[53]
Chen et al., 2020	China	Jan-Feb 2020	Genome sequencing	10	10	S, N gene	[60]
Akhter et al., 2020	Bangladesh	May-June 2020	NGS	3	3	S, N gene	[71]
Badua et al., 2020	Worldwide	Jan-May 2020	NGS	151	151	S, N gene	[37]
Barret et al., 2020	USA	Dec-May 2020	NGS	119	119	S-gene	[68]
Bartolini et al., 2020	Italy	Feb-March 2020	qRT-PCR	9	9	S, N gene	[63]
Delmir et al., 2020	Turkey	March-May 2020	qRT-PCR	63	63	S, N gene	[74]
Du et al.,2020	China	Jan-April 2020	NGS	102	102	S, N gene	[61]
Gong et al.,2020	Taiwan	Jan-March 2020	NGS	20	19	S, N gene	[75]

Gupta et al., 2020	Worldwide	Jan-April 2020	NGS	87	87	S, N gene	[43]
Hartely et al., 2021	USA	March-June 2020	NGS	200	173	S-gene	[69]
Hassan et al., 2020	India	May 2020	NGS	128	128	S, N gene	[54]
Islam et al., 2020	Worldwide	May 2020	NGS	444	404	S, N gene	[44]
Jary et al., 2021	France	Jan-Feb 2020	ONT	1	1	N-gene	[76]
Kim et al., 2020	Worldwide	_	NGS	178	178	S, N gene	[45]
Laanarti et al., 2020	Morocco	_	NGS	6	6	S-gene	[77]
Leung et al., 2021	Hong Kong	Feb 2020	NGS	50	50	S-gene	[51]
Ling et al., 2020	Sweden	Feb-May 2020	NGS/WGS	348	348	S, N gene	[78]
Nagy et al., 2021	Worldwide	Dec-Sep 2020	NGS	149,061	149,061	S, N gene	[46]
Parvez et al., 2021	Bangladesh	August 2020	NGS	311	311	S, N gene	[72]
Saha et al., 2020	Bangladesh	April-July 2020	NGS	41	41	S, N gene	[73]
Saha et al., 2020	India	_	NGS	566	566	S, N gene	[55]
Velsco et al., 2020	Philippines	April-July 2020	NGS	23	23	S, N gene	[79]
Wang et al., 2020	USA	July 2020	NGS	24,715	24,715	N-gene	[70]
Toryano et al., 2021	Worldwide	September 2020	NGS	105,276	101,376	S, N gene	[31]

Table 2: Sub group analysis of "S and N" gene mutation in COVID-19

Location	Number of studies	P-value	Prevalence (%)
United states	5	0.000	98.3
Italy	4	0.569	98
India	4	0.244	99.6
Bangladesh	3	0.095	98.7
China	3	0.348	97.5
Hong Kong	2	0.000	60.3
Israel	1	1.000	94.4
Romania	1	1.000	98.1
Sweden	1	0.034	98.8
Morocco	1	1.000	92.9
Spain	1	0.053	99.6
Taiwan	1	1.000	95
Philippians	1	1.000	97.9
Korea	1	1.000	90.0
Worldwide	8	0.000	90

ISSN: 2349-8889
Volume-9, Issue-1 (January 2022)

https://doi.org/10.31033/ijrasb.9.1.4

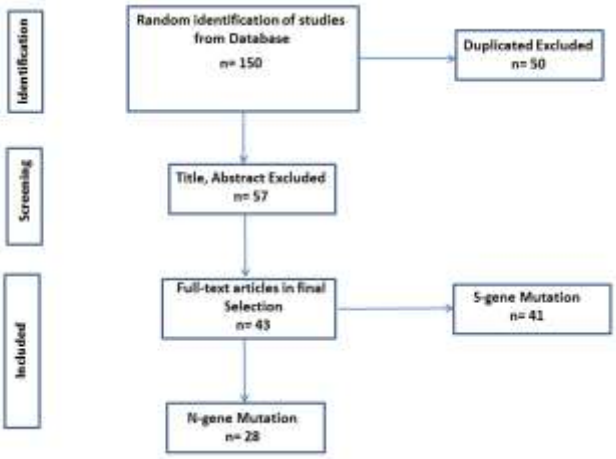

Figure 1: Consort diagram of selected studies

The SARS-CoV-2 S and N gene mutations and its identical variability in clinical trials:

