Evolutionary tracking of SARS-CoV-2 genetic variants highlights intricate balance of stabilizing and destabilizing mutations

Jobin John Jacob[1*], Karthick Vasudevan[1*], Agila Kumari Pragasam[1], Karthik Gunasekaran[2], Gagandeep Kang[3], Balaji Veeraraghavan[1#], Ankur Mutreja[4#]

[*]Contributted equally to the work

[#]Joint corresponding author

[1]Department of Clinical Microbiology, Christian Medical College, Vellore, India

[2]Department of General Medicine (Unit-V), Christian Medical College, Vellore, India

[3]Division of Gastrointestinal Sciences, Christian Medical College, Vellore, India

[4]Cambridge Institute of Therapeutic Immunology & Infectious Disease (CITIID) Department of Medicine, University of Cambridge, Cambridge, UK

jjj1986@gmail.com

karthick.1087@gmail.com

agilakpragasam@gmail.com

karthikgunasekaran@yahoo.com

gkang@cmcvellore.ac.in

vbalaji@cmcvellore.ac.in

am872@medschl.cam.ac.uk

Abstract

The currently ongoing COVID-19 pandemic caused by SARS-CoV-2 has accounted for millions of infections and deaths across the globe. Genome sequences of SARS-CoV-2 are being published daily in various forums and this big data availability has allowed unprecedented access into the mutational patterns of SARS-CoV-2 evolution. Genomic information has further been utilised in public health decision making through molecular epidemiology, transmission tracking and vaccine design. We made use of the same sea of genomic information for conducting a detailed phylogenetic analysis and identifying lineage-specific mutations. The catalogued mutations were analysed for their stabilizing or destabilizing impact on E, M, N and S viral proteins that are key for infection, virulence and pathogenesis. The same proteins make crucial components of vaccines recently approved and in development. We recorded positive natural selection of D614G, S477N, A222V V1176F variants and a global expansion of the PANGOLIN variant B.1. In addition, a positive natural selection of Q57H (B.1.X), R203K/G204R (B.1.1.X), T85I (B.1.2-B.1.3), G15S+T428I (C.X) and I120F (D.X) variants was observed. Overall, we recorded a striking balance between stabilizing and destabilizing mutations, therefore incredibly well maintained protein structures. With selection pressures in the form of newly developed vaccines and therapeutics to mount soon in coming months, the task of mapping of viral mutations and recording of their impact on key viral proteins would be crucial to pre-emptively catch any escape mechanism that SARS-CoV-2 may evolve for.

Key Words: SARS-CoV-2, mutations, stabilizing/destabilizing, vaccine, evolution

Introduction

The emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Wuhan, China and the subsequent global spread has brought the world to a standstill (WHO, 2020). During the course of 11 months the coronavirus disease 19 (COVID-19) pandemic has caused more than 72 million confirmed cases in 217 countries with close to 1,620,000 fatalities. (Worldometer 14/12/2020). While initially, and rightly too, the efforts were focused on minimising the number of cases and deaths due to COVID-19 (WHO 2020). This included fast tracking the search and development of novel treatment and prevention options (Li & De Clercq, 2020). Today, however, as vaccine candidates have started showing promising results, there is a cautious shift towards assessing the efficacy of vaccine candidates with

respect to the circulating diversity of SARS-CoV-2 and its continuously evolving genetic variants (Hodgson et al., 2020).

Functional mutations that help the virus to adapt to the recent host-shift events are hypothesised to drive the evolution of transmissibility and virulence in SARS-CoV-2 (Sironi et al., 2020). Shortly after the first isolated SARS-CoV-2 genome from China was published, >30,500 distinct mutations were catalogued in the CoV-GLUE (http://cov-glue.cvr.gla.ac.uk/) among globally circulating strains of this virus (Singer et al., 2020). Variation in the genetic makeup are key determinants in measuring the evolutionary distance and stability of SARS-CoV-2 from the first sequenced isolate (Hu et al., 2020). Moreover, tracking the evolution of SARS-CoV-2 since its introduction in humans is a high priority undertaking to prevent future waves of this pandemic from escaping the global preparedness (van Dorp et al., 2020). Since many vaccine candidates currently under development are derived from the first available SARS-CoV-2 sequences, recurrent genetic changes may have an unforeseen impact on their sustained effectiveness in longer term (Dearlove et al., 2020).

The availability of whole genome sequences of SARS-CoV-2 in public repositories such as Global Initiative on Sharing All Influenza Data (GISAID) and real-time data visualisation pipeline NextStrain (https://nextstrain.org) offers a great opportunity for scientists to track the evolutionary path of this virus. Phylogenetic Assignment of Named Global Outbreak LINeages tool (PANGOLIN) has been the most widely used tool for lineage assignment to newly emerging variants. PANGOLIN has also been deployed in establishing the transmission patterns of various clones of this virus (Rambaut et al., 2020). Due to the fast-evolving nature of this virus, the evolution and respective lineage assignment has not been easy (Tang et al., 2020; Rambaut et al., 2020). As a result, multiple studies that tracked the evolution of SARS-CoV-2, have been hugely controversial. For example, doubts have been cast on the claim of finding more aggressive L type emerging from S type strains (Tang et al., 2020). Similarly, the hypothesis of rapid spread of D614G variant of SARS-CoV-2 indicating a possible fitness advantage has been questioned (Plante et al., 2020; Korber et al., 2020; Hoffmann et al., 2020). Therefore, in the current and highly sensitive global circumstances due to this pandemic, having a detailed map of mutations highlighting their prospective role in therapeutics and vaccine development can prepare us better for the future waves of continuously evolving SARS-Cov-2. In this study, we present a catalogue of the most important genomic mutations recorded between December 2019 – November 2020 in SARS-

Cov-2 and their possible impact on stability of protein candidates that form crucial part of vaccines and also constitute the most common therapeutic targets.

Methods

Data acquisition and curation

In total, we have retrieved 7000 genomes from GISAID EpiCoV database (https://www.gisaid.org/). Datasets that were flagged as complete (>28,000 bp) were screened and subsequently manually curated for excluding low quality/coverage sequences and duplicates. Sequence metadata (Supplementary Table S1) was retrieved and genome containing sampling time and location were only chosen for the study. Lineages were assigned from alignment file using the Phylogenetic Assignment of Named Global Outbreak LINeages tool PANGOLIN v1.07 (https://github.com/hCoV-2019/pangolin). We selected a subset of 513 genomes that belongs to all major PANGOLIN lineages and common mutations for the optimal output of the phylogenetic tree.

