ACTA SCIENTIFIC MICROBIOLOGY (ISSN: 2581-3226)

Volume 4 Issue 5 May 2021

Review Article

Mutants and Variants of SARSCoV-2 Across the Globe - A Comprehensive Review

Anju Kaushal[1-3]*

[1]*Central Research Institute, Kasauli, India*

[2]*Panjab University, Chandigarh, India*

[3]*New Zealand Organization for Quality, New Zealand*

***Corresponding Author:** Anju Kaushal, Experienced Microbiologist/Biotechnologist and Quality Professional, Auckland, New Zealand.

Received: March 19, 2021

Published: April 27, 2021

Abstract

Variants of SARSCoV-2 could have the potential to disrupt virus to elevate drug resistance or enhance their ability to escape the neutralization antibodies and innate immune response. Changes occurred by mutation in the spike protein are quite significant for host receptor and antibodies binding. Presence of insufficient protective immunity in the population help transmitting the virus faster; could enhance its capacity for antigenic drift to produce relevant mutations interfering with clinical efficacy of vaccines. The spike protein sequenced from Wuhan strain (the index strain) are presently used in testing reagents and immunogen production. Currently, the relevant mutations reported to provide the specific functional significance are D614G, N501Y, K417N, E484K, L452R etc. D614G is the more transmissible form. Emerging new mutations within two viruses may recombine in a host to make it more virulent; could enforce to develop new anti-viral therapies considering multiple recombinants. Various variants emerged during past months in the ongoing COVID-19 pandemic, as B.1.1.7 (UK), B.1.351 (South Africa), P.1 lineage (Brazilian) and B.1.427, B.1.429 and B.1.617. Deleterious mutations at the immunodominant epitopes could make the Abs susceptible and lead to facilitate the ADE.

This article describes the prevalent mutations turning into recombinants and variants during pandemic and their implications on transmission and infection rates to implement new control measures. Performance of vaccines and immune sera against the mutated forms tracking mutations to inform urgency on developing the strategies for precise vaccine construct, antibody therapeutics and diagnostic reagents

Keywords: Mutation; SARSCoV-2; Variant; D614G; Genome Sequencing; Phylogenetic Alignment

Abbreviations

AP-1- Activator Protein-1; ARDS- Acute Respiratory Distress Syndrome; A, G, T, C, U- Adenine, Thymine, Guanine, Cytosine, Uracil; Abs- Antibodies; ACE2- Angiotensin Converting Enzyme 2; ADE- Antibody Dependent Enhancement; BALF- Bronchoalveolar Lavage Fluid; Lineage B.1.1.7 – UK variant, 20I/501Y.V1- UK variant, VOC-202012/01- variant of concern UK variant; B.1.1.248 lineage- Brazil variant; Lineage P.1- Brazil variant; 20J/501Y.V3- Brazil variant ; B.1.106-South African variant; 501Y.V2 – South Africa variant; CDC US- Centre for disease control US; CT-Cycle threshold (CT); B.1.427-California variant , B.1.429-California variant, CAL 20C- California variant; CCL2, CCL3, CCL5- Chemokine ; CXCL8, CXCL10- Chemokine C-X-C motif; B.1.351 lineage -South Africa variant ; B.1.617-Double mutant Indian Variant; DNA- Deoxyribonucleic acid; D614G- emerged spike mutation; E- envelope protein gene; FDR- genes with false discovery rate; IFN I- type I interferon ; IL-1β- Interleukin - 1β; IL-6- Interleukin-6 ; IRF7- Interferon Regulatory Factor 7; IFN-1-Interferon 1; IRF3- Interferon Regulatory factor 3; GISAID- Global initiative on sharing influenza data; ΔG- Gibbs free energy; mRNA-LNP platform- mRNA Lipid nanoparticle; MHC-I- Major Histocompatibility Complex I; MOI- Multiplicity of Infection; M-Membrane Protein Gene; MyD88- Myeloid Differentiation Factor 88 ; N- Nucleocapsid Protein Gene; NSP- Non-Structural Proteins; NTD- N-Terminal Domain; NCBI database- National Centre for Biotechnology Information; NIID-The National Institute of Infectious Diseases ; ORF- Open Reading Frame; PANGO lineages-Phylogenetic Assignment of Named Global Outbreak Lineages; P.2 lineage- B.1.1.28.2 variant originated from Rio de Janeiro; RNA- Ribonucleic acid; RT-PCR- Reverse Transcriptase Polymerase chain reaction; p38MAPK- p38 Mitogen Activated protein Kinases

pathway; RBM-Receptor-Binding Motif; RdRp- RNA-dependent RNA polymerase; RBD- Receptor Binding Domain; RBM-Receptor-Binding Motif; S- Spike; SNPs - Single Nucleotide Polymorphs; SGTF- S-Gene Target Failure; SARSCoV-2- Severe Acute Respiratory Syndrome Coronavirus-2; SARS CoV- Severe Acute Respiratory Syndrome Coronavirus -1; 5'UTR – 5' Untranslated Region; TNFα-Tumor Necrosis Factor-α; TRAF6- Tumor necrosis factor associated factor-6; TLRs- Toll Like Receptor ; VOC- Variant of concern; WHO- World Health Organization.

Introduction

SARSCoV-2 virus is the causative agent of COVID-19 pandemic. The outbreak first exploded in Wuhan, China in Dec 2019. WHO declared it as a public health emergency of international concern on 30 Jan 2020 and confirmed as a pandemic on 11 March 2020. It has affected 219 countries and territories around the globe. As of 4th May 2021, more than 154 million people are the confirmed cases, with more than 3.2 million deaths attributed to COVID-19, making it deadliest pandemic in history [2].

People contract the viral infection mainly through inhalation. Severity of the disease depends upon how patient's immune system reacts to the virus; it could be asymptomatic and symptomatic. Virus gets detected in the clinical samples (nasopharyngeal) by RT-PCR. Genomic sequencing is being done for phylogenetic analysis for epidemiological purpose [7,8].

Case fatality rates were reported to be varied in different countries during pandemic, due to diversified demography and implementation of types of control measures are considered in different countries to prevent the COVID-19.

Mutations are the cause of great deal of genetic diversity, which is responsible for error occurred in RdRp in any genome replication [3]. SARSCoV-2 virus has lower mutation rates comparatively to other RNA viruses and its stable markers found to be useful for contact tracing. The spike protein mutations can result into higher transmission and infection rates, even help escaping the antibodies. The detailed repercussions of these mutations are yet to be done in terms of escaping the vaccines efficacy naturally in host.

Variability in a genome is induced and accomplished by different recombinational events. Multiple variants in a population are denoted as 'Quasispecies', have the capability to enhance drug resistance/ or escaping the neutralizing antibodies to lapse the developed protective immunological response. The innate and adaptive immunity evasion in immunocompromised hosts help prolong the infection might accelerate the process of generating mutants. The viruses have the unique ability to evolve themselves to cross the species barriers, resulting in zoonotic infections [3].

Recombination of two and more strains have generated the highly mutated hybrid forms of SARSCoV-2, eventually started the new waves of pandemic from October 2020 on in different geographies. These forms are classified in pangolin lineage e.g., B.1.1.7 (501Y.V1) identified in UK, observed with important changes in N501Y. This mutation in combination with 69-70 deletion appeared to have the two-fold higher infectivity. The South African variant B.1.351(501Y.V2) is defined by having mutational pattern of K417N, E484K, and N501Y with special interest with spike protein [34]. The variant of concern in US, B.1.427/ B.1.429 (California variant), is driven by mutation L452R [49]. In April 2021, a double mutant appeared in India to be classified as B.1.617, which accumulated the mutations E484Q and L452R [52].

Variants to initiate the new phases in pandemic with highest transmission rate and different pathogenic outcomes in patients, prompted the new lockdown measures.

This review highlighted the various mutational patterns of SARS CoV-2 clades in different geographies around the globe. The mutational analysis by genome sequencing assist implementing the required containment/ lockdown measures in the communities to break the chains of virus transmission. In addition to that these studies help support scientific community working in fields of immunology and vaccines & sera development. Effectiveness of monoclonal antibodies, convalescent sera from recovered individuals and sera obtained from vaccinated individuals, who are immunized with different vaccines prepared on different platforms, could instruct the promptness on updating the gene sequences and immunogens required for vaccine construct and diagnostics.

Insufficient immunological experience on SARS CoV-2

Our immune system doesn't recognize the new zoonotic viruses until they infect us and we develop some immunity against them. This escalates their transmission making difficult to contain the virus. Hence, virus undergoes antigenic drift and can spread the lethal mutations potentially in the population. These mutations conceivably reduce the effectiveness of vaccines. At present, the testing reagents and vaccine immunogens are based on the initial sequence obtained from Wuhan. SARS CoV-2 and SARS CoV both use ACE2 receptors, but SARSCoV-2 S-protein is 10-20 times higher than SARS CoV [4,5,20].

This is the first experience of pandemic with SARSCoV-2 to humankind. Insufficient studies are available in context to the immunological experience with the virus. Asymptomatic individuals have the innate immune response to control the virus and CD8+ T-cells provide the cell mediated immunity. In case of symptomatic individuals, SARSCoV-2 escape the cell mediated response. It has been hypothesized that viral RNA inhibits the two important intracellular pathways i) MyD88 and IRF7 to produce IFN-1 ii) TRAF6 and IRF3 to activate T-cells. These pathways are recognized by TLRs (7/8,3). In opposite, the third intracellular pathway iii) p38MAPK and AP-1 – regulate the production of proinflammatory molecules, gets activated. Proinflammatory molecules activate the macrophages mostly discovered in ICU patients with high level of cytokines and chemokines (IL-1β, IL-6, TNFα, CCL2, CCL3, CCL5, CXCL8, CXCL10). Patients developed with pneumonia turns into ARDS due to hyperinflammation. The secondary hemophagocytic-lymphohistiocytosis (HLH) with hypercytokinemia with multiorgan failure leading to the death [7].