It has been demonstrated that differing national legislation for quarantine and transportation limitations, as well as different ethnicities or genetic origins of humans, may influence COVID-19 mortality and infection rates in different geographic areas [83]. Furthermore, viral alterations over time COVID-19 is considered to be affected by propagation and severity [83]. Yao et al. used Eleven SARS CoV-2 isolates from patients to infect Vero E6 cells (which have a good resemblance to humans in terms of their ACE2 or viral-gate proteins) were used to test the effect of a mutation on virus efficiency. They discovered that SARS-pathogenicity CoV-2's was affected by alterations in the RBD of the S protein, one of the virus's most polymorphic parts. S-D614G has a considerable favourable impact on viral multiplication [84]. the findings of Yao et al., investigation found that the S-protein is more altered, with some mutations, including D614G, having the ability to enhance viral-illness activity and disease stringency [85].According to another study, the most prevalent mutations up to April 5, 2020 are 29095 C > T in the N gene, 28144 T > C in ORF/8 (missense mutation), 8782 C > T in ORF1/ab (synonymous mutation), and 28144 T > C in ORF/8 (missense mutation), which may have an impact on COVID-19 propagation and frequency in most areas. Remarkably, genome sequencing analysis revealed that the shift from C-T is presently the most common nucleotide-variation, with synonymous and missense mutations occurring at considerably higher frequencies than deletion and inclusion mutations [86].

Researchers used in-silico techniques to assess the most harmful mutations in the viral Spike protein's RBD (Leu to Tyr), (Gln to Asn), (Phe to Leu), (Arg to Ile), (Asn to Thr) and (Gln to Tyr), domains, as well as the heptad repeat-1 domain known as (HR1) [87] A further study analysed 80 SARS-CoV-2 variations and 26 mutants to look into S protein mutations and their biological significance. They discovered that variations with the p.614 (Asp- Gly) transformation were significantly more pathogenic, while those with a mutation in the RBD region were significantly less pathogenic. Most neutralising antibodies were resistant to sequences comprising (Val - Ala) (Leu - Arg) , (Ala - Val) and (Phe to Leu) [88].

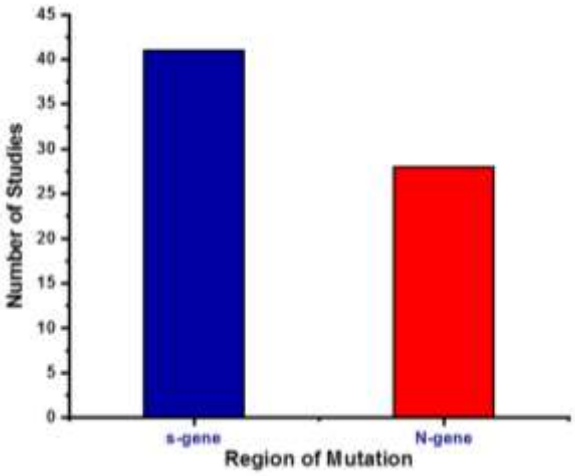

Figure 2: Reported S and N gene mutation region of SARS-CoV-2 in Covid-19

II. DISCUSSION

The COVID-19 epidemic broke out nearly at the end of 2019 and has impacted negatively on medical, economic and social systems all over the world. COVID-19 does not have a therapeutic intervention or vaccination at this time [89]. In order to understand the to track the growth of this novel coronavirus and ensure the efficacy of new screening procedures and vaccines