Phylogenetic analysis

Genome sequences were aligned against the original Wuhan-Hu-1 genome (Accession: NC_045512) using multiple genome sequence alignment tool MAFFT (v6.240). Subsequently, the error prone 5'-UTR and 3'-UTR regions were masked and the genome size was adjusted without losing key sites. Maximum likelihood tree was generated using IQTREE v.1.6.1 (http://www.iqtree.org/) under the GTR nucleotide substitution model with 1000 bootstrap replicates. The ML tree was visualised and labelled using the interactive tree of life software iTOL v.3 (Letunic & Bork, 2019).

Mutation Profiling

In order to identify the genetic variants, assembled genomes were mapped against the reference (Wuhan-Hu-1: Accession: NC_045512) using Snippy mapping and variant calling pipeline (https://github.com/tseemann/snippy). Among the SNPs, missense SNPs (nonsynonymous) was extracted using custom written bash scripts and manually curated as per CoV-GLUE database (http://cov-glue.cvr.gla.ac.uk/). Specifically, we considered 24 major missense mutations in four major structural proteins: Envelope protein (E), Membrane glycoprotein (M), Nucleocapsid phosphoprotein (N) and Spike protein (S). In addition, stability of 38 amino acid substitutions in the RBD domain and/or Spike protein were further

analysed to examine the potential impact of these mutations on protein stability of S and binding affinity of RBD, both being key antigens in COVID vaccines.

Protein homology modelling

The complete genome sequence of SARS-CoV-2 (Accession: NC_045512.2) was obtained from the National Centre for Biotechnology Information (NCBI). The 3D structure of the major proteins (E, M, N and S) was modelled using SWISS MODEL (https://swissmodel.expasy.org/). When no reliable templates were available for homology modelling, the 3D structures were downloaded from I-Tasser (https://zhanglab.ccmb.med.umich.edu/COVID-19/). The quality of models was validated using Ramachandran plot.

Stability analysis

The stability of 62 missense mutations in four major proteins (E, M, N and S) was analysed utilizing Site Directed Mutator (SDM) and DUET servers (Pandurangan et al., 2017; Pires et al., 2014a). The impact of variations is predicted based on the difference in free energy between folded and unfolded states. The stability of the protein is represented by changes in Gibbs free energy. The positive and negative values of ΔΔG° indicate the stabilizing and destabilizing mutations. SDM predicts changes in protein stability upon amino acid substitution using 3D structure of the protein as input while DUET is an integrated support-vector machines (SVM) based approach that combines two complementary approaches (mCSM and SDM) and in a consensus prediction (Pires et al., 2014b). Structural stability of the proteins was confirmed only when both the servers predicted similar outputs. The change in vibrational entropy energy (ΔΔSVib ENCoM) between wild-type and mutant protein was calculated using DynaMut (Rodrigues et al., 2018).

Results

Diversity of SARS-CoV-2 Genomes

Of the 7000 SARS-CoV-2 genomes screened, we constructed a robust phylogenetic tree on strategically selected 513 genomes that reflected the most complete diversity among the isolates by covering all the PANGOLIN lineages. Lineage assignment based on PANGOLIN tool indicated the circulation of seven distinct lineages and/or sub-lineages such as A, B.1,

B.1.1, B.1.1.1, B.2, B.3, B.4 and B.6. This is in line with the phylogenetic groupings by GISAID (S, L, V, O, G, GH and GR) (**Figure 1**). As the epidemic has progressed and mutations have accumulated, further subdivision of major lineages into sub-lineages has been observed. Overall, a total of 61 lineages and sub-lineages have been found to be circulating concurrently in multiple countries around the world. In general, numerous introductions of different variants were observed across the globe with a few sub-lineages (C.2, D.2) being restricted to certain regions. While B.1.113 lineage, for example, has been exclusively reported from India, lineages C.2 and D.2 geographically have been confined to South Africa and Australia, respectively.

Major amino acid substitutions

Mutation mapping showed a total of 106 amino acid substitutions (missense mutations in >5 genomes) from a representative set of 513 genomes. The analysis also revealed 36 mutations that were found in >5% of genome sequences while 12 major substitutions were lineage defining mutations (**Figure 1)**. The first major mutation to appear was L84S in ORF8, (present in 8.6% of the genomes) that has defined A lineage (i.e. clade S in GISAID classification). The subsequent amino acid substitutions L37F in ORF3a and G251V in *nsp6* were found to be present in 13.3% and 1.4% of genomes, respectively. The combination of G251V and L37F, which was initially considered as a defining mutation pattern for B.2 – B.6 lineage (clade V in GISAID classification), more detailed analysis has shown that isolates carrying G251V mutation are distributed in other lineages too. The predominant lineage defining mutations in the whole dataset were D614G (85.5%) and P323L (85.5%), after originally appearing in late January 2020 (**Figure 2**). Other major mutations noted are Q57H (26.5%), R203K/G204R (33%), G15S (12%), I120F (11.5%) and T85I (14%).

Dominance of D614G variant

Two mutations have become consensus: D614G in S (nucleotide 23,403, A to G) and P323L (also known as P4715L) in nsp12 (nucleotide 14,143, C to T). These mutations were present in 80.5% of the sequences and have defined the B.1 lineage (G in GISAID classification). The widely discussed D614G variant is speculated to have been introduced in Europe at the end of January (EPI_ISl_422424) before becoming globally dominant. Genomes with D614G mutations were assigned as B.1 by PANGOLIN or GH/GR by GISAID. Notably, founder lineage B.1 and its sub lineages B.1.X, B.1.1.X, D.X and C.X that carry both D614G and

P323L mutations have become the dominant variants across the world (87% of global collection as per CoV-GLUE as on 30th November 2020).

As the pandemic has progressed several other major substitutions affecting the protein structure have appeared. These are Q57H (nucleotide 25,563, G to T) in ORF3a, R203K + G204R combination (nucleotide 28,881, GGG to AAC) in Nucleocapsid and T85I (nucleotide 1059 C to T) in ORF1a. The region-specific sub lineages C.1, C.2, D.1 and D.2 were found to cumulatively harbour multiple mutations. Amino acid substitution such as T428I and G15S in ORF1a were reported in sub lineages C.1 and C.2, and S477N substitution in S protein along with I120F in nsp2, which specifically established the sub lineage D.2 (**Figure 1**).