Most of the asymptomatic individuals appear to be recovered and develop detectable antibodies against virus within 7-10 days.

The main factors hinder the immunological studies are:

- How asymptomatic individuals develop the humoral and cellular immunity.
- How lymphocytes decreased in disease progression and attain abnormal functionality.
- Impaired regulatory mechanism to interfere with protective immune cells, help propagate the virus faster.

These deteriorated immunological interactions should be addressed to develop safe and effective vaccine/ sera therapeutics to recover systemic interaction.

There is no specific drug is available particularly against COVID-19. Most of the drugs are being used for repurposing trials e.g., dexomethsone, lopinavir, ritonavir, tocilizumab, remdesivir etc. [7]

COVID-19 vaccines developed amid pandemic on novel platforms, completed the clinical trials successfully and received approval for mass vaccination. Vaccines are designed to train the protective immune system. T-cells detect immunogen and alarm the T and B-cell. Helper T-cells manufacture antibodies. 1st and 2nd doses prime the immune system. The active immunity remains for months and it lowered down later on. Memory cells retain the information for many years / or rest of the life to combat same pathogenic attack [1].

Although the diversity among the pandemic sequences is low, but its wider spread across the globe has given the virus great opportunities for natural selection to act upon rare but favourable mutations. If epidemic extends for a longer period would certainly allow the selection of advantageous antigenic determinants to confront the developed immunity. The best example of antigenic drift is to manufacture new influenza vaccine in every season of the year [4,14,17].

Mutations reported in UK, Europe, South Africa, Brazil, US, India and other countries are mainly from G, GR and GH clades of SARSCoV-2. Each mutation occurred are not responsible for increased transmissibility and severity of the disease.

Genomic structure and current phylogenetic classification of SARS CoV-2

SARSCoV-2 is a member of genus Betacoronavirus and is closely related to SARSCoV. CoVs are spherical with size ~100-120 nM, enveloped with lipids. The petal shaped 20-40 nm long spike glycoproteins protrude from envelope. The other shorter projection is haemagglutinin (HA) esterase protein. The nucleocapsid contain single stranded non-segmented positive sense RNA genome. M protein provide structural support and E protein is a small membrane protein.

Figure 1: Schematic diagram of the genomic structure of the 29.3 kilobase 2019-novel coronavirus (SARSCoV-2) gene and domain structure of the1273amino acid. S: Spike Glycoprotein; E: Envelope Protein Gene; M: Membrane Protein Gene; N: Nucleocapsid Protein Gene; RBM: Receptor-Binding Motif; RdRp: RNA-Dependent RNA Polymerase; S: Spike Protein Gene.

Metagenomic Next Generation Sequencing elucidated the SARSCoV-2 genome structure in January 2020. The positive single stranded genome nucleotides are 29.3 kb long. Sub-genomic RNA strands are transcribed from negative strand RNAs. The genome transcribed from 5' of RNA in a non-contiguous manner, where RdRp skips from one part of the genome to the next to generate RNAs. The spike protein (S) is a 1273 amino acid homotrimeric class I fusion protein allows the viral membrane to attach to the host cell membrane. The 3rd open reading frame in the genome encodes spike protein. It is divided into number of functional domains. The first head contains S1-N-terminal RBD and stem has S2 C-terminal membrane fusion domain followed by two heptad regions (HR1 and HR2), transmembrane domain (TM) and cytosolic tail. RBD S1 is ~200 amino acid long. At the time of binding S1 exposes S2 domain. The glycosylated ectodomain of spike protein present in the prefusion and post fusion conformation. Cysteine residue of cytosolic tail binds to the ACE2 [53].

GISAID tracks the occurrence and spread of variants. Genomic epidemiology is largely depending upon the clades and lineages based on their marker mutations. The outbreak studies around the globe, assist in building the nomenclature system.

PANGOLIN lineages provide the more detailed outbreak cluster information (Table:1). The lineage 'L' relates to the reference genome NC_045512.2 and its loci are also described in all clades G, GH, GR, S and V. The early phylogenetic analysis is characterized by two major clades, first is 'S' named after the mutation in ORF8 L84S, also defined by silent genomic mutation C8782T and second clade is 'V' from the ORF3a: G251V mutation co-occurred with the NSP6:L37F events. Clusters of all genomic mutations highlights the five major phylogenetic groups G, GH, GR, S and V and their constituting mutations. 'O' clade belongs to the sequences which don't match with the clades in phylogenetic tree [6,9].

Distribution of SARSCoV-2 clades in different geographies during pandemic

The first mutated virus appeared in clade 'S' alongside with clade 'O' in sequences dataset. Around mid -January the clade 'V' was mutated in ORF3a and NSP6 region, right on the same time when the original clade 'G' appeared. In mid -February clade G with sub clades GH and GR were increased in Europe and North America (Figure 6). The spike mutation D614G and ORF3a Q57H were found in 50% of sequences of clade GH in North America. Oceania observed mixed clades, but clade 'G' was present in Africa [Fig 6C]. The clade 'G' with its sub parts GH and GR were appeared with the gradual disappearance of clade 'L' and 'V'. Clade 'S' was found to be in adequate minority, despite its declining trend in US and Spain [6].

Table 1: Current phylogenetic classification provided by GISAID. Current definition of characterizing mutations of SARS-CoV-2 phylogenetic categorization systems [GISAID clades and PANGOLIN lineages (Rambaut A., *et al*. 2020) (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7387429/table/T2/?report=objectonly).

GISAID Clade	Pangolin lineage	Nucleotide features	Given effect on protein sequence
G	B.1	C241T C3037T C14408T A23403G	5'UTR NSP3:F106F NSP12b:P314L S:D614G
GH	B.1.*	C241T C3037T C14408T A23403G G25563T	5'UTR NSP3:F106F NSP12b:P314L S:D614G ORF3a:Q57H
GR	B.1.1	C241T C3037T C14408T A23403G GGG28881AAC	5'UTR NSP3:F106F NSP12b:P314L S:D614G N:RG203KR
S	A	C8782T T28144C	NSP4:S76S ORF8:L84S
V	B.2	G11083T G26144T	NSP6:L37F ORF3a:G251V
L - Reference of all defining clades G, GH, GR, S, and V			
O - Others			

Clade G (variant of spike protein S -D614G), Clade V (variant of ORF3a coding protein NS3-G251), and clade S (variant ORF8-L84S) were studied in the beginning to accomplish the genomic studies [9]. Daniele et al, have analyzed and highlighted all the mutations within subclades and genomic variables; in order to implement the new control measures on regional basis and to find the solution for vaccine design.

He obtained the mutational genome sequences from North America, South America, Africa, Europe and Asia to be aligned with reference genome of SARSCoV-2 (NC_045512.2). Annotated algorithm noticed the 353,000 mutational events. The average mutation rate was 7.23 continentwise, but it was different in countrywise. The variability occurred was due to the different sampling conditions at different places, including the lower mutations in the beginning [6].

Figure 2: A) Number of mutational events per sample for all SARS CoV-2 genome analysed. B) Distribution of number of mutations per continent across the average number of mutations per sample, worldwide. The box plot rectangles are between the 1st and 3rd quartile + 1.5interquartile range. C) Sequenced genomes in countries have shown higher(red) mutations and blue (lower) mutations. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7387429/figure/F1/?report=objectonly).

Amino acid (aa) changing events were noticed more than 50%, with highly observed SVPs (single nucleotide polymorphs) along with indel (insertion and deletion) events. Mutational SNPs were found in the coding regions, and others are in 5' to 3' UTR region with short frameshift deletion of indel. These mutations were accounted for shortening the protein length with no addition of stop codons (which is a rare event). All the continents were providing the similar pattern (Figure 3A).

Most of the mutational events are C>T transition (>50.0%) and A>G transitions (>14.0%) in America, Europe and Africa. C>T transition event picked up in Asia and Oceania too. ATG codon deletion is the common indel with >1200 occurrences and substitution of GGG found worldwide. The clade GR is formed from nucleocapsid in the phylogenetic tree. 'T' represents 'U' of RNA sequences.

Protein profile by mutations is similar in continents. Most popular mutation is A23403G forming clade G, which was responsible to carryout changes from Aspartate to Glycine at protein position 614 (D614G), a major 'aa' change in spike. This mutation strengthens the attachment of spike with ACE2. The four mutations, recognized within G clade co-occurring with similar frequency were discovered as C241T, C3037T, C14408T and A23403G. Another mutation G11083T specifically reported in Asia from Dec 2019 to March 2020 [5,6,10].

Figure 3: A)'aa' changing events in continents by SNP, deletion, and insertion. B) Nucleotide changes across the continents. 'T' is a counterpart of 'U'. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7387429/figure/F2/?report=objectonly).

Figure 4: A) Nucleotide changes in the most frequent mutational events in different continents B) Protein changes in most frequent mutational events in different continents. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7387429/figure/F3/?report=objectonly).

The spike D614G change & GH derivative, is featured by ORF3a:057H mutation and GR, all together affected by RG203KR nucleocapsid mutation. These mutations cooccurred in C241T, C3037T, C14408T, and A23403G which relates to the current phylogenetic classifications (Table 1 &2, Figure 5 and 12).