and therapeutics against COVID-19, researchers must track SARS CoV-2 genomic diversity and mutations in development of this current epidemic. The aa diversity of the four SARS-CoV-2 structural proteins is examined in the biggest set of global sequences in this descriptive statistical analysis. Despite the remarkable sustainability of SARS CoV-2 structural- proteins (99.97 percent in S, 99.99 percent in M, 99.77 percent in N and 99.98 percent in E), we found that almost all proteins had altered positions in different strains, with varied temporal patterns. The S protein (88.9% of 1273 aa) had the greatest number of aa locations permitting alterations, followed by that of the E protein (86.6 percent of 75 aa), and the N protein (85.7 percent). The most notable aa alteration in the S protein was D614G, which was found in eight out of ten worldwide spike examined genomes. This mutation has also been spreading at an increasing rate both worldwide and regionally, establishing itself as the dominant polymorphism in all six regions examined. The D614G mutation is found near the cleavage point in S1's NTD, which is located between RBD and S2 where TMPRSS2 and furin cleave both S1 and S2 [12, 90]. As a result, this mutation may improve SARS CoV-2 viral infectivity by improving the RBD's ability to connect to human ACE2 while lowering associations between S1 and S2 [43]. According to a recent study, D614G improves pathogenicity in cells or cells from the lungs of humans expressing ACE2 receptor in bats or pangolins in cultured cells. The D614G alters the Spike protein shape towards another ACE2 binding fusion competent position [91]. Some other study scanned a database for non - structural ORFs and structured E, M, N, and S proteins using genomic alignments searching. Four mutations have been discovered in ten countries: A870V in India, S221W in Korea, F737C in Sweden, S247R in Australia. According to the report, a single amino acid substitution (L294Q) in the Spike protein is enough to eliminate the phenotype, and the strain develops well above and below 32 degrees Celsius [92]. The D614G mutant has also been found in people from all across the world, encompassing North America, European Countries, Asia, and other areas with less COVID-19 cases [93].

After 21 March 2020, the D614G mutation in the S gene of SARS-CoV increased considerably comparable to before 1 March 2020. Before 1 March 2020, there have been many SARS-CoV2 cases with the D614G mutation in Europeans, but few in China and North America. Between, March 21 to 30, 2020, the D614G mutation is prevalent all around the world.

This review investigated at the mutational pattern of SARS-CoV-2 from 45 studies throughout the globe from December 2019 September 2020. The random-effect model was used to determine a 93.9 % prevalence of SARSCoV-2 mutations in COVID-19 clinical samples. The significant heterogeneity may also be due to the various methods employed to identify mutations, particularly in the Worldwide. The significant heterogeneity may probably be attributable to the multiple parts of the SARS-CoV-2 gene especially Spike protein and Nucleocapsid type poly-protein, that were examined as well as the types of samples employed in researches. The previous researches have been revealed that most mutations occurred at the region of spike-gene along with nucleocapsid-gene region [83].

Our extract data indicate that the large number of 41 spike-gene mutations studies was recorded out of 43 eligible studies which may be due to significance transmissibility and pathogenicity of SARS-CoV-2 infectious virus. D614G was found in the European region from mid-February to late-February 2020, according to our database. It has quickly spread to the North America and United States by early March [65]. In early March 2020, the D614G mutation was detected in Thailand patient diagnosed with COVID-19 disease [94]. The variation was identified in samples taken in China during January to April 2020 [61]. D614G was identified in each and every sample analysed globally until June 2020. The prevalence of the Spike 614G, on the other hand, was not linked to disease severity or COVID-19 morbidity [57]. Owing to lack of data reported from the included papers, this study has a few limitations, such as the difficulty to analyse the influence of the structural detected mutations on patients' infection rates, intensity of the disease, and infectivity. Understanding the effects of mutations on these factors would be extremely useful. Moreover, the majority of previous work only collected viral genetic information from COVID-19 participants from the NCBI and GSAID and science direct databases, limiting our accessibility to demographic information like age, gender as well as clinical dataset like infection rates, manifestations, and severity of disease. To fully comprehend consequences of the aa alternations revealed, more research into related mutations, Protein structure is necessary for virus-host protein interactions. The information presented here could be useful for future it's for a diagnostic, therapeutic, or vaccine purpose, strategies targeted directly at these structural SARS-CoV-2 proteins, which could aid disease prevention and control efforts, as well as a deeper understanding of how this virus spreads in multiple nations or geographical areas. To fully comprehend the global and regional consequences of the aa modifications previously revealed, more research into Virus host protein linkages, as well as associated proteins, are all necessary. The information presented here could be useful for future diagnosing, restorative, and sometimes vaccination strategies that target these structural SARS CoV 2 proteins effectively, assisting disease prevention and control efforts as well as a clearer picture about how this virus spreads across multiple countries or geographical areas.

III. CONCLUSION

COVID-19 disease has a wide range of clinical manifestations due to the virus's mutations and genetic diversity. In this review paper, previous studies especially based on S-gene and N-gene mutation regions in COVID-19 patients across worldwide were examined. The large number of S-gene mutations 41 out of 45 were find out as compared to N-gene mutations. Moreover, the global prevalence of SARS-CoV-2 mutations was calculated at 93.9 % in this review.