Balance of stabilizing and destabilizing mutations

The impact of mutations in key proteins is widely accepted as an important factor in driving the clinically meaningful evolution, which potentially allows any pathogen to escape available treatment and prevention regimes. Recent experimental evidence (Zhang et al., 2020; Korber et al., 2020; Fernández et al.,2020) suggests that the emergence of D614G variant led to stable phenotype with increased structural stability and enhanced infectivity. Among the 24 major missense mutations, our analysis using SDM predicted 13 missense mutations as stabilizing while using DUET analysis predicted 11. Among the conservatively selected 14 missense mutations, 10 were predicted to be stabilizing by both SDM and DUET server (**Table 1**).

Spike protein in full or part is the key vaccine candidate in today's vaccine landscape. Of the additional 38 amino acid substitutions that have been checked particularly for the stability of spike protein, 13 were found to be stabilizing mutations while 16 were destabilizing the protein structure. Nine mutations have given inconclusive results as SDM and DUET server provided contrary ΔΔG values (**Table 1**). Overall, from both the datasets, 48 amino acid substitutions in spike protein were tested for stability of which 16 were stabilizing, 20 were destabilizing and 12 showed inconclusive results. When checked for amino acid substitutions, the stabilizing mutation in S protein predicted an increase in the rigidity of its structure (**Figure 3; Figure 4**). The increased rigidities of the structure may provide a stable conformation to the protein that may positively influence the binding of spike protein to ACE2 receptor (Ou et al., 2020). Major mutations D614G and S477N were located at

potential epitope regions (Codons 469–882) with S477N particularly positioned in the receptor-binding domain (RBD) of the S protein (319 – 541).

Most frequent amino acid substitutions were observed in the N protein, in which the variants S194L, D103Y, P13L, S197L, M234I, and S188L were predicted to be stabilizing according to both the analytical servers (**Table 1**). In contrast, M and E proteins accounted for the least number of amino acid substitutions. The amino acid changes in M (T175M) indicated a stabilizing effect, while E does not account for any stabilizing variant. Structural analysis of double (D614G + S477N; D614G + A222V) and triple (D614G + S477N + A222V) mutation patterns in S protein indicated ΔΔG values of 0.228, 0.195 and 0.129, respectively (**Table 2**). This signifies that accumulation of spike mutation in D614G bearing lineages could potentially be affecting the stability of the spike and therefore may influence the binding affinity towards ACE2 receptor.

Discussion

Since the beginning of the ongoing COVID-19 pandemic, whole genome sequence based phylogenetic inference has been heavily utilized in tracing viral origins and transmission chains (Munnink et al., 2020). However, as the virus has evolved with time, genomic data is being increasingly used in guiding infection risk and control strategies. Several genomic mutations have been mapped that seem to be of advantage to the virus (Gómez-Carballa et al., 2020). In parallel, numerous vaccine candidates have been designed using genomic data from the original SARS-CoV-2 strain of Wuhan and many are now approved for use or at late-stage trials (Dong et al., 2020; Alturki et al., 2020). Based on immunological data obtained from infected and recovered patients, rightly, almost all COVID-19 vaccine candidates of today are based on the original SARS-CoV-2 spike protein or its RBD domain (Corbett et al., 2020; Sahin et al., 2020; Poland et al., 2020). However, as vaccines are introduced and successful treatment options become available, it is vital that we carefully monitor the mutations in the immunogenic region of SARS-CoV-2 genome (Li et al., 2020). Mapping these changes on protein structure will allow pre-emptive forecasting of the direction of change in vaccine effectiveness and guide future preparedness efforts. We analysed the impact of recurrent amino acid replacements in the genomic evolution and proteome stability of SARS-CoV-2 since its introduction in December 2019 to November 2020. Our analysis found an intriguing balance of stabilizing and destabilizing mutations, allowing the SARS-CoV-2 to evolve and persist without losing pathogenicity.

SARS-CoV-2 is considered a slowly-evolving virus as it possesses an inherent proofreading mechanism to repair the mismatches during its replication. This is believed to have a crucial role in maintaining the stability and integrity of the viral genome (Robson et al., 2020; Bar-On et al., 2020). Our analysis confirmed previously recorded positive natural selection of D614G, S477N (Starr et al., 2020), S477N, A222V and V1176F (Farkas et al., 2020) variants and a global expansion of the PANGOLIN variant B.1 (Rambaut et al., 2020). In addition, we also observed a positive natural selection of Q57H (B.1.X), R203K/G204R (B.1.1.X), T85I (B.1.2-B.1.3), G15S+T428I (C.X) and I120F (D.X) variants (**Figure 2**).

Apart from the clade defining mutations, some of the major missense mutations were in the four structural proteins (E, M, N and S). When analysed for their impact in the respective protein structure, spike glycoprotein, more specifically its RBD domain was found to be most vulnerable to frequent mutations. This may be due to the immunological observation that most neutralizing anti-SARS-CoV-2 antibodies have been found to target the RBD domain of the S protein (Ju et al., 2020; Liu et al., 2020). Consistent with this finding, a total of 4170 missense mutations have been reported in the spike protein, with 683 on the RBD domain alone (CoV-Glue accessed on 12th December 2020). Computational prediction to understand the effect of amino acid substitutions in E, M, N and S proteins revealed a balance of stabilization and destabilization of the proteins. While viral population carrying mutations with higher stabilizing effects (Positive $\Delta\Delta G$ values) would be expected to become the dominant variant, it is interesting to note that destabilization mutations in the major protein targets of SARS-CoV-2 have also generated variants that have been hugely successful. For example, many of the favourably selected variants such as L18F, L5F, R203K, G204R, A220V (Nucleocapsid) were found to be destabilizing the respective protein structure (Table 1). As destabilizing mutations are known for their crucial functional roles, a trade-off between stabilizing and destabilizing mutations may balance the protein function and structure in ways that are not fully understood yet (Laha et al., 2020; Teng et al., 2020).