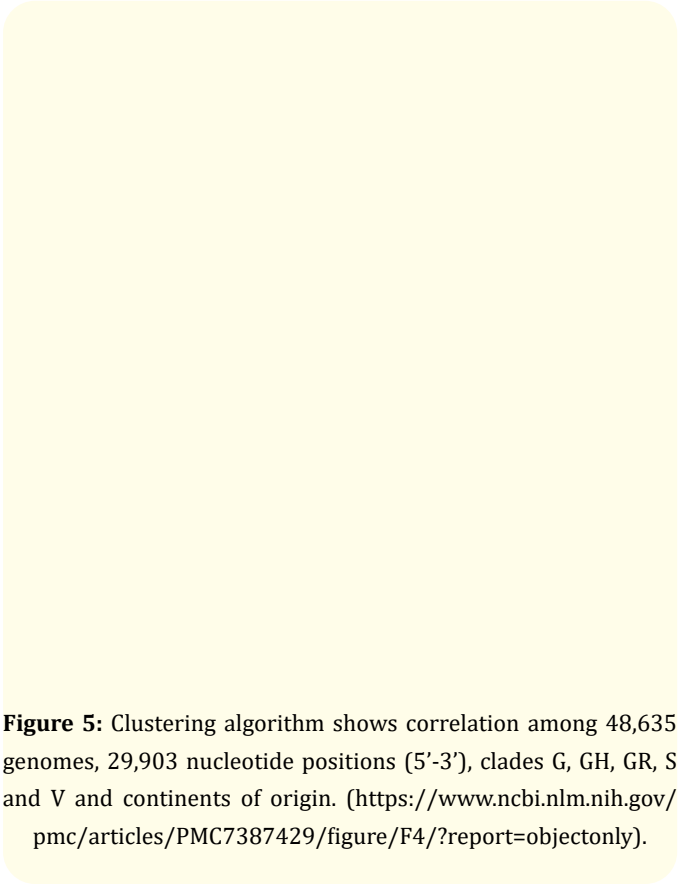

Figure 5: Clustering algorithm shows correlation among 48,635 genomes, 29,903 nucleotide positions (5'-3'), clades G, GH, GR, S and V and continents of origin. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7387429/figure/F4/?report=objectonly).

The G and GR clades are prevalent in Europe and S and GH are in Americas. The clade L is a reference one represented by most sequences from Asia [12]. More than 70.0% world's sequences were covered by clade G (GH and GR). However, GR represents both D614G spike and RG203KR nucleocapsid mutations worldwide. Clade L accounts for 7% genome sequences and the other clades S and V have the similar frequencies globally.

Antara Sengupta has analyzed the isolates from ten different countries around Asia. The circulating clades were found to be G, GH, GR, L, S, O and V. The clade GR and GH have the highest number of SNPs (Figure 7). Most deleterious mutations occurred in isolates belong to clade GH is described in mutational pattern of this manuscript. They collected the DNA sequences from NCBI to compared with GISAID mutated sequences [14].

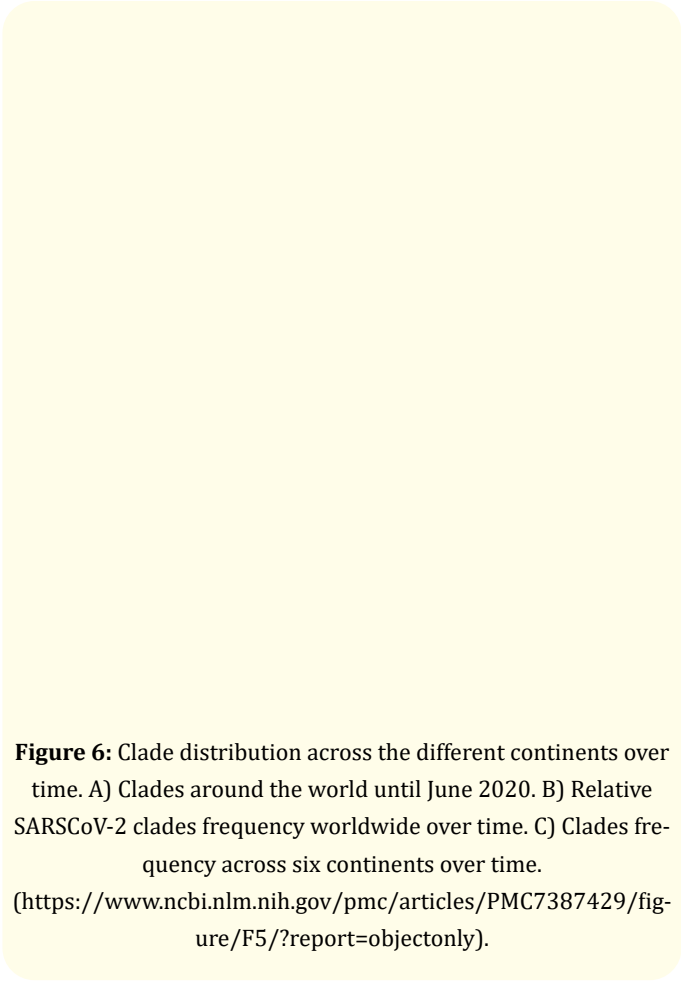

Figure 6: Clade distribution across the different continents over time. A) Clades around the world until June 2020. B) Relative SARSCoV-2 clades frequency worldwide over time. C) Clades frequency across six continents over time. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7387429/figure/F5/?report=objectonly).

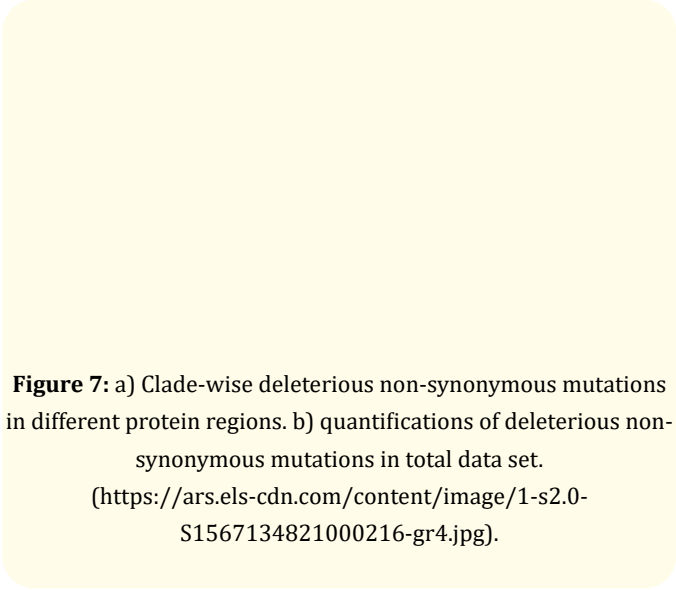

Figure 7: a) Clade-wise deleterious non-synonymous mutations in different protein regions. b) quantifications of deleterious non-synonymous mutations in total data set. (https://ars.els-cdn.com/content/image/1-s2.0-S1567134821000216-gr4.jpg).

The structural changes in amino acids caused the instability of isolates, which increased the magnitude of virulence and vulnerability to the disease. Changes in the Gibbs free energy due to single point mutations has taken as an impact factor for protein's function and stability. It is observed that the structural protein mutations on the spike are more deleterious mutations than others, analyzed by the score obtained from PROVEAN (Protein Variation Effect Analyzer). The accessary protein NSP3 is most affected. NSP2, NSP5, and NSP12 are the non-structural proteins with occurred deleterious mutations (Figure 12). The deleterious mutations in structural and accessary proteins belong to clade GH. Deleterious mutations in the non-structural proteins occurred in clade GH and GR [14,15].

Clades associated with severe disease in US, France, Italy, and Brazil had low frequencies in Africa [33].

A phylogeographic analysis have shown that the chain transmission originated from South American countries in Late February, 2020. Genetic analysis suggested the dominance of S and G clades (G, GH, GR). A lethal outcome of SARS-CoV-2 infection significantly correlated with arterial hypertension, kidney failure, and ICU admission (FDR < 0.01), but not with any mutation in a structural or non-structural protein, such as the spike D614G mutation. The genetic, phylodynamic, and clinical correlation data interpreted the disease patterns and regional distribution in the pandemic of Latin America [16].

Characteristics and Global Distribution of D614G:

Viral mutations have selective advantage to increase the transmission rate and attaining resistance to the current interventions. Mutations are evaluated for positive selection; by exploring their implications through structural modelling. Virus began to spread in Europe in Feb 2020 and the D614G spike mutation was found to be of urgent concern. Kober et al, has explained the recombination between local circulating strains are the indicative of multiple strain infections in a host. She has highlighted the implications of D614G form of SARSCoV-2 on transmission, pathogenesis and immune intervention therapy.

Evaluation of spike variants in pandemic is a prognostic factor in a dynamically changing pattern in the spike genes. Recombination is significant in SARS evolution. GISAID data was assessed to analyze the structural modelling of the sites of interest, and their experimental evolution. The positive selection made the D614G form dominant in many countries [17].

Figure 8: Phylogenetic tree prepared by Kober, 2020, by aligning 4535 genome sequences obtained from GISAID https://www.biorxiv.org/content/biorxiv/early/2020/04/30/2020.04.29.069054/F1.large.jpg.

The three base changes have defined the clade G for D614G mutation. In early April, G614 was more common than the original D614 isolated from Wuhan, and it started spreading around the globe rather than being restricted to Europe. The phylogenetic tree (Figure 8) has shown that the mutations predominantly appeared as a part of single lineage were P1263L (Orange in UK and Australia), and A831V (red in Iceland). A rare mutation L5F originated separately and scattered around the places. A mutation in sequences S943P from the same geographic location but arising in distinct lineages in the phylogeny was only observed in Belgium. The relative frequency appeared in Wuhan have changed into mutation at G614 in Europe [17,18]. The mutations were tracked in the phylogenetic tree by informed evolutionary trails in the full genome.

The observed signal for shifts in positive selection of mutational frequencies are as below:

- Increased frequency of sequences for a specific mutation.
- Recurrent identification of specific mutation in different geographic regions over time.
- Different codons to encode the same recurrent amino acid.
- Tight cluster of mutations in linear structural sequence space.

The trail of evolution in spike protein has been figured out for its impacts on the structural modelling to learn about its binding sites, trimer stability, and glycosylation patterns. Mutations on the spike with antibody epitopes along with RBD receptor binding domain are of special concern. The mutation D614G is located at 4 Å site in the 3D structure of spike. The calibrated site is allowed for further characterizations e.g., experimental evolution, studies on antigenicity, neutralization, sensitivity and capacity to bind to ACE2. Total sites of interest included 14 positions as of April 13th 2020 and one local cluster mutations. The spike mutations of particular interest, D614G and S943P, are at same length. D614G is increasing in frequency at the alarming rate for the rapid spread. Mutation S943P is present in the fusion core region have potential for recombination to enable the spreading.