REFERENCES

[1] Bakhshandeh, B., et al., *Mutations in SARS-CoV-2; Consequences in structure, function, and pathogenicity of the virus.* Microbial Pathogenesis, 2021: p. 104831.

[2] Lauring, A. and E. Hodcroft, *Variantes genéticas del SARS-CoV-2:¿ qué significan.* JAMA, 2021. 325: p. 529-531.

[3] Letko, M., A. Marzi, and V. Munster, *Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nat Microbiol. 2020; 5: 562–9*. PUBMED.

[4] Perlman, S. and K. McIntosh, *155-Coronaviruses, Including Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS)*, in *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*. 2020, University of Iowa. p. 2072-2080. e3.

[5] Lam, T.T.-Y., et al., *Identification of 2019-nCoV related coronaviruses in Malayan pangolins in southern China. bioRxiv 2020.* Google Scholar, 2020.

[6] Sohrabi, C., et al., *World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19).* International journal of surgery, 2020. 76: p. 71-76.

[7] Andersen, K.G., et al., *A origem proximal do SARS-CoV-2.* Nat Med, 2020. 26(4): p. 450-452.

[8] Ardura, M., et al., *Addressing the impact of the coronavirus disease 2019 (COVID-19) pandemic on hematopoietic cell transplantation: learning networks as a means for sharing best practices.* Biology of Blood and Marrow Transplantation, 2020. 26(7): p. e147-e160.

[9] Qiu, H., et al., *Clinical and epidemiological features of 36 children with coronavirus disease 2019 (COVID-19) in Zhejiang, China: an observational cohort study.* The Lancet infectious diseases, 2020. 20(6): p. 689-696.

[10] Belouzard, S., et al., *Mechanisms of coronavirus cell entry mediated by the viral spike protein.* Viruses, 2012. 4(6): p. 1011-1033.

[11] Zhou, P., et al., *& Shi, ZL (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin.* Nature. 579(7798): p. 270-273.

[12] Ahmadpour, D., P. Ahmadpoor, and L. Rostaing, *Impact of circulating SARS-CoV-2 mutant G614 on the COVID-19 pandemic.* Iran J Kidney Dis, 2020. 14(5): p. 331-334.

[13] Walls, A.C., et al., *Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer.* Nature, 2016. 531(7592): p. 114-117.

[14] Ortega, J.T., et al., *Role of changes in SARS-CoV-2 spike protein in the interaction with the human ACE2 receptor: An in silico analysis.* EXCLI journal, 2020. 19: p. 410.

[15] Rambaut, A., et al., *A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol. 2020 Jul 15; 1–5*.

[16] Kaufmann, S.H., et al., *Host-directed therapies for bacterial and viral infections.* Nature reviews Drug discovery, 2018. 17(1): p. 35-56.

[17] Grubaugh, N.D., M.E. Petrone, and E.C. Holmes, *We shouldn't worry when a virus mutates during disease outbreaks.* Nature microbiology, 2020. 5(4): p. 529-530.

[18] Pachetti, M., et al., *Impact of lockdown on Covid-19 case fatality rate and viral mutations spread in 7 countries in Europe and North America.* Journal of Translational Medicine, 2020. 18(1): p. 1-7.

[19] Flores-Alanis, A., et al., *Molecular epidemiology surveillance of SARS-CoV-2: mutations and genetic diversity one year after emerging.* Pathogens, 2021. 10(2): p. 184.

[20] Chen, Y., et al., *Structure analysis of the receptor binding of 2019-nCoV.* Biochemical and biophysical research communications, 2020. 525(1): p. 135-140.

[21] Chan, J.F.-W., et al., *Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan.* Emerging microbes & infections, 2020. 9(1): p. 221-236.

[22] Li, F., *Structure, function, and evolution of coronavirus spike proteins.* Annual review of virology, 2016. 3: p. 237-261.

[23] Peck, K.M. and A.S. Lauring, *Complexities of viral mutation rates.* Journal of virology, 2018. 92(14): p. e01031-17.

[24] of the International, C.S.G., *The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2.* Nature microbiology, 2020. 5(4): p. 536.

[25] Wu, A., et al., *Genome composition and divergence of the novel coronavirus (2019-nCoV).* 2020.

[26] Kim, D., et al., *The architecture of SARS-CoV-2 transcriptome.* Cell, 2020. 181(4): p. 914-921. e10.

[27] Chen, Y., Q. Liu, and D. Guo, *Emerging coronaviruses: genome structure, replication, and pathogenesis.* Journal of medical virology, 2020. 92(4): p. 418-423.