The currently circulating spike variants or RBD variants need to be taken into account while evaluating the vaccine candidate or neutralizing monoclonal antibodies against SARS-CoV-2 (Weisblum et al., 2020). Mapping the viral mutations that escape antibody binding is essential for accessing the efficacy of therapeutic and prophylactic anti-SARSCoV-2 agents (Starr et al., 2020a; Starr et al., 2020b). Recently generated experimental evidence suggests that leading vaccines (mRNA-1273, BNT162b1 and ChAdOx1a) and two potent neutralizing

antibodies (REGN10987 and REGN10933) are unlikely to be affected by the dominant variant D614G (Corbett et al., 2020; Sahin et al., 2020; Voysey et al., 2020; Baum et al., 2020; McAuley et al., 2020). As all three candidate vaccines encode RBD or the part of spike protein as antigens, the viral population is expected to try and escape by altering the positioning of the respective antigens (Greaney et al.,2020) when vaccination induced selection pressure would be on. Notably, complete escape mutation map of 3,804 of the 3,819 possible RBD amino acid mutations against ten human monoclonal antibodies are already in place (Starr et al., 2020a; Greaney et al.,2020). The antigenic effect of key RBD mutations against REGN-COV2 cocktail (REGN10933 and REGN10987) showed N439K and K444R variants escaped neutralization only by REGN10987, while E406W escaped both individual REGN-COV2 antibodies and the cocktail (Starr et al., 2020b). Similar strategies should be adopted to map all antibody resistance mutations against neutralizing antibodies elicited after vaccination. Once mutation escape maps are available for all successful vaccine candidates, vaccine roll out strategies should be carefully planned to dodge geographically confined escape mutants.

Our study highlights the importance of continued genomic surveillance, mutation mapping, stability analysis and potential escape mutation cataloguing in the pre- and post-vaccination period of SARS-CoV-2 in designing epidemiologically best vaccination programs.

Conclusion

The currently observed mutation pattern and subsequent phylogenetic diversification of SARS-CoV-2 seem to be strongly influenced by the negative and positive selection pressures. The overall variation in SARS-CoV-2 sequences is currently low compared to many other RNA viruses. One of the possible reasons for the slow rate of mutations can be attributed to the widespread absence of neutralizing antibodies or the selective pressure. Once the virus population is challenged with the vaccine candidates or therapeutic monoclonal antibodies the currently known epitopes on surfaces of SARS-CoV-2 proteins are likely to undergo rapid forced change for survival. Thus, the prevalence of such possible escape mutations needs to be monitored even more carefully after vaccination if we are to remain ahead of this pandemic curve.

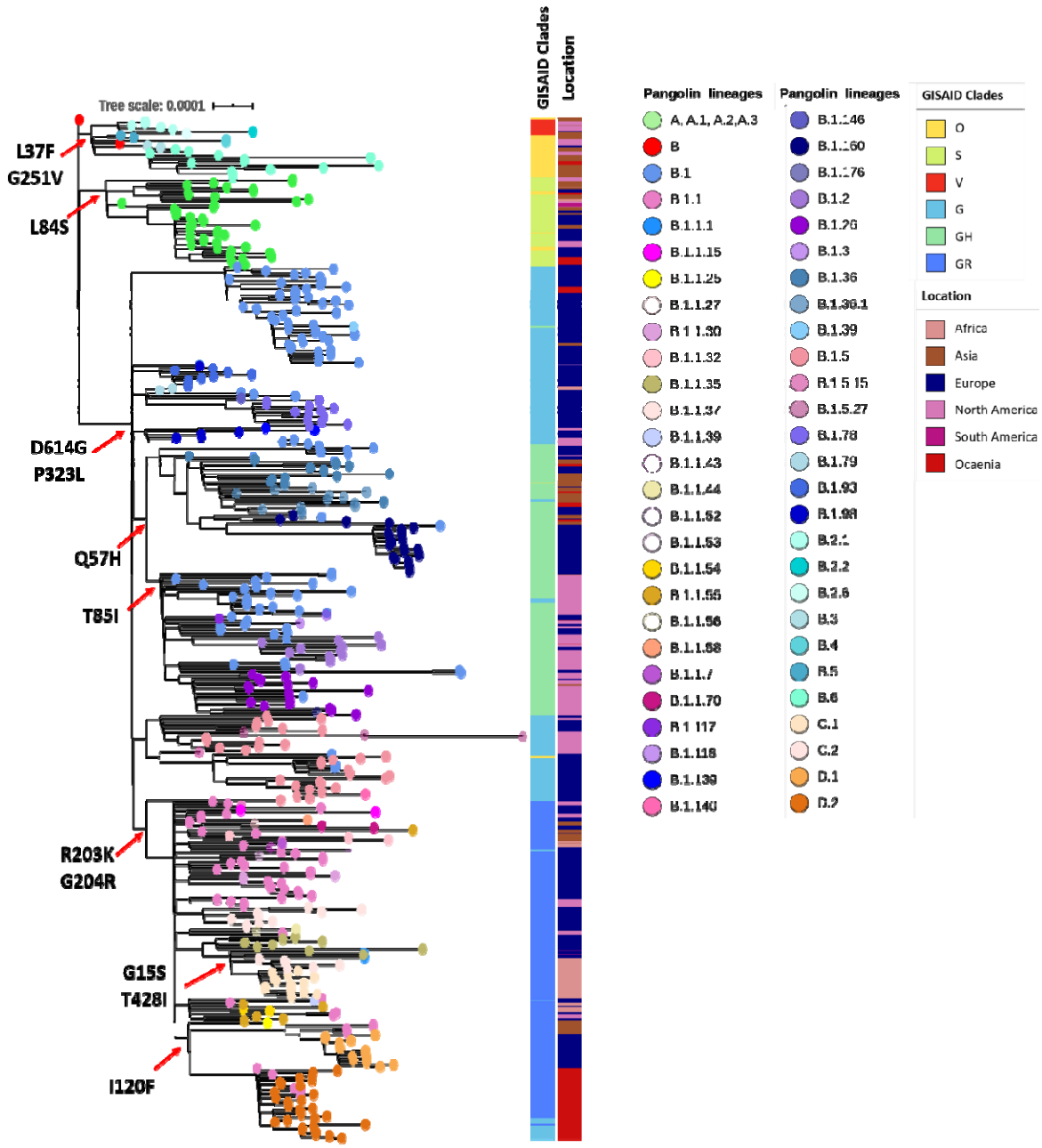


Figure 1: Maximum Likelihood phylogenetic tree inferred from 513 SARS-CoV-2 genomes. The tree was constructed using multiple genome sequence alignment (MAFFT) by mapping against the Wuhan-Hu-1 strain (Accession: NC_045512). Tips are coloured the major lineages assigned by PANGOLIN. Respective lineages assigned by GISAID and origin of sequence are labelled as colour strips. The scale bar indicates the distance corresponding to substitution per site.