The spotted mutations were mainly in the N-terminal domain, S2 membrane fusion subunit and HR1 first heptad regions of the spike. Up/down conformations refer to a change in state in which the up conformation is exposed the RBD (Table 3). In the first epidemic with SARS CoV; the immunodominant linear antibody epitope observed in natural infection and in animal models [19].

The other mutations remained silent as compared to D614G. Nevertheless, efforts are being done to monitor their immunological impact and increased frequency regionally and globally in the pandemic.

D614G mutation have G-to-A change at position 23403 first discovered. As March 2020 began, the D614 was noticed to be dominated, but changed in to G614 in a matter of few weeks in April. In Europe, D614 and G614 forms were circulating in the early epidemic, but in Canada and USA the infections were established with G614 forms dominating. In Iceland, the epidemic started with G614 form and persisted at constant level. Asian samples have D614 form dominated, which was expanded into G614 form later on. This is the example of dramatic shifts of selective advantage over a period of several weeks. In Germany C-to-T mutation at 3037 was discovered in the beginning, but not at 14409 position. These findings have provided the clue that D614G mutations may have arisen independently [17].

The genomic diversity from the samples obtained throughout New Zealand presented the global viral population. The genomes were aligned with residue at spike protein have shown SD614 (24% Aspartic acid) and SG 614 (>70% Glycine). Preliminary studies conducted on cell culture, phylodynamic approaches to increase in the glycine were giving the indication that D614G mutation can enhance the viral infectivity. Notably. Increase in the glycine reported in imported samples in New Zealand, rather than selection of mutation with in the country [45].

D614G- The most competent form

On structural basis

D614 is located on the surface of spike protein promoter where it can form contacts with neighboring promoter. A cryo-EM structure indicates that the side chain of D614 can potentially form hydrogen bond with T859 of the neighboring promoter. The strength of Aspartic acid- Threonine hydrogen bonding as the promoter-promoter hydrogen bonding may be of critical importance to bring together the S1 of one promoter to S2 units of neighbor promoter. The protein brackets the dibasic furin- and S2-cleavage sites; hence it is possible that D614G mutation diminishes the interaction between S1 and S2 units, facilitating the shedding of S1 from viral membrane bound S2. An alternative structure-based hypothesis is that this mutation may impact RBD-ACE-2 binding. RBD is in 'up' position to engage with ACE-2 receptor, it is possible to allosterically alter the transition 'up' and 'down'. In known SARSCoV-2 spike protein with all RBD structure in down position, the distance between D614 and T-859 remains the same. However, in one structure with RBD in 'up' position the distance between neighboring promoters is longer than the rest of the promoters. These slight changes are within the conformational fluctuations of a dynamic spike trimer. More studies are required to accomplish the entire effect of this mutation on RBD transitions [20,21,23].

Generated immune response by D614 mutation

D614 mutation could have its immunologic impacts. The immunodominant peptide of SARS CoV spike epitope (encoded s597-603), has high level (64%) of serologic reactivity, and induced long term B-cell memory response in convalescent sera obtained from individuals who were infected during the previous SARS CoV in 2002 epidemic. Antibodies against this peptide mediate antibody-dependent enhancement (ADE) of SARS CoV infection using epitope-sequence dependent mechanism, both in vitro and in vivo in rhesus macaques [5].

The core epitope for ADE mediating Abs in SARS CoV was LYQDVNC (SARS CoV S597-603), and this peptide is immediately proximal to the peptide that targets the potent neutralizing antibodies (SARS CoV S604-625). The ADE target peptide spans the SARS CoV-2 site, and is identical to the equivalent region in SARS CoV is S611-617. Wang et al, discovered that the proximity of this epitope to the RBD speculated the bonding of Abs may mediate the conformational change in spike that enhance RBD-ACE-2 interaction resulting in severe effect. This process is different from the mechanisms of alteration of Fc receptor domain to cause ADE reported by other authors which could occur in presence and absence of ACE-2 [5,22,23].

Hence, D614G mutation may increase the spike's infectivity by several ways. This could be due to improved receptor binding, fu-

Figure 9: Structural mapping key mutational sites in the Spike protein. A) Spike protein with structural S1 and S2 units with mutational sites. The RBD promoter is in 'UP' position to engage binding with ACE2 receptor. Mutational site balls are in red color. B) Mutational sites near the RBD (blue)-ACE2 (yellow) binding interface. The interfacial region is shown as a surface. C) The proximity of D614 to T859 from the neighboring protomer. The white dashed lines indicate the possibility for forming hydrogen bonds. D) The schematic representation of potential protomer-protomer interactions shown in (C) brings together D614 from S1 unit of one protomer to the T859 from S2 unit of the neighboring protomer. E) Cluster of mutations, S937-S943, in the HR1 region (Heptad repeat region) of the Spike protein. These residues occur in a region that undergoes conformational transition during fusion. The left and right images show the pre-fusion and post-fusion conformations of this HR1 region. (https://www.biorxiv.org/content/biorxiv/early/2020/04/30/2020.04.29.069054/F4.large.jpg).

sion activation, or the development of ADE antibodies. Shifting to G614 simply implies the antigenic drift mediating antibody escape. The current vaccines showed effective antibody immune response against the virus harboring G614 mutation.

D614G mutation and its impact on the severity of disease

The pattern of mutation observed in, Sheffield Teaching Hospital NHS Foundation Trust, SARSCoV-2 sequences generated from 453 individuals starting out with D614, and shifting predominantly G614 by the end of March. Sheffield data included age, gender, data of sampling, cycle threshold (CT) for positive signal in E gene-based RT-PCR. There was, however, no significant correlation found between D614G status and hospitalization status. Although the G614 mutation was slightly enriched among the ICU subjects [17,24]. Consequently, D614G status didn't contribute to modelling hospitalization as an outcome, except there was a marginally significant interaction with PCR CT ($p = 0.04$).

While D614G didn't predict hospitalization, but with fewer threshold of PCR cycles being required for G614 as compared to D614 (Figure 10). This also provides a signal that patients carrying G614 mutation had higher viral loads. There was uncertainty in regards to the time of infection at which samples were taken, and in fact the indirect PCR measured the viral load. Despite observing these drawbacks, there was still a clear-cut observation shifting the D614 to G614 [17].

Figure 10: Clinical study outcome in Sheffield, England on D614 and G614. A) G614 is overtaking D614 in the course of epidemic B) Age distribution of people visiting hospitals is similar between the two forms i.e. (D 614- Median age of 60 (IR 44-80)) (G614 had the median age of 59(IR 43-77)). C) D614G status was not statistically associated with hospitalization status. D) G614 was associated with fewer rounds of PCR required for detection, indicating that people infected with G614 virus had higher viral loads. (https://www.biorxiv.org/content/biorxiv/early/2020/04/30/2020.04.29.069054/F5.large.jpg).

Susceptibility of D614G to neutralization

Drew Weissman demonstrated that vaccinated mice, non-human primates, and humans using the nucleoside-modified mRNA-LNP vaccine platform encoding four different SARS-CoV-2 spike immunogens generate antibody responses that not only recognize the G614 mutation, but also gained the stronger titers of neutral-

ization to this virus variant. The mechanism appears to be that the mutation increases the up formation of the RBD in the spike trimer, increasing the exposure of neutralization epitopes. Most of the vaccine immunogens are used in the clinical trials, either derived from the initial D614 virus or contain D614G in the spike. Although the G614 variant has replaced the original D614 sequence in the SARS-CoV-2 spike throughout much of the world. But it is discovered that this is not an escape mutation and, in fact it is better neutralized by sera from mice, NHPs, and humans immunized with vaccines derived from the D614 viral spike. This mutation alleviated a major concern regarding the current efforts to develop an effective SARS CoV-2 vaccine [11,17,27].

The studies were conducted during this time as below:

- The four variations studied in the spike immunogen against one vaccine platform, nucleoside modified mRNA-LNP platform [27,28]. Moderna and Pfizer/BioNTech had conducted successful clinical trials to demonstrate the safety and neutralizing response [29].
- High significant correlations were observed using pseudovirus neutralization assay for both the convalescent plasma and human mAbs. These tests are typically performed in the cell lines, natural target cells of the respiratory system are technically challenging, low throughput, and difficult to standardize.
- This study was performed in furin cleavage -deficient spike ectodomain, mimic the native spike with RBD 'up' conformation between D614 and G614 producing an allosteric effect of the D614 mutation on the RBD conformation. However, these structures are still different from the native spike.
- Incorporation of spike is another mechanism for increased neutralization -susceptibility of D614G spike pseudovirus and susceptibility cell host.

Observed mutational patterns

The GH clade received novel mutations at NSP2 (C1059T), which is a non-recurring mutation from majority of sequences. Apart from D614G mutation observed in the spike protein, the 2nd most common amino acid changing mutation P314L, affecting the structural protein 12 (NSP 12), the viral RdRp. The other two mutations don't seem to affect the protein sequence, as they are silent mutations targeting the 106th codon of NSP3 (a viral predicted phosphoesterase) and the 5'UTR in position 241(Table 2) [6,9].

Table 2: Highlights of 20 most frequent mutational events observed worldwide in sequence of SARSCoV-2 genomes (Daniele., *et al*. 2020) (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7387429/table/T1/?report=objectonly).