[28] Wu, A., et al., *Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China.* Cell host & microbe, 2020. 27(3): p. 325-328.

[29] Astuti, I., *Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response.* Diabetes & Metabolic

Syndrome: Clinical Research & Reviews, 2020. 14(4): p. 407-412.
[30] Yin, C., *Genotyping coronavirus SARS-CoV-2: methods and implications.* Genomics, 2020. 112(5): p. 3588-3596.
[31] Troyano-Hernáez, P., R. Reinosa, and Á. Holguín, *Evolution of SARS-CoV-2 envelope, membrane, nucleocapsid, and spike structural proteins from the beginning of the pandemic to September 2020: a global and regional approach by epidemiological week.* Viruses, 2021. 13(2): p. 243.
[32] Satarker, S. and M. Nampoothiri, *Structural proteins in severe acute respiratory syndrome coronavirus-2.* Archives of medical research, 2020. 51(6): p. 482-491.
[33] Yang, J., et al., *Molecular interaction and inhibition of SARS-CoV-2 binding to the ACE2 receptor.* Nature communications, 2020. 11(1): p. 1-10.
[34] Huang, Y., et al., *Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19.* Acta Pharmacologica Sinica, 2020. 41(9): p. 1141-1149.
[35] Andersen, K.G., et al., *The proximal origin of SARS-CoV-2.* Nature medicine, 2020. 26(4): p. 450-452.
[36] Chang, T.-J., et al., *Genomic analysis and comparative multiple sequences of SARS-CoV2.* Journal of the Chinese Medical Association, 2020. 83(6): p. 537-543.
[37] Badua, C.L.D., K.A.T. Baldo, and P.M.B. Medina, *Genomic and proteomic mutation landscapes of SARS-CoV-2.* Journal of medical virology, 2021. 93(3): p. 1702-1721.
[38] Kang, S., et al., *Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites.* Acta Pharmaceutica Sinica B, 2020. 10(7): p. 1228-1238.
[39] Peng, T.Y., K.R. Lee, and W.Y. Tarn, *Phosphorylation of the arginine/serine dipeptide-rich motif of the severe acute respiratory syndrome coronavirus nucleocapsid protein modulates its multimerization, translation inhibitory activity and cellular localization.* The FEBS journal, 2008. 275(16): p. 4152-4163.
[40] Andres, C., et al., *Naturally occurring SARS-CoV-2 gene deletions close to the spike S1/S2 cleavage site in the viral quasispecies of COVID19 patients.* Emerging Microbes & Infections, 2020. 9(1): p. 1900-1911.
[41] Becerra-Flores, M. and T. Cardozo, *SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate.* International journal of clinical practice, 2020. 74(8): p. e13525.
[42] Yuan, F., et al., *Global SNP analysis of 11,183 SARS-CoV-2 strains reveals high genetic diversity.* Transboundary and emerging diseases, 2021. 68(6): p. 3288-3304.
[43] Gupta, A.M., J. Chakrabarti, and S. Mandal, *Non-synonymous mutations of SARS-CoV-2 leads epitope loss and segregates its variants.* Microbes and infection, 2020. 22(10): p. 598-607.
[44] Islam, M.R., et al., *Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity.* Scientific reports, 2020. 10(1): p. 1-9.
[45] Kim, S., et al., *The progression of sars coronavirus 2 (sars-cov2): mutation in the receptor binding domain of spike gene.* Immune network, 2020. 20(5).
[46] Nagy, Á., S. Pongor, and B. Győrffy, *Different mutations in SARS-CoV-2 associate with severe and mild outcome.* International journal of antimicrobial agents, 2021. 57(2): p. 106272.
[47] Benvenuto, D., et al., *Evidence for mutations in SARS-CoV-2 Italian isolates potentially affecting virus transmission.* Journal of medical virology, 2020. 92(10): p. 2232-2237.
[48] Cusi, M.G., et al., *Whole-genome sequence of SARS-CoV-2 isolate Siena-1/2020.* Microbiology Resource Announcements, 2020. 9(39): p. e00944-20.
[49] Micheli, V., et al., *Geographical reconstruction of the SARS-CoV-2 outbreak in Lombardy (Italy) during the early phase.* Journal of Medical Virology, 2021. 93(3): p. 1752-1757.
[50] Ip, J.D., et al., *Intra-host non-synonymous diversity at a neutralizing antibody epitope of SARS-CoV-2 spike protein N-terminal domain.* Clinical Microbiology and Infection, 2021. 27(9): p. 1350. e1-1350. e5.
[51] Leung, K.S.-S., et al., *Territorywide study of early coronavirus disease outbreak, Hong Kong, China.* Emerging infectious diseases, 2021. 27(1): p. 196.
[52] Jacob, J.J., et al., *Genomic evolution of severe acute respiratory syndrome Coronavirus 2 in India and vaccine impact.* Indian journal of medical microbiology, 2020. 38(2): p. 210-212.
[53] Devendran, R., M. Kumar, and S. Chakraborty, *Genome analysis of SARS-CoV-2 isolates occurring in India: Present scenario.* Indian journal of public health, 2020. 64(6): p. 147.
[54] Hassan, S.S., et al., *Missense mutations in SARS-CoV2 genomes from Indian patients.* Genomics, 2020. 112(6): p. 4622-4627.
[55] Saha, O., M.S. Hossain, and M.M. Rahaman, *Genomic exploration light on multiple origin with potential parsimony-informative sites of the severe acute respiratory syndrome coronavirus 2 in Bangladesh.* Gene reports, 2020. 21: p. 100951.
[56] Kozlovskaya, L., et al., *Isolation and phylogenetic analysis of SARS-CoV-2 variants collected in Russia during the COVID-19 outbreak.* International Journal of Infectious Diseases, 2020. 99: p. 40-46.
[57] Volz, E., et al., *Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity.* Cell, 2021. 184(1): p. 64-75. e11.
[58] Ziegler, K., et al., *SARS-CoV-2 samples may escape detection because of a single point mutation in the N gene.* Eurosurveillance, 2020. 25(39): p. 2001650.
[59] Zhang, Y., et al., *Genomic characterization of SARS-CoV-2 identified in a reemerging COVID-19*