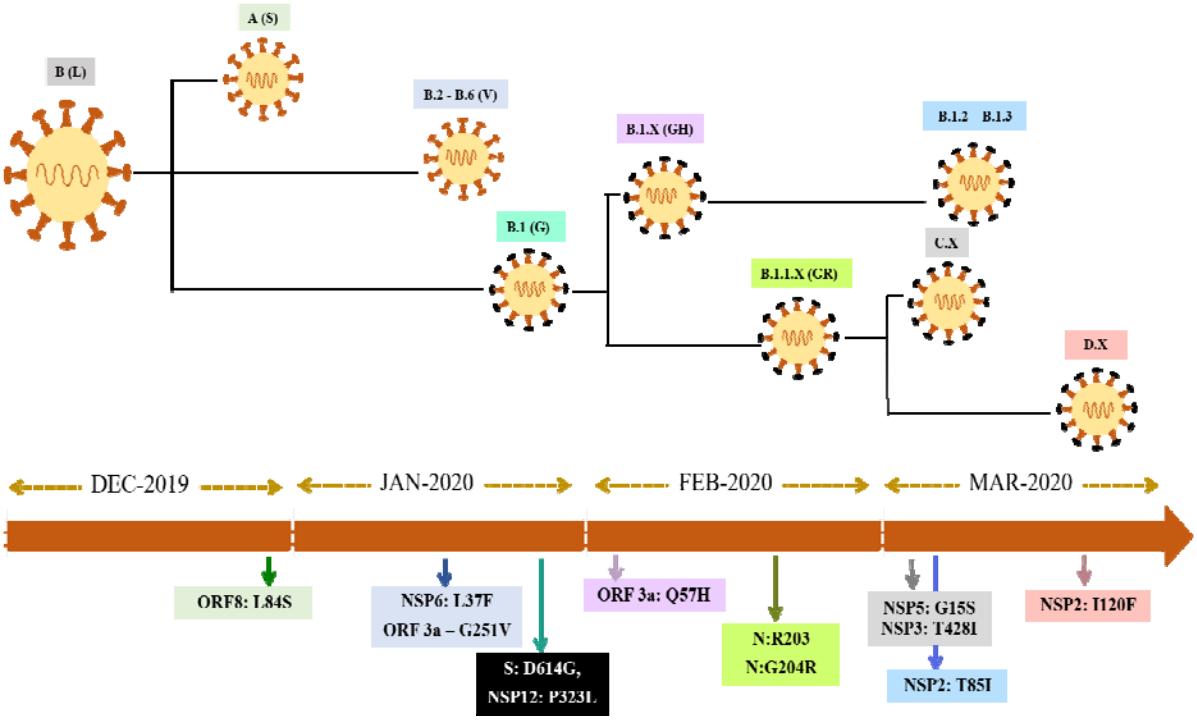


Figure 2: Schematic representation of the major evolutionary events/ amino acid substitution that give rise to SARS-CoV-2 variants in sequential order.

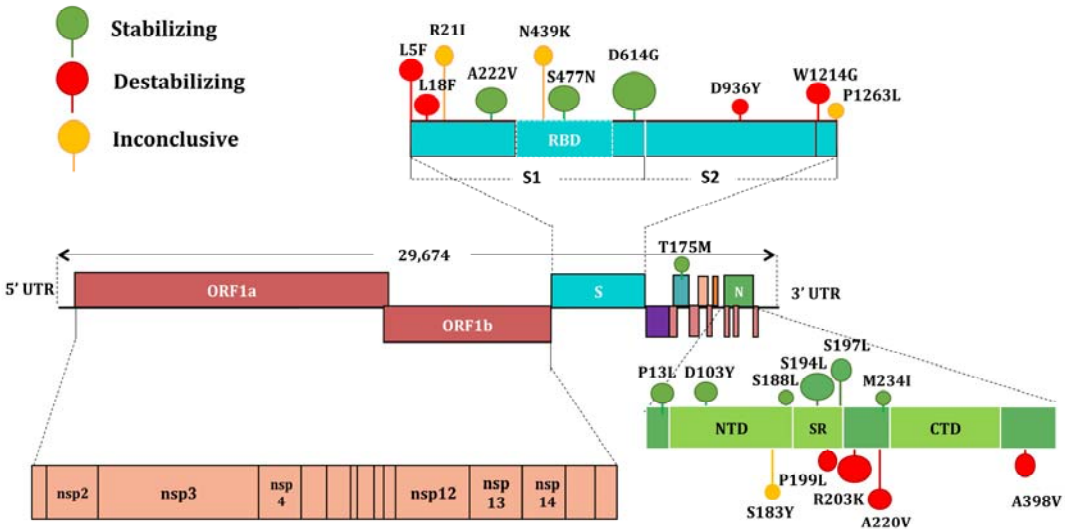


Figure 3: Schematic representation of SARS-CoV-2 genome organization, the major amino acid substitutions and stability of amino acid changes. Stabilizing mutations are coloured

green, Destabilizing mutations are coloured red and mutations that neither stabilize nor destabilize are coloured in yellow