S.no	Genomic Coordinate Mutation	Effect of protein/ UTR	Class	Region in the Genome
1	A23403G	S: D614G	aa-changing SNP	Spike protein
2	C14408T	NSP12b: P314L	aa-changing SNP	Non-structural protein 12, post-ribosomal frameshift (RNA-dependent RNA polymerase)
3	C3037T	NSP3:F106F	silent SNP	Non-structural protein 3 (predicted phosphoesterase)
4	C241T	5'UTR:241	5'UTR SNP	5' UnTranslated Region
5	GGG28881AAC	N: RG203KR	aa-changing SNP triplet	Nucleocapsid protein
6	G25563T	ORF3a: Q57H	aa-changing SNP	ORF3a protein
7	C1059T	NSP2:T85I	aa-changing SNP	Non-structural protein 2
8	G11083T	NSP6:L37F	aa-changing SNP	Non-structural protein 6 (transmembrane protein)
9	C14805T	NSP12b: Y446Y	silent SNP	Non-structural protein 12, post-ribosomal frameshift (RNA-dependent RNA polymerase)
10	T28144C	ORF8:L84S	aa-changing SNP	ORF8 protein
11	G26144T	ORF3a: G251V	aa-changing SNP	ORF3a protein
12	C8782T	NSP4:S76S	silent SNP	Non-structural protein 4
13	A20268G	NSP15:L216L	silent SNP	Non-structural protein 15 (endoRNAse)
14	C18060T	NSP14:L7L	silent SNP	Non-structural protein 14 (3'-5' exonuclease)
15	C23731T	S: T723T	silent SNP	Spike protein
16	G10097A	NSP5:G15S	aa-changing SNP	Non-structural protein 5 (protease)
17	A17858G	NSP13:Y541C	aa-changing SNP	Non-structural protein 13
18	C17747T	NSP13:P504L	aa-changing SNP	Non-structural protein 13
19	C2558T	NSP2:P585S	aa-changing SNP	Non-structural protein 2
20	A2480G	NSP2:I559V	aa-changing SNP	Non-structural protein 2

aa- stands for amino acids

NSP- Non-structural proteins

The nucleocapsid protein N: RG203KR is affected by trinucleotide mutation and 'aa' amino acid change. G15S change in viral protease NSP5, which instigate the protease activity and folding, is the worldwide event occurred 3.7%. Although the mutational changes in NSP2, NSP6 and NSP13, ORF3a, ORF8 are not fully comprehended and explained [6,11].

Antara Sengupta observed the changes in amino acids, due to change in Gibbs free energy at single point mutation. This process resulted with great impact on functionality and stability of proteins. The accessary protein NSP3 is most affected. NSP2, NSP5 and NSP12 have shown the deleterious mutations could be the cause of increased virulence (Figure 7) [14,15].

The detailed phylogenetic analysis has shown that the spike mutations were maximally related to D614G. Nevertheless, the other spike mutational events were noticed to be 1% only. One of the spike mutation N439K present in the spike and ACE2 interaction domain accounted for 0.7% of the viruses. The envelop protein appears to be more conserved, but have some more frequent mutations in C-terminus observed in around 0.2% population. Membrane protein has presented T175M mutation besides RG203KR mutation which characterizes the GR clade. The several non-silent mutations 1%, appeared particularly as P13L, D103Y, S194L, and S197L (Figure 11).

Table 3: Notable spike mutations [26] (Kober B., *et al*. 2020) (https://www.biorxiv.org/content/10.1101/2020.04.29.069054v1.full#T1).

No. of Mutations 13 April, 20	Spike Mutation	Possible Impact (Spike Location)	Geographic Sampling	Phylogenetic Pattern
3577	D614G	SARS-CoV epitope Interprotomer stabilization	Global	One main lineage and recurrent emergence
37	L5F	Signal Peptide	13 Countries	Recurrent emergence
18	L8V/W	Signal Peptide	Hong Kong	
12	H49Y	S1 NTD domain	China	
10	Y145H/del	S1 NTD domain	6 Countries	
8	Q239K	S1 NTD domain	Europe	
12	V367F	Up/Down conformations	Europe/Hong Kong	
8	G476S	Directly in RBD	Washington, USA	
21	V483A	Up/Down conformations	Washington, USA	
13	V615I/F	In SARS-CoV ADE epitope	Wales	
28	A831V	Potential fusion peptide in S2	Iceland	One lineage
27	D839Y/N/E	S2 subunit	Europe	
23	S943P	Fusion core of HR1	Belgium	Local recombination
49	P1263L	Cytoplasmic Tail	UK/Iceland/Australia	One lineage
65	Mutational Cluster 937-943	Fusion core of HR1	Different forms in 9 countries	

Sites L5F and L8V

The L5F mutation is attractive because of its recurrence in many lineages throughout the SARS CoV-2 phylogenetic tree, and in many different countries throughout the world (Fig 8). It is found to be established independently, regionally transmitted and contributed to the multiple small local clusters. For example, cluster of 5 infections in Iceland carrying identical 5F, and several comparable clusters in different states of USA. It has maintained the frequency of 0.6% globally with no increment in frequency, despite its recurrences [17].

L8V mutation mostly found in single lineage in Hong Kong with declined global frequency.

Sites V367F, G476S and V483A

V367F, G476S and V483A are found in RBD domain (Table 3). G476S is at the end of the binding interface of RBD and peptidase

domain of ACE2. The Q24 residue of ACE2 closest to G476. The G476S mutation may contribute to the rigidification of the local loop region in RBD near the binding interface. The mutation at site V483 doesn't directly contact ACE2 although it is on the same face of RBD binds to ACE2 receptor. Site V367F is at the opposite end of RBD binds to ACE2 receptor; It is on the same face as epitope CR3022 (a neutralizing Ab isolated from convalescent sera), when two RBD regions of the spike trimer are 'up'. Although there is no direct contact between V367 and CR3022 are observed [17, 25, 26].

Sites V367F and G476 were identified in early mutations appears to be diminishing in overall global frequency in later samples. V367F research continued as its potential interactions with ACE2.The mutation V483A has predominately appeared in Washington State. This mutation maintains the low presence and not increasing overtime [17].

Sites H49Y, Y145H/del, Q239K

These mutations are located in the S1 N-terminal domain (NTD) with low frequency, and appeared to be diminishing later on. They recur in different countries, especially K239 is found predominantly in Netherlands [17].

Site A831V and D839Y/N/E

These are located in S2 near the potential fusion peptide. A831V is found in Iceland and in a single lineage that is a stable frequency over time.

Site P1263L

This mutation is not included in SARSCoV-2 structure and it is present near the cytoplasmic tail of the spike protein. It is maintaining its frequency both globally as well as locally in the UK [17].

Sub Saharan Africa (SSA) accounted for 3-4% in P323L and Q57H mutations in Africa. Higher disease pressure occurred in countries viz., South Africa, Tunisia, Morocco and Egypt with higher D614G frequencies (52-98%). Other mutations like G476S, D364Y, V367F and V483A are normally associated with ACE2 receptor binding and severe disease have not been observed in Africa. Most of the RdRp mutations were deaminations leading to CpG depletion and possible attenuation of virulence [33].

The dominating variant in Australia was lineage B.1.1.25 on January 2020. The mutation S477N was prevalent in Australia, which occurred separately [46].

Prediction of Secondary Structure of SARSCoV-2 5'UTR (nt1-265)' WT (wild type) and variant C241T

The protein sequences were not found to be altered in C241T mutation spotted in SARSCoV-2 5'UTR. There was not any significant difference, because of no participation in forming hydrogen bonds with other nucleotide in the secondary structure of wild type (WT) genome and the C241T variant (Fig11). These sequences could bring the consequences on the secondary RNA structure due to the minimum free energy prediction, by influencing the RNA replication and speed of infection [13].

Figure 11: A) Silent (blue) and protein sequence affecting mutations(red) occurred in four structural proteins of SARSCoV-2 i.e., S (spike), E (envelop), M (membrane), and N (nucleocapsid). On the x-axis amino acid coordinate of the mutation and on the y-axis, the log 10 number of samples (maximum limit is 48,635 samples in dotted line). The top 5 a-a changing mutations are indicated in spike protein (D614G, L5F, N439K, D936Y, P1263L). B) Minimum free energy prediction of the secondary structure of SARSCoV-2 5'UTR (nt1-265)' WT (wild type) and variant C241T. Positional entropy express the base reliability, is presented in colors. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7387429/figure/F6/?report=objectonly).

Recombination among pandemic strains

Recombination is another evolution phenomenon plays a critical role in SARS CoVs. The current standard methods couldn't detect these changes in the pandemic samples. When two viral strains (carrying different mutations) infect the same host at a time, could have the ability to generate recombined sequences. This theory is mainly implied to the local geographical population having co-infection occurs with two different strains. The computational RAPR has been used to detect the potential recombinational events in local geographical area [17].

The original Wuhan D614 form carried out the bases 'C-C-G' at this position, and the G614 form was discovered to carry the bases 'T-T-A". These two forms co-circulating in many communities over the month of March; hence 'C-C-G' or 'T-T-A' patterns was the first indication of recombination.

Kober et al, pinpointed the relevant mutations, by comparing the three sample sets and run-length statistics for significance in every set. The pattern shared sites have shown serial spontaneous mutations/ or recombination in vivo or possible recombination in vitro could be the result of contamination during PCR. The sequenced samples were defined by three mutational events. These has been obtained from areas with co-circulation of both haplotypes of D614G mutation. The C- to -T mutation at position 3037, a C-to -T mutation at position 14408, and the G- to- A base change at position 23403 that gives rise to the D614G amino acid change in spike (Figure 12) [17].

SARSCoV-2 sequences were analyzed from Washington State, Iceland and Netherlands [17].

Figure 12: Recombination events detected in the three different regions, Washington State, Netherlands and Iceland. Black arrows show the D614G trio of mutations at positions 3037, 14408 and 23403 (D614G in the spike gene). The p-values are based on a run-length statistic and are corrected for three datasets. RAPR software is used to detect the recombination events.

(https://www.biorxiv.org/content/biorxiv/early/2020/05/05/2020.04.29.069054/F6.large.jpg)

Variants of SARSCoV-2

The highly transmissible and lethal mutations alarmed to enter the new phases of pandemic of SARSCoV-2. They are generated as a consequence of recombination of strains in a highly mutated hybrid forms e.g., B.1.1.7 and B.1.429 variants.