outbreak in Beijing's Xinfadi market in 2020. Biosafety and Health, 2020. 2(4): p. 202-205.
[60] Chen, J., et al., *Epidemiological and Genomic Analysis of SARS-CoV-2 in 10 Patients From a Mid-Sized City Outside of Hubei, China in the Early Phase of the COVID-19 Outbreak.* Frontiers in public health, 2020. 8.
[61] Du, P., et al., *Genomic surveillance of COVID-19 cases in Beijing.* Nature communications, 2020. 11(1): p. 1-9.
[62] Zuckerman, N.S., et al., *Comprehensive analyses of SARS-CoV-2 transmission in a public health virology laboratory.* Viruses, 2020. 12(8): p. 854.
[63] Bartolini, B., et al., *SARS-CoV-2 phylogenetic analysis, Lazio region, Italy, February–march 2020.* Emerging infectious diseases, 2020. 26(8): p. 1842.
[64] Surleac, M., et al., *Molecular epidemiology analysis of SARS-CoV-2 strains circulating in Romania during the first months of the pandemic.* Life, 2020. 10(8): p. 152.
[65] Taboada, B., et al., *Genomic analysis of early SARS-CoV-2 variants introduced in Mexico.* Journal of virology, 2020. 94(18): p. e01056-20.
[66] McNamara, R.P., et al., *High-density amplicon sequencing identifies community spread and ongoing evolution of SARS-CoV-2 in the Southern United States.* Cell reports, 2020. 33(5): p. 108352.
[67] Jenjaroenpun, P., et al., *Two SARS-CoV-2 genome sequences of isolates from rural US patients harboring the D614G mutation, obtained using Nanopore sequencing.* Microbiology Resource Announcements, 2020. 10(1): p. e01109-20.
[68] Barrett, C., et al., *Multiscale feedback loops in SARS-CoV-2 viral evolution.* Journal of Computational Biology, 2021. 28(3): p. 248-256.
[69] Hartley, P., et al., *Genomic surveillance revealed prevalence of unique SARS-CoV-2 variants bearing mutation in the RdRp gene among Nevada patients.* medRxiv, 2020.
[70] Wang, R., et al., *Decoding asymptomatic COVID-19 infection and transmission.* The journal of physical chemistry letters, 2020. 11(23): p. 10007-10015.
[71] Akter, S., et al., *Coding-complete genome sequences of three SARS-CoV-2 strains from Bangladesh.* Microbiology Resource Announcements, 2020. 9(39): p. e00764-20.
[72] Parvez, M.S.A., et al., *Genetic analysis of SARS-CoV-2 isolates collected from Bangladesh: Insights into the origin, mutational spectrum and possible pathomechanism.* Computational biology and chemistry, 2021. 90: p. 107413.
[73] Ghosh, N., et al., *Genome-wide analysis of 10664 SARS-CoV-2 genomes to identify virus strains in 73 countries based on single nucleotide polymorphism.* Virus research, 2021. 298: p. 198401.
[74] Demir, A.B., et al., *Identification of the nucleotide substitutions in 62 SARS-CoV-2 sequences from Turkey.* Turkish Journal of Biology, 2020. 44(SI-1): p. 178-184.
[75] Gong, Y.-N., et al., *SARS-CoV-2 genomic surveillance in Taiwan revealed novel ORF8-deletion mutant and clade possibly associated with infections in Middle East.* Emerging microbes & infections, 2020. 9(1): p. 1457-1466.