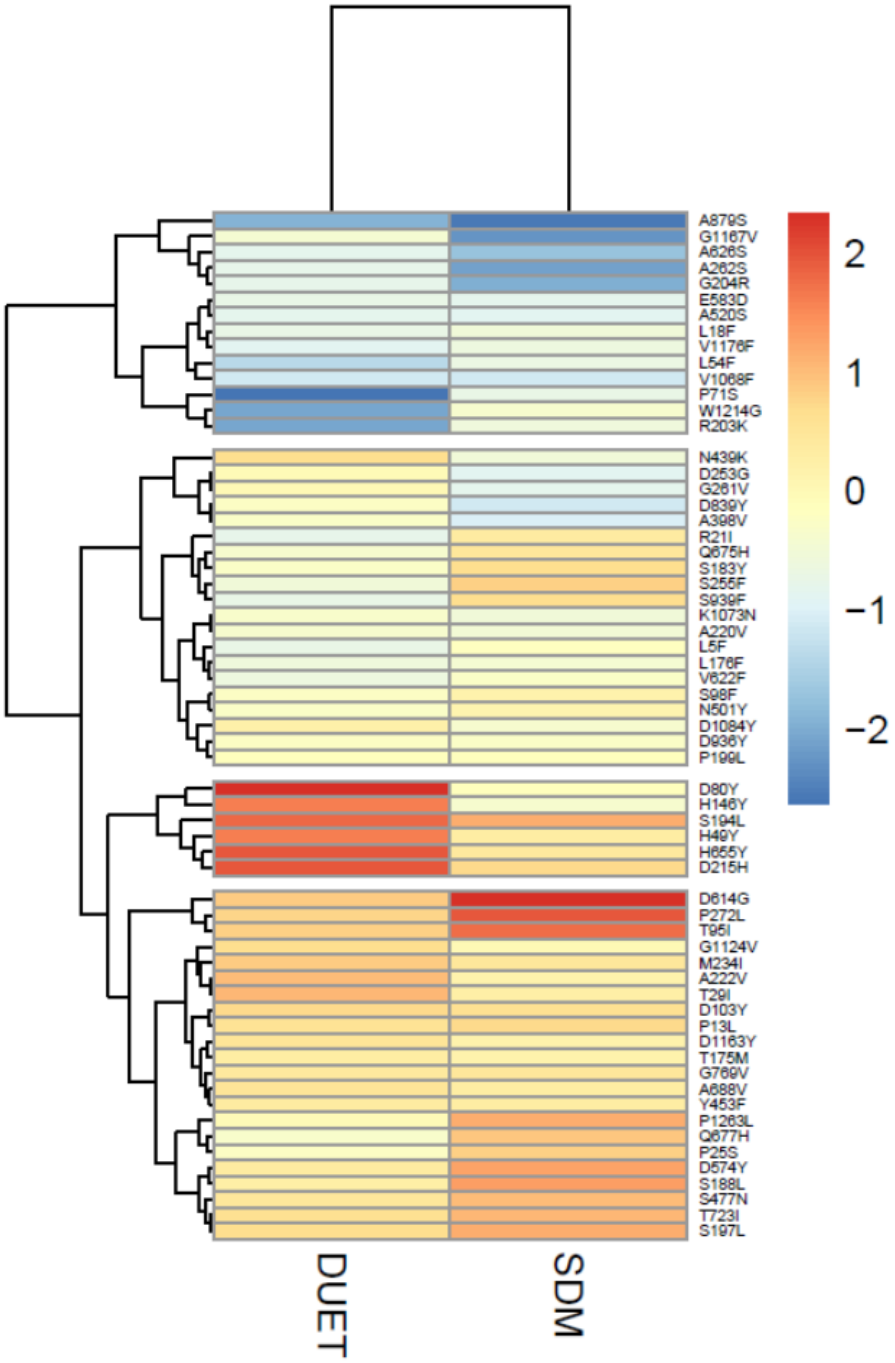
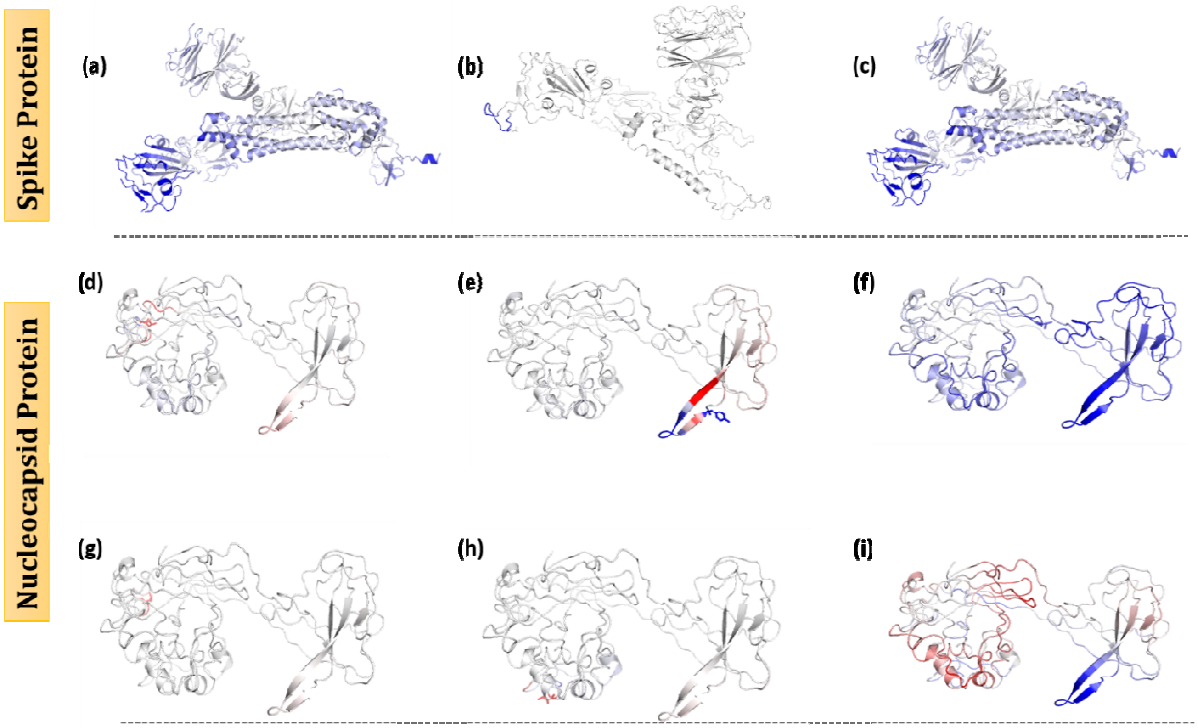


Figure 4: Heat map showing the stabilizing and destabilizing mutations of SARS-CoV-2 proteins based on the predicted ΔΔG values



Suppl Fig 1: Functional effect of mutated residues and conformational changes in major structural proteins: (a) D614G, (b) S477N and (c) A222V (d) S194L, (e) D103Y, (f) P13L (g) S197L, (h) M234I and (i) S188L Amino acids colored according to the vibrational entropy change upon mutation. **Blue** represents a rigidification of the structure and **Red** a gain in flexibility

Table 1: Predicted effect of protein stability in the presence of amino acid mutations in the SARS-COV-2 genomes.