The first identified C lineage, C.1, of SARS CoV-2 had 16 mutations as compared to Wuhan sequence; found geographically widespread lineage in South Africa. Despite causing infection in multiple provinces, the B.1.106 lineage (identified in April 20) became extinct after successful control of the nosocomial outbreak [34].

86 of the 120 variants associated with immune escape were noticed in a total of 26,917 genomes from 63 countries, reported by Jolly B. et al. 19 out of 86 genetic variants were found in genomes from India. The S: N440K variant was observed to have a frequency of 2.1% in India and high prevalence in state of Andhra Pradesh (33% of 272 genomes). The variant site was homoplastic, belonging to the different clades and haplotypes. The time scale analysis suggested that the variant emerged; in recent months; reported to have more cases of re-infection in North India [54].

Lineage B.1.1.7

The lineage B.1.1.7 or 201/501Y.V1 designated as 'variant of concern' (VOC-202012/01). Its transmissibility rate is increased by 40-80% than the Wuhan strain. All the genomic sequencing was carried out in UK as of January 2021 [49].

The genetic information of altered S-gene target failure (SGTF) is received by a community-based PCR diagnostic test. The pattern of SGFT and non-SGFT have revealed the higher transmissibility of 'VOC' than 'non-VOC lineages with estimated ratio of reproduction numbers varying between 1.4 to 1.8. This resulted in applying the higher level of social distancing in England [31].

The lineage 501Y, a mutation in RBD, without amino acid deletion Δ69/Δ70 in circulation from early Sept. to mid Nov. 2020 was

10% (6-13%) transmissible than 501N lineage. The dominant form 501Y lineage with amino acid deletion Δ69/Δ70 (HV 69-70 deletion- a deletion of amino acid histidine and valine at positions 69 and 70 respectively) circulating since late Sept. 2020, was found to be more transmissible 75% (70-80%) than 501N lineage [32].

The genome sequencing. B.1.1.7 represented the 23 mutations: 14 non synonymous mutations, 3 deletions, and 6 synonymous mutations. There are 17 mutations that change the proteins and six that don't. Public health also reported limited number of E484K mutations in Feb 2021, which they named as 'Variant of Concern'(VOC-202102/02). This mutation is largely present in Brazil and South Africa, making 50-70% more transmissible and may reduce the vaccine's effectiveness [49].

The critical change N501Y in B.1.1.7 lineage, represents a change from Aspargine 'N' to Tyrosine 'Y' in amino acid position at 501. This position is inside the RBD, more precisely inside the RBM, could lead to increase the virulence. This mutation with the deletion at 69/70 in the N-terminal domain (NTD) is appeared to have the two-fold higher infectivity over a single round of infection compared to SARS CoV-2 pseudovirus [31,37].

Table 4: Investigation of novel SARSCoV-2 'Variant of Concern' 202012/01 on the basis of nucleotides and amino acids change (Chand et al, 2020) (https://en.wikipedia.org/wiki/Lineage_B.1.1.7).

Mutational Profile of 501Y.V1		
Gene	**Nucleotide**	**Amino acid**
ORF1ab	C3267T	T1001I
	C5388A	A1708D
	T6954C	I2230T
	11288-11296 deletion	SGF 3675-3677 deletion
Spike	21765-21770 deletion	HV 69-70 deletion
	21991-21993 deletion	Y144 deletion
	A23063T	N501Y
	C23271A	A570D
	C23604A	P681H
	C23709T	T716I
	T24506G	S982A
	G24914C	D1118H
ORF8	C27972T	Q27stop
	G28048T	R52I
	A28111G	Y73C
N	28280 GAT→CTA	D3L
	C28977T	S235F

People infected with B.1.1.7 variant are more likely to get admitted to the hospitals as compared with people infected by non-B.1.1.7 SARSCoV-2 lineage.

Mutation Q27 cut off the ORF8 protein and render this inactive. The variants with deleted ORF8 gene are reported to have the milder symptoms and better disease outcomes. ORF8 protein is like an immunoglobulin (Ig) which modulates the pathogenesis and mediates the major histocompatibility complex I (MHC-I) degradation. It also suppresses the type I interferon (IFN)-mediated innate antiviral response.

Chand *et al*, concluded that the amino acid position 501 inside the spike protein affect the neutralization of the virus. Although, there is no neutralization data on N501Y available from polyclonal sera from natural infection. The HV 69-70 deletion has been associated with RBD changes and observed to elude the immune response in some immunocompromised patients. Nonetheless, rapid antigen and vaccine effectiveness is still working against the Variant.

This variant is reported in many countries. As of Feb 28, it become the dominant COVID-19 variant for 12 countries: UK, Ireland, Slovakia, Israel, Luxembourg, Netherlands, Portugal, Denmark, Norway, Italy, Belgium and France [49].

Lineage B.1.351

The South Africa variant B.1.351lineage also known as 501Y.V2/or 20H/501Y.V2/ or 20C/501Y.V2. It was first detected in Eastern Cape province of South Africa on 18 December, 2020 [50].

501Y.V2 is characterized by eight lineage mutations in the spike protein, including three at RBD e.g., K417N, E484K and N501Y, could have functional significance. The importance of this lineage is associate with rapid transmission of the virus [34]. This strain was predicted to become the predominant strain by March in US, a modelling provided by CDC US. Some of the mutations viz. E484K and K417N, changed the surface protein of spike, and shown the reduce the neutralizing response by monoclonal antibodies in the lab. The potency of convalescent sera was found to be 10-fold lesser. But it doesn't necessarily mean that the mutation would cause people's immunity by new strain to drop 10-fold [47].

The first three mutations, K417N, E484K and N501Y, are in the RBD help variants to attach human cells easily. The two mutations N501Y (a change from Aspargine (N) to Tyrosin (Y) in amino acid position 501), and E484K are within the receptor-binding motif (RBM) of the receptor-binding domain (RBD). The E484K amino acid change in RBD possibly alter the antigenicity and referred as the escape mutation. It is also linked with cases of re-infection with the 501.V2 variant.

Table 5: Mutational profile of 501Y.V2 on the basis of nucleotides in the gene and amino acids (Tegally H., *et al*. 2020) (https://en.wikipedia.org/wiki/501.V2_variant).

Mutational profile of 501Y.V2		
Gene	**Nucleotide**	**Amino Acids**
ORF 1ab	C1059T	T265I
	G5230T	K1655N
	C8660T	H2799Y
	C8964T	S2900L
	A10323G	K3353R
	G13843T	D4527Y
	C14408T[1]	P4715L
	C17999T	T5912I
Spike	C21614T	L18F
	A21801C	D80A
	A22206G	D215G
	G22299T	R246I
	G22813T	K417N
	G23012A	E484K
	A23063T	N501Y
	A23403G[1]	D614G
	G23664T	A701V
ORF 3a	G25563T	Q57H
	C25904T	S171L
E	C26456T	P71L
N	C28887T	T205I

The other five spike mutations L18F, D80A, D215G, R246I, and A701V were of less concern. Indel is also represented by insertion of K1655N, SGF3675-3677 and deletion of P71L, and T205I, away from the spike [34].

The sequencing of the whole genome of the new variant was carried out at the University of KwaZulu-Natal Research Innovation, followed by the submission of the same to GISAID.

This variant is reported to become prevalent more in young people with no underlying conditions, and responsible to cause more frequent serious illness.

Pfizer and BioNTech (BNT162b2) mRNA vaccine have shown slightly less effectiveness against 501.V2 variant mutations. Johnson and Johnson (Ad26.COV2. S) adeno vectored vaccine trials in South Africa, reported the level of protection 57%. But the same vaccine provided 72% efficacy in US. Novavax (NVX-CoV2373) protein-based vaccine is 60% effective (for HIV negative participants) in South Africa compared to 90% efficacy in Britain. AstraZeneca (AZD1222) adeno-vectored vaccine was found to provide minimal protection in severe cases of COVID-19. Pfizer and BioNTech vaccine produced ~66% antibodies to combat the variant B.1.351 than Wuhan strain, but still be successful to neutralize the virus in lab [50].

As of 5 march 2021, the cases with 501Y.V2 confirmed in countries are South Africa, United Kingdom, Switzerland, Finland, Japan, Australia, Zambia, France, South Korea, Sweden, Norway, Botswana, china, Ireland, The Netherlands, Canada, Israel, New Zealand, Germany, Belgium, Taiwan, Denmark, Spain, Kenya, Portugal, Panama, UAE, Comoros, Mayotte, USA, Mozambique, Vietnam, Luxembourg, Turkey, Austria, Bangladesh, Ghana, Greece, Malta, Singapore, Thailand, Reunion, Italy, Rwanda, Costa Rica, Slovenia, Philippines, Croatia, India and Malawi [50].

Lineage B.1.128.1/P.1

Variant P.1/ B.1.128.1 also denoted as 20J/501Y.V. It is also named as 'variant of concern' 202101/02(VOC-202101/02). This variant has 17 unique amino acid changes, 10 of which are in its spike protein. The three mutations detected are of specific concern N501Y, E484K, and K417T. This variant is detected first in Japan on 6th January 2021. It was declared to be circulated in Brazil [39,51].

Another P.2 lineage initiated from Rio de Janeiro, also named as B.1.128.2 or VUI-202101/01. It carries E484K mutations only.

Table 6: Mutations found in lineage P.1 (Faria., *et al*. 2021) (https://en.wikipedia.org/wiki/Lineage_P.1).

Mutations in the Lineage P.1	
Gene	**Amino Acid**
ORF Iab	synT733C
	synC2749T
	S1188L
	K1795Q
	del11288-11296 (3675-3677 SGF)
	synC12778T
	synC13860T
	E5665D
Spike	L18F
	T20N
	P26S
	D138Y
	R190S
	K417T
	E484K
	N501Y
	H 655Y
	T1027I
ORF 8	E92K
	ins28269-28273
N	P80R

ORF 1a and 1b have the eight mutations, while spike protein has 10 including N501Y and E484K. It has 2 mutations with one insertion in ORF8 region and one in N region.