[76] Jary, A., et al., *Evolution of viral quasispecies during SARS-CoV-2 infection.* Clinical Microbiology and Infection, 2020. 26(11): p. 1560. e1-1560. e4.
[77] Laamarti, M., et al., *Genome sequences of six SARS-CoV-2 strains isolated in morocco, obtained using oxford nanopore minion technology.* Microbiology Resource Announcements, 2020. 9(32): p. e00767-20.
[78] Ling, J., et al., *Spatio-temporal mutational profile appearances of Swedish SARS-CoV-2 during the early pandemic.* Viruses, 2020. 12(9): p. 1026.
[79] Velasco, J.M., et al., *Coding-complete genome sequences of 23 SARS-CoV-2 samples from the Philippines.* Microbiology Resource Announcements, 2020. 9(43): p. e01031-20.
[80] Aljahdali, B.A., *VITAMIN D STATUS AND AUTOIMMUNE DISEASE (HASHIMOTO'S THYROIDITIS) IN SAUDI ARABIAN WOMEN.* 2018.
[81] Botelho, I.M.B., et al., *Vitamin D in Hashimoto's thyroiditis and its relationship with thyroid function and inflammatory status.* Endocrine journal, 2018. 65(10): p. 1029-1037.
[82] Roman, G.S., et al. *Vitamin D deficiency and Hashimoto's thyroiditis*. in *Endocrine Abstracts*. 2021. Bioscientifica.
[83] Pachetti, M., et al., *Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant.* Journal of translational medicine, 2020. 18(1): p. 1-9.
[84] Yao, H.-P., et al., *Patient-derived mutations impact pathogenicity of SARS-CoV-2.* 2020.
[85] Korber, B., et al., *Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2.* BioRxiv, 2020.
[86] Khailany, R.A., M. Safdar, and M. Ozaslan, *Genomic characterization of a novel SARS-CoV-2*. Gene reports, 2020. 19: p. 100682.
[87] Ahamad, S., H. Kanipakam, and D. Gupta, *Insights into the structural and dynamical changes of spike glycoprotein mutations associated with SARS-CoV-2 host receptor binding.* Journal of Biomolecular Structure and Dynamics, 2020: p. 1-13.
[88] Li, Q., et al., *The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity*. Cell, 2020. 182(5): p. 1284-1294. e9.
[89] Shang, J., et al., *Cell entry mechanisms of SARS-CoV-2.* Proceedings of the National Academy of Sciences, 2020. 117(21): p. 11727-11734.
[90] Bhattacharyya, C., et al., *Global spread of SARS-CoV-2 subtype with spike protein mutation D614G is shaped by human genomic variations that regulate expression of TMPRSS2 and MX1 genes.* BioRxiv, 2020.

ISSN: 2349-8889
Volume-9, Issue-1 (January 2022)

https://doi.org/10.31033/ijrasb.9.1.4

[91] Yurkovetskiy, L., et al., *Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant.* Cell, 2020. 183(3): p. 739-751. e8.
[92] Tsai, P.-H., et al., *Genomic variance of Open Reading Frames (ORFs) and Spike protein in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).* Journal of the Chinese Medical Association, 2020. 83(8): p. 725.
[93] Korber, B., et al., *Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus.* Cell, 2020. 182(4): p. 812-827. e19.
[94] Yap, P.S.X., et al., *An overview of the genetic variations of the SARS-CoV-2 genomes isolated in Southeast Asian countries.* 2020.