Protein	Mutations	ΔΔG SDM (kcal/mol)	ΔΔG DUET (kcal/mol)	Stability
Spike	D614G	0.53	2.51	Stabilizing
	A222V	0.65	0.27	Stabilizing
	S477N	0.2	1.16	Stabilizing
	L18F	-0.801	-0.46	Destabilizing
	N439K	0.35	-0.48	Inconclusive
	L5F	-0.801	-0.1	Destabilizing
	W1214G	-1.913	-0.28	Destabilizing
	R21I	-0.856	0.46	Inconclusive
	A262S	-0.84	-2.11	Destabilizing
	S98F	-0.4	0.29	Inconclusive
	D1163Y	0.21	0.26	Stabilizing
	G1167V	-0.58	-2.25	Destabilizing
	D936Y	-0.38	-0.19	Destabilizing
	P272L	0.42	2.12	Stabilizing
	D80Y	1.73	-0.06	Inconclusive
	E583D	-0.82	-0.74	Destabilizing
	P1263L	-0.231	1.29	Inconclusive
	K1073N	-0.48	-0.48	Destabilizing
	D253G	-0.21	-0.84	Destabilizing
	T723I	0.3	1.21	Stabilizing
	A688V	0.21	0.41	Stabilizing
	A626S	-0.9	-1.69	Destabilizing
	L54F	-1.33	-0.61	Destabilizing
	H655Y	1.44	0.52	Stabilizing
	G769V	0.14	0.6	Stabilizing
	L176F	-0.7	-0.33	Destabilizing
	G1124V	0.36	0.13	Stabilizing
	V622F	-0.71	-0.2	Destabilizing
	S255F	-0.64	0.94	Inconclusive
	H49Y	1.14	0.4	Stabilizing

	D839Y	-0.389	-1.08	Destabilizing
	V1176F	-0.92	-0.55	Destabilizing
	D215H	1.41	0.83	Stabilizing
	H146Y	1.139	-0.29	Inconclusive
	A879S	-1.8	-2.57	Destabilizing
	Q677H	-0.48	1.06	Inconclusive
	D1084Y	-0.02	-0.3	Destabilizing
	V1068F	-1.13	-1.05	Destabilizing
	P25S	-0.392	0.93	Inconclusive
	A520S	-0.88	-0.84	Destabilizing
	G261V	-0.17	-0.78	Destabilizing
	D574Y	0.12	1.38	Stabilizing
	T29I	0.7	0.35	Stabilizing
	Y453F	0.12	0.42	Stabilizing
	N501Y	-0.44	0.19	Inconclusive
	S939F	-0.8	0.8	Inconclusive
	T95I	0.49	1.91	Stabilizing
	Q675H	-0.53	0.58	Inconclusive
Nucleocapsid	R203K	-0.500	-1.919	Destabilizing
	G204R	-1.950	-0.840	Destabilizing
	A220V	-0.390	-0.505	Destabilizing
	S194L	1.280	1.316	Stabilizing
	D103Y	0.750	0.380	Stabilizing
	P13L	0.840	0.287	Stabilizing
	S197L	1.280	0.349	Stabilizing
	A398V	-0.990	-0.435	Destabilizing
	P199L	-0.040	-0.326	Destabilizing
	M234I	0.570	0.499	Stabilizing
	S188L	1.450	0.033	Stabilizing
	S183Y	0.770	-0.457	Inconclusive
Membrane	T175M	0.290	0.070	Stabilizing
Envelope	P71S	-0.660	-2.403	Destabilizing

Table 2: Impact of double and triple mutation in the spike protein

Protein	Combinations	Mutations	**ΔΔG (pred)**	**C (pred)**
Spike	Independent	D614G	0.422	0.892
	DOUBLE	D614G+S477N	0.228	0.896
		D614G+A222V	0.195	0.889
	TRIPLE	D614G+S477N+A222V	0.129	0.129

Reference

Alturki SO, Alturki SO, Connors J, Cusimano G, Kutzler MA, Izmirly AM, Haddad EK. The 2020 pandemic: current SARS-coV-2 vaccine development. Frontiers in immunology. 2020;11.

Bar-On YM, Flamholz A, Phillips R, Milo R. Science Forum: SARS-CoV-2 (COVID-19) by the numbers. Elife. 2020 Mar 31;9:e57309.

Baum A, Fulton BO, Wloga E, Copin R, Pascal KE, Russo V, Giordano S, Lanza K, Negron N, Ni M, Wei Y. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. Science. 2020 Aug 21;369(6506):1014-8.

Corbett KS, Edwards D, Leist SR, Abiona OM, Boyoglu-Barnum S, Gillespie RA, Himansu S, Schafer A, Ziwawo CT, DiPiazza AT, Dinnon KH. SARS-CoV-2 mRNA Vaccine Development Enabled by Prototype Pathogen Preparedness. Nature, 567–571(2020).

Dearlove B, Lewitus E, Bai H, Li Y, Reeves DB, Joyce MG, Scott PT, Amare MF, Vasan S, Michael NL, Modjarrad K. A SARS-CoV-2 vaccine candidate would likely match all currently circulating variants. Proceedings of the National Academy of Sciences. 2020 Sep 22;117(38):23652-62.

Dong Y, Dai T, Wei Y, Zhang L, Zheng M, Zhou F. A systematic review of SARS-CoV-2 vaccine candidates. Signal transduction and targeted therapy. 2020 Oct 13;5(1):1-4.

Farkas C, Mella A, Haigh JJ. Large-scale population analysis of SARS-CoV2 whole genome sequences reveals host-mediated viral evolution with emergence of mutations in the viral Spike protein associated with elevated mortality rates. medRxiv. 2020 Jan 1.

Fernández A. Structural impact of mutation D614G in SARS-CoV-2 spike protein: enhanced infectivity and therapeutic opportunity. ACS medicinal chemistry letters. 2020 Aug 17;11(9):1667-70.

Gómez-Carballa A, Bello X, Pardo-Seco J, Martinón-Torres F, Salas A. Mapping genome variation of SARS-CoV-2 worldwide highlights the impact of COVID-19 super-spreaders. Genome Research. 2020 Oct 1;30(10):1434-48.

Greaney AJ, Starr TN, Gilchuk P, Zost SJ, Binshtein E, Loes AN, Hilton SK, Huddleston J, Eguia R, Crawford KH, Dingens AS. Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. Cell Host & Microbe. 2020 Nov 19.

Hodgson SH, Mansatta K, Mallett G, Harris V, Emary KR, Pollard AJ. What defines an efficacious COVID-19 vaccine? A review of the challenges assessing the clinical efficacy of vaccines against SARS-CoV-2. The Lancet Infectious Diseases. 2020 Oct 27.

Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu NH, Nitsche A, Müller MA. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell. 2020 Mar 5.

Hu B, Guo H, Zhou P, Shi ZL. Characteristics of SARS-CoV-2 and COVID-19. Nature Reviews Microbiology. 2020 Oct 6:1-4.

Ju B, Zhang Q, Ge J, Wang R, Sun J, Ge X, Yu J, Shan S, Zhou B, Song S, Tang X. Human neutralizing antibodies elicited by SARS-CoV-2 infection. Nature. 2020 Aug;584(7819):115-9.

Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi EE, Bhattacharya T, Foley B, Hastie KM. Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell. 2020 Aug 20;182(4):812-27.

Laha S, Chakraborty J, Das S, Manna SK, Biswas S, Chatterjee R. Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission. Infection, Genetics and Evolution. 2020 Nov 1;85:104445.

Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic acids research. 2019 Jul 2;47(W1):W256-9.

Li G, De Clercq E. Therapeutic options for the 2019 novel coronavirus (2019-nCoV)

Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, Zhao C, Zhang Q, Liu H, Nie L, Qin H. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. Cell. 2020 Sep 3;182(5):1284-94.

Liu L, Wang P, Nair MS, Yu J, Rapp M, Wang Q, Luo Y, Chan JF, Sahi V, Figueroa A, Guo XV. Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. Nature. 2020 Aug;584(7821):450-6.

McAuley AJ, Kuiper MJ, Durr PA, Bruce MP, Barr J, Todd S, Au GG, Blasdell K, Tachedjian M, Lowther S, Marsh GA. Experimental and in silico evidence suggests vaccines are unlikely to be affected by D614G mutation in SARS-CoV-2 spike protein. NPJ vaccines. 2020 Oct 8;5(1):1-5.

Munnink BB, Nieuwenhuijse DF, Stein M, O'Toole Á, Haverkate M, Mollers M, Kamga SK, Schapendonk C, Pronk M, Lexmond P, van der Linden A. Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. Nature medicine. 2020 Sep;26(9):1405-10.

Ou J, Zhou Z, Dai R, Zhang J, Lan W, Zhao S, Wu J, Seto D, Cui L, Zhang G, Zhang Q. Emergence of RBD mutations in circulating SARS-CoV-2 strains enhancing the structural stability and human ACE2 receptor affinity of the spike protein. bioRxiv. 2020 Jan 1.

Pandurangan AP, Ochoa-Montaño B, Ascher DB, Blundell TL. SDM: a server for predicting effects of mutations on protein stability. Nucleic acids research. 2017 Jul 3;45(W1):W229-35.

Pires DE, Ascher DB, Blundell TL. DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. Nucleic acids research. 2014 Jul 1;42(W1):W314-9.

Pires DE, Ascher DB, Blundell TL. mCSM: predicting the effects of mutations in proteins using graph-based signatures. Bioinformatics. 2014 Feb 1;30(3):335-42.

Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, Zhang X, Muruato AE, Zou J, Fontes-Garfias CR, Mirchandani D. Spike mutation D614G alters SARS-CoV-2 fitness and neutralization susceptibility. bioRxiv. 2020 Jan 1.

Poland GA, Ovsyannikova IG, Crooke SN. SARS-CoV-2 vaccine development: current status. InMayo Clinic Proceedings 2020 Jul 30. Elsevier.

Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol 2020. Preprint] July. 2020;15.

Robson F, Khan KS, Le TK, Paris C, Demirbag S, Barfuss P, Rocchi P, Ng WL. Coronavirus RNA proofreading: molecular basis and therapeutic targeting. Molecular Cell. 2020 Aug 4.

Rodrigues CH, Pires DE, Ascher DB. DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. Nucleic acids research. 2018 Jul 2;46(W1):W350-5.

Sahin U, Muik A, Derhovanessian E, Vogler I, Kranz LM, Vormehr M, Baum A, Pascal K, Quandt J, Maurus D, Brachtendorf S. COVID-19 vaccine BNT162b1 elicits human antibody and TH 1 T-cell responses. Nature. 2020 Sep 30:1-9.

Singer J, Gifford R, Cotten M, Robertson D. CoV-GLUE: a web application for tracking SARS-CoV-2 genomic variation.

Sironi M, Hasnain SE, Phan T, Luciani F, Shaw MA, Sallum MA, Mirhashemi ME, Morand S, González-Candelas F. SARS-CoV-2 and COVID-19: A genetic, epidemiological, and evolutionary perspective. Infection, Genetics and Evolution. 2020 May 29:104384.

Starr TN, Greaney AJ, Addetia A, Hannon WH, Choudhary MC, Dingens AS, Li JZ, Bloom JD. Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. bioRxiv. 2020 Jan 1.

Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KH, Dingens AS, Navarro MJ, Bowen JE, Tortorici MA, Walls AC, King NP. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. Cell. 2020 Sep 3;182(5):1295-310.

Statens Serum Institut. Mutations in the mink virus. 5 Nov 2020. https://www.ssi.dk/aktuelt/nyheder/2020/mutationer-i-minkvirus

Tang X, Wu C, Li X, Song Y, Yao X, Wu X, Duan Y, Zhang H, Wang Y, Qian Z, Cui J. On the origin and continuing evolution of SARS-CoV-2. National Science Review. 2020 Mar.

Teng S, Sobitian A, Rhoades R, Liu D, Tang Q. Systemic Effects of Missense Mutations on SARS-CoV-2 Spike Glycoprotein Stability and Receptor Binding Affinity. bioRxiv. 2020 Jan 1.

van Dorp L, Acman M, Richard D, Shaw LP, Ford CE, Ormond L, Owen CJ, Pang J, Tan CC, Boshier FA, Ortiz AT. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. Infection, Genetics and Evolution. 2020 May 5:104351.

Voysey M, Clemens SA, Madhi SA, Weckx LY, Folegatti PM, Aley PK, Angus B, Baillie VL, Barnabas SL, Bhorat QE, Bibi S. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. The Lancet. 2020 Dec 8.

Weisblum Y, Schmidt F, Zhang F, DaSilva J, Poston D, Lorenzi JC, Muecksch F, Rutkowska M, Hoffmann HH, Michailidis E, Gaebler C. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. Elife. 2020 Oct 28;9:e61312.

World Health Organization. Coronavirus disease 2019 (COVID-19): situation report, 82.

World Health Organization: Rolling Updates on Coronavirus Disease (COVID19). Available at: https://www.who.int/ emergencies/diseases/novel-coronavirus2019/events-as-they-happen. Accessed 18 May 2020.

Worldometer, 2020 COVID-19 Coronavirus. Available at: https://www.worldometers.info/coronavirus/ (Cited date Dec 01, 2020).

Zhang L, Jackson CB, Mou H, Ojha A, Rangarajan ES, Izard T, Farzan M, Choe H. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. bioRxiv. 2020 Jan 1.