As of 5 March 2021, cases confirmed in countries are: Japan, Brazil, South Korea, Faroe Islands, Denmark, Germany, Italy, Ireland, USA, Peru, The Netherlands, Colombia, Croatia, turkey, France, Canada, Argentina, Belgium, French Guiana, Spain Switzerland, Mexico, Sweden, United Kingdome and India [51].

Lineage B.1.427/ B.1.429

Scientists have identified many more variants, but there is still uncertainty over which of these mutations are relevant and to what extent. Variant identified in California has emerged as a "variant of concern" in US and named as B.1.427/ B.1.429. The country emerged from devastating winter wave of infections, hospitalizations and deaths [40].

A mutation is known as L452R in the variant B.1.427/ B.1.429 give advantage in binding to receptors in human cells. Researchers have identified another mutation in the virus called Q677, which initially was detected in Louisiana and New Mexico. It has emerged independently in seven variants could be related to "convergent evolution." The nasopharyngeal viral loads obtained from B.1.427/ B.1.429 infected persons found to be significantly higher and the virus infecting the lung cells in the lab more efficiently [40].

Lineage B.1.617

A spike in COVID-19 cases were reported more than 200,000 per day. This surge is mainly caused by a new double mutant discovered from 200 samples received for genome sequencing. The double mutant is classified as lineage B.1.617 could be the 'Variant of Concern'. This double mutation refers to the specific changes are denoted by E484Q (glutamate is replaced by glutamine at the 484th point of spike protein) and L452R (substitution of leucine with arginine at the 452nd position).

E484Q mutation is similar to mutation E484K found in lineage B.1.1.7 and lineage B.1.351. The mutation L452R was responsible for more spreading of the variant in California B.1.427 and B.1.429. These both mutations strengthen the binding of spike to ACE2 receptors. The mutation L452R effectively enhance the viral replication. These mutations could evade antibodies.

Most of the patients reported to be asymptomatic with this variant B.1.617. The long- term goal to tackle the COVID-19 pandemic is for people to get vaccinated [52].

Discussion

Quasispecies are missing in the evolutionary events studied by Daniele et al, 2020; it was impossible to analyse the multiple evolutions in the subpopulations with in same patient. Therefore, the actual mutation rate could be much higher than 7.23. The evolutionary mRNA editing polypeptide enzymes (apolipoprotein B) by APOBEC mechanism help generating the protein diversity from C to U editing. These evolutionary mRNA editing enzymes could be hypothesized for the transitions observed in mutational events [6].

There are five distinct clades of SARSCoV-2 exist currently, could be increased further in future.

The most common mutational event is 'aa'-changing SNPs in 5'UTR. Due to the repercussions of silent events may change the codon usage and translation efficiency. 5'UTR SNP mutations may affect the transcription and replication /or folding of genomic ssRNA of the virus. This process has been partially elucidated.

Transcriptome dynamics of SARSCoV-2 suggest the mechanism of mutation by single nucleotide transition, which is associated with RdRp by which RNA editing triggered by the host cell defense mechanism. The largest indel reported is 80 nucleotide deletion (Arizona sample) in the genome [6].

Recombination among the pandemic SARSCoV-2 strains is not surprising, it is also found among more distant coronaviruses with higher diversity level. The SARSCoV-2 virus evolves continuously to emerge in mutated forms, with clinical and pharmacological repercussions. D614 mutation is likely to be associated with the higher fatality rate. The other important strategy is that combining the details of epidemiological information and clinical features of COVID-19 patients, could be very useful to identify the new therapeutics to minimize the burden of the disease [6,43].

The emerging new mutations and recombination of two viral strains could enforce to develop new antiviral therapies considering multiple recombinants. Especially the diversity in the spike protein must be taken in to account. Main limitations of emerging new recombinational strains are:

- Recombination can't be detected without simultaneous coinfection of distinct viruses in one host. It doesn't give the clear picture whether it is happening prior to the development of adaptive immune response, or in a series of events of re-infection occurred after the first one stimulated the response initially.
- Recombination could be more common in communities where less stringent preventive measures are applied, especially in less stringent patient isolation in hospitals where all the pa-

tients assumed to be infected in same region. In those places, the antigenic drift could begin to enable the serial infections to produce more resistant form of the virus.

- Recombination process provides more opportunities for the virus to make the recombinant forms and by multiple mutations to confer distinct fitness to thrive independently, but that could be carried out separately in the two parental strains.

The prevalent mutation D614G doesn't seem to affect the interaction domain with ACE2, responsible for the virus entry into epithelial cells [5]. There are some more mutations have been noticed in that area as N439K present in 0.7% of the sequence. The new mutations and clades could confer the evolutionary advantage to SARSCoV-2. It is pivotal to monitor the mutations constantly to track the movement of the virus across individuals and geographic area. One of such tools is NextStrain, which allows for scalable phylogenetic analyses and real time monitoring.

SARS CoV-2 is continuously changing its characteristics and degree of infectivity. Mutations P323L (in NSP12) and D614G (in spike protein) are the most occurred events and found to be globally dominant. Mutations change the structural stability by making the changes in free energy creating impact on the biological functions of the proteins [14]. However, all the mutations are not deleterious.

G614 form is highly transmissible than D614, associated with low RT-PCR cycle thresholds but didn't increase the disease severity [17]. On the other hand, the P323L mutation is located in the interface domain of the RNA-dependent RNA polymerase (RdRp) and it changes the intramolecular interactions in proteins as its stability changes. The other non-synonymous mutations, S194L in N protein and Q57H in NS3 are deleterious.

Clade wise analysis of mutations have shown that they mainly belong to clade GR and GH in Asia. Mutations in the clade GH are deleterious in nature by changing structural and functional changes in the proteins as the free energy(ΔG) changes. Human genome can carry large number of mutations, which as whole can also make significant contributions to the fatal outcome in disease.

A comprehensive analysis of SARSCoV-2 specific CD4+ and CD8+ cell responses from COVID-19 convalescent subjects and the recipients of Moderna (mRNA-1273)/ & Pfizer/BioNTech (BNT162b2) mRNA vaccines recognizing the parental strain than variant lineages B.1.1.7, B.1.351, P.1, and CAL.20C. The vast majority of the sequences SARSCoV-2 T- cell epitopes are not affected by the mutations found in the variants analyzed. The CD4+and CD8+ T cell responses are not substantially affected by the mutations found in the SARSCoV-2 variants. The bioinformatic analysis predicted the impact of mutations on various variants with sets of previously reported CD4+ and CD8+ T cell epitopes derived from the ancestral reference sequence [41]. T-cell functionalities were performed by AIM assay and FluoroSPOT assay. Therefore, there was a negligible impact of SARSCoV-2 variant on CD4+ and CD8+ T cells. The molecular basis revealed that 93% of CD4+ T cells and 97% CD8+ T cell epitopes are completely conserved in the variants. They found that the epitopes affected by the single mutations, not affecting the HLA binding capacity in majority of cases.

Undoubtedly, variants with increased transmissibility have been associated with decreased susceptibility to neutralizing antibodies from infected and vaccinated individuals. In contrast, the T-cell responses are largely unaffected by variants, but it is not anticipated the circulatory T memory cells would be effective in preventing the SARSCoV-2 infection. Hence, T-cell may contribute to limiting the COVID-19 severity. This could have the potential implications for engineering coronavirus vaccines with broader protective immunity against variant of concern [41].

The neutralization activity performed in cell culture of monoclonal and serum derived polyclonal Abs against the variants B.1.1.7 isolate, chimeric Washington strains with a South African (Wash SA-B.1.351) or Brazilian (Wash BR-B-1.1.248) spike gene, and isogenic recombinant variants with designed mutations or deletions at position 69-70, 417, 484, 501, 614 and/or 681 of spike protein. It has shown substantial diminishing of the neutralization potency of Abs against the SARS CoV-2 strains or isogenic variants containing a mutation at position 484. Reduced neutralization activities were obtained using highly neutralizing mAbs engaging the receptor -binding domain or N-terminal domain, convalescent sera and mRNA vaccine-induced immune sera showed reduced inhibitory activity against virus containing an E484K spike mutation. Abs to N-terminal domain demonstrate diminished neutralization potency in vitro against some emerging variants [42].

50% plaque reduction neutralization testing was performed using 20 serum samples after the second dose of BNT162b2 vaccine. All the serum samples efficiently neutralized USA-WA1/2020 and all the viruses with variant spikes i.e., USA-WA1/2020, B.1.1.7-spike, P.1-spike, B.1.351-spike, B.1.351(-Δ242-244+D614G), and B.1.351(-RBD+D614G). The neutralization titer against the B.1.351-spike virus was lower. This suggests that the mutations that result in amino acid substitutions K417N, E484K, and N501Y

in the receptor binding site have a greater effect on neutralization than the 242-244 deletion affecting the N-terminal domain of the spike protein [48].

A spike of COVID-19 cases in Maharashtra state, India has created greater concern to public health experts. Yadav,P.D. et al, has isolated and characterized the VUI lineage B.1.617 propagated in VeroCCL81 cells . Convalescent sera taken from COVID-19 recovered patients and recipients of BBV152 (Covaxin) were found to neutralize the VUI B.1.617. Further studies are underway to assess the clinical efficacy of B.1.617 variants [35].

Conclusion

It is significant to carry out mutational analysis for SARSCoV-2 clades in various geographic locations with high concern as an early warning to consider for further developments to implement the appropriate lockdown measures to break the transmission chains and conduct vaccine efficacy studies. GISAID provides rich data to study evolutionary relationships among the viral sequences obtained from different countries on different time periods.

Kober studied and analyzed the real-time evolution of SARSCoV-2 spike protein amid pandemic. She also tracked the mutational events especially in spike proteins with primary focus to meet the urgency to develop the strategies for vaccines and antibody therapy. This would inform making precise vaccine constructs & diagnostic reagents. Understanding D614G mutations overtaking the pandemic and the recombination impacting the evolution process of the virus would instruct responding in order to control the epidemic spread and resurgence.

Deleterious mutation at the immunodominant epitopes could make the Abs susceptible leading to facilitate the ADE in SARS CoV-2.

Diversity in infection from clades in Asia has shown more than 30% mutations in GR and GH clades. Frequently mutated amino acids were Glutamine and Serine and they both are deleterious. More than 80% deleterious mutations are responsible to affect the biological function (14). These characteristics provide the insight to develop clade specific vaccines.

Mutation's impact was seen on the 10 pseudoviruses harbouring the mutations of receptor binding domain (RBD). Surprisingly, 5 of them including N501Y, E484K, and K417N/T were found to be highly resistant to neutralization. The detailed clinical impact of these mutations is yet to be explored. These studies emphasize the need to develop broadly protective interventions against the evolving pandemic [44].

As virus have the more space/ more hosts to spread, it is highly likely that the virus emerges into new recombinants and variants. The constantly emerging variants could resist the current therapeutics. Under the less rigorous control measures, the viruses tend to spread faster comparatively.

Implementation of strict lockdown measures in New Zealand have successfully controlled the multiple chains of transmission and eliminated the virus several times. The ongoing genomic surveillance is an integral part of the national response to monitor any re-emergence of the virus, especially from the border and quarantine facilities [45].

Vaccines have provided enough protection so far and it is more convenient to manufacture vaccines using new variants' sequences in short period of time on the advanced technology platforms.

Cocktails of antibodies targeting the distinct structural and functional domains of spike protein of SARSCoV-2, are being currently developed considering redundant mechanism of targeting immunodominant portion to minimize the risk of escape mechanism [54].

The mandatory preventive measures and guidelines implemented by health authorities should be executed to contain the virus, especially in the epidemic areas. More than 70-80% of the current population, should get vaccinated to build the enough herd immunity throughout world to keep everyone safe.

Recommendations to be followed on pandemic preparedness response guidelines strengthen the disease control activities, applying risk-based & evidence-based approaches, including epidemiological surveillance, strategic testing plans, contact tracing to implement the public health and social measures to control the spread. Genomic surveillance programs should be strengthened to identify the emerging variants of concern.

Acknowledgements

Author extended her sincere gratitude to all the scientists and researchers for their great contributions in the literature help carry out this study. She is also thankful to other colleagues and seniors for sharing the knowledge.

Conflict of Interest

There is no conflict of interest exists.

Bibliography

1. Kaushal A. "mRNA Technology a Promising Strategy for Rapid Response Applications Against the Emerging Infectious Diseases". *Acta Scientific Microbiology* 4.3 (2020): 199-212.

2. Wikipedia COVID-19 Pandemic.

3. Adi Stern and Raul Andino. "Viral Evolution: Viral Pathogenesis (Third Edition)" (2016).

4. Lu R., *et al*. "Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding". *Lancet* 395 (2020): 565-574.

5. Wang Q., *et al*. "Structural and functional basis of SARS-CoV-2 entry by using human ACE2". *Cell* 181 (2020): 894-904.e9.

6. Daniele Mercatelli and Federico M Giorgi. "Geographic and Genomic Distribution of SARS-CoV-2 Mutations". *Frontiers in Microbiology* 11 (2020): 1800

7. Kaushal A. "Immune response and pathogenesis of COVID-19 and the strategies for developing target drugs". *Acta Scientific Microbiology* 3.9 (2020): 92-102.

8. Kaushal A. "Current and Future diagnostics tests for COVID-19- Challenges and Recommendations". *Acta Scientific Microbiology* 3.9 (2020): 103-119.

9. Forster P., *et al*. "Phylogenetic network analysis of SARS-CoV-2 genomes". *Proceedings of the National Academy of Sciences of the United States of America* 117 (2020): 9241-9243.

10. Seiya Ozono. "SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity". *Nature Communications* 12 (2021): 848.

11. Zhang L., *et al*. "Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors". *Science* 368 (2020): 409-412.

12. Rambaut A., *et al*. "A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology". *bioRxiv* (2020).

13. Kim D., *et al*. "The architecture of SARS-CoV-2 transcriptome". *Cell Press* 181 (2020): 914-921.e10.

14. Antara Sengupta., *et al*. "Clade GR and clade GH isolates of SARS-CoV-2 in Asia show highest amount of SNPs". *Infection, Genetics and Evolution* 89 (2021): 104724.

15. Capriotti E., *et al*. "I-mutant2. 0: predicting stability changes upon mutation from the protein sequence or structure". *Nucleic Acids Research* 33 (2005): W306-W310.

16. Victoria Elizondo., *et al*. "SARS-CoV-2 genomic characterization and clinical manifestation of the COVID-19 outbreak in Uruguay". *MedRxiv* (2020).

17. Korber B., *et al*. "Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2". *BioRxiv* (2020).

18. Kurtz S., *et al*. "Versatile and open software for comparing large genomes". *Genome Biology* 5 (2004): R12-R12.

19. Kirchdoerfer RN., *et al*. "Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis". *Scientific Report* 8 (2018): 15701-15701.

20. Wrapp D., *et al*. "Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation". *Science* 367 (2020): 1260.

21. Walls AC., *et al*. "Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein". *Cell* 181 (2020): 281-292. e286.

22. Wan, Y., *et al*. "Molecular Mechanism for Antibody-Dependent Enhancement of Coronavirus Entry". *Journal of Virology* 94 (2020): e02015-02019.

23. Yip MS., *et al*. "Antibody-dependent enhancement of SARS coronavirus infection and its role in the pathogenesis of SARS". *Hong Kong Medical Journal* 22 (2016): 25-31.

24. Corman VM., *et al*. "Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR". *Eurosurveillance* 25 (2020): 2000045.

25. Yan R., *et al*. "Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2". *Science (New York, NY)* 367 (2020): 1444-1448.

26. Yuan, M., *et al*. "A highly conserved cryptic epitope in the receptor-binding domains of SARS-CoV-2 and SARS-CoV". *Science* (2020): eabb7269.

27. Drew Weissman., *et al*. "D614G Spike Mutation Increases SARS CoV-2 Susceptibility to Neutralization". *Cell Host and Microbe* 13. 29.1 (2021): 23-31.e4.

28. Pardi N., *et al*. "Nucleoside-modified mRNA vaccines induce potent T follicular helper and germinal center B cell responses". *Journal of Experimental Medicine* 215 (2018): 1571-1588.

29. Jackson L.A., *et al*. "mRNA-1273 Study Group An mRNA Vaccine against SARS-CoV-2 - Preliminary Report". *The New England Journal of Medicine* 383 (2020): 1920-1931.

30. Schmidt F., *et al*. "Measuring SARS-CoV-2 neutralizing antibody activity using pseudotyped and chimeric viruses". *Journal of Experimental Medicine* 217 (2020)"217.

31. Chand VM., *et al* "Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data". *MedRxiv* (2020).

32. Kathy Leung., *et al*. "Early empirical assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom". *MedRxiv* (2020).

33. Angus A Nassir. "Mutation Landscape of SARS COV2 in Africa". *MedRxiv* (2020).

34. Tegally H., *et al*. "Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa". *MedRxiv* (2020).

35. Yadav PD., *et al.* "Neutralization of variant under investigation B.1.617 with sera of BBV152 vaccinees". *BioRxiv* (2021).

36. Rambaut Andrew., *et al*. "Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations (Report)". Written on behalf of COVID-19 Genomics Consortium UK (2020).

37. Chand, Meera., *et al*. "Investigation of novel SARS-COV-2 variant: Variant of Concern 202012/01 (PDF) (Report)". Public Health England (2020).

38. Francis., *et al*. "Oxford/AstraZeneca COVID shot less effective against South African variant: study". Reuters (2021).

39. Faria Nuno., *et al*. "Genomics and epidemiology of a novel SARS-CoV-2 lineage in Manaus, Brazil". *medRxiv* (2021).

40. Medicaloid (Healthcare around the Globe). "California covid variant B.1.427/B.1.429 is more transmissible studies show" (2021).

41. Tarke A., *et al* "Negligible impact of SARS-CoV-2 variants on CD4+ and CD8+ T cell reactivity in COVID-19 exposed donors and vaccinees". *bioRxiv* (2021).

42. Rita E. Chen., *et al*. "Resistance of SARS-CoV-2 variants to neutralization by monoclonal and serum-derived polyclonal antibodies". *Nature Medicine* (2021).

43. Becerra-Flores M., *et al*. "SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate". *International Journal of Clinical Practice* (2020): e13525.

44. Wilfredo F Garcia-Beltran., *et al*. "Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral Immunity". *Cell* (2021).

45. Jemma L Geoghegan., *et al*. "Genomic epidemiology reveals transmission patterns and dynamics of SARS-CoV-2 in Aotearoa New Zealand". *Nature Communications,* Article number: 6351 (2020).

46. Albert Tian Chen. "COVID-19 CG: Tracking SARS-CoV-2 mutations by locations and dates of interest". *bioRxiv* (2020).

47. Summer E Galloway, CDC. "Emergence of SARS-CoV-2 B.1.1.7 Lineage — United States". *CDC* 70.3 (2021): 95-99.

48. "Neutralizing Activity of BNT162b2-Elicited Serum". *The New England Journal of Medicine* (2021).

49. Wikipedia UK variant Lineage B.1.1.7.

50. Wikipedia South Africa variant Lineage B.1.351.

51. Wikipedia Brazil variant LineageP.1.

52. Kaul R. "India's double mutant could be considered a variant of concern". *India News* (2021).

53. Tahir S Pillay. "Gene of the Month : 2019-nCoV/ SARS CoV-2 novel coronavirus protein". *Journal of Clinical Pathology* 73 (2020): 7.

54. Jolly Bani., *et al*. "Genetic epidemiology of variants associated with immune escape from global SARS-CoV-2 genomes". *bioRxiv* (2021).