Emergence of RBD and D614G Mutations in Spike Protein: An Insight from Indian SARS-CoV-2 Genome Analysis

Vinayagam Samyuktha, Venkatesan Naveen Kumar[#]

ImmuGenix Biosciences Pvt Ltd, Perambur, Chennai-600011, Tamil Nadu, India

[#]**Corresponding author**

Dr. V. Naveen Kumar Ph.D.,
Founder & Managing Director,
ImmuGenix Biosciences Pvt Ltd,
No 16/2 Nattal Garden 1st street,
Perambur, Chennai-600011,
Tamil Nadu, India.
Email: naveenkumar.nsd1527@gmail.com

Abstract:

Currently, entire world is crumbled due to COVID-19 caused by novel SARS-CoV-2. Globally, over 5 million people are infected by SARS-CoV-2 with 6% fatality rate. The surface spike (S) protein plays a key role in the pathogenesis of SARS-CoV-2 by mediating viral entry through human angiotensin converting enzyme 2 (hACE2) receptors on the host cell and there is a big global race to find viral neutralizing antibodies and vaccine against S protein of SARS-CoV-2. Since SARS-CoV-2 evolved into 10 different clades in a very short span, a study on sipke protein mutation is essential to have effective vaccine coverage globally. Based on the mutation analysis of S protein from 166 Indian SARS-CoV-2 genome, a total of 40 different SNPs comprising of 14 synonymous and 26 non-synonymous mutations were observed, and notably, Indian S protein diverged into two major clusters, D614 and G614, with 11 different types. Majority of Indian strains fall in A2a and O clusters. Alarmingly, we have observed six SNPs at RBD and notably two of them at RBM (S438F and S494P). S494P SNP, similar to Bat–SARS like-CoV, may indicate a low ACE2 binding affinity. Interestingly 38% of Indian strains harbor a characteristic D614G SNP which was found predominantly in A2a cluster, mostly comprising USA and European strains with high disease severity. The association of disease severity with

D614G SNP is well-correlated in states with high death rate except Maharashtra. Notably, more than 50% of D614G mutation were observed in Northern part of India and 14% in Southern part but not in Kerala and Tamil Nadu strains. Highly conserved motif, D614 (608-VAVLYQ**D**VNCT-618) in upstream and also few downstream, of S1/S2 furin cleavage site may indicate specific key role in efficient interaction with host proteases in pathogenesis. Further studies are warranted to clarify the impact of SD614G SNP association to disease severity . Interestingly, C2367T (Y789Y) synonymous SNP is observed in 37% of Indian strains and notably similar SNPs with degeneracy bases were observed which is a key indication for the possibility of misdiagnosis by Real-Time PCR and revised strategies are needed for the precise diagnosis. Circulation of high number of signature SNPs [D614G and C2367T (Y789Y)] in certain states may be an early indication of emergence of community transmission in India. Further large genome sequence data from India will aid in deep understanding on the diversity of circulating SASR-Cov-2 and its impact on disease severity, origin of imported cases to India, community spread, effect on diagnosis and vaccine coverage.

Keywords: SARS, Spike protein, COVID-19, Mutation, hACE2 Receptor, Real-time PCR, Vaccine

Introduction

Currently, a novel betacoronavirus known as **S**evere **A**cute **R**espiratory **S**yndrome-**Co**rona **V**irus-**2** (SARS-CoV-2) is spreading globally which causes a disease called **Co**rona **Vi**rus **D**isease -20**19** (COVID-19) (Wu et al., 2020a). SARS-CoV-2 first evolved in Wuhan, Hubei province of China in December 2019, and later announced as pandemic in 2020 by WHO, since it started spreading to other countries very effectively (WHO, 2020a). As of 25th May 2020, 5.3 million people were infected globally, 2.1 million recovered and 0.3 million lost life due to COVID-19 (WHO, 2020b). The current global case fatality rate (CFR) is 6% and it varies based on ethnicity, age, and associated co-morbid conditions (Tahir ul Qamar et al., 2020; Bhattacharyya et al., 2020). Based on genome analysis, it is clear that SARS-CoV-2 belong to Sarbecovirus and found to be a sister virus to SARS-CoV since the genome homology is about 89.10% (Shi et al., 2020; Wu et al., 2020b; Tahir ul Qamar et al., 2020). Compared to SARS

(R0- 2to3) and MERS (R0-0.5), COVID-19 (R0-3.5) spreads rapidly and hence referred as super spreader. However, the mortality rate of COVID-19 (6%) is less than that of SARS-CoV (11%) and MERS-CoV (34%) (Lipsitch et al., 2003; Bauch et al., 2005; Cauchemez et al., 2016; WHO, 2020c; Read et al., 2020; Zhao et al., 2020; Majumder and Mandl 2020; Sun et al., 2020). Similar to its preceders, SARS-CoV-2 is also believed to have originated from bats, mainly horseshoe bats (Bat-CoV-RaTG13, *Rhinolophus affinis* from Yunnan Province), and, like civets in SARS-CoV, camels in MERS-CoV, pangolins are considered to be an intermediate host for SARS-CoV-2 (Cui et al., 2019; Zhang et al., 2020; Jia *et al.,* 2020). The genome analysis of SARS-CoV-2 reveals an identity of 96% and 91.02% to Bat-CoV-RaTG13 and Pangolin-CoV, respectively. Interestingly, the spike (S) protein was found to be highly similar to Pangolin-CoV (97.5%) compared to Bat-CoV-RaTG13 (95.4%) (Zhang et al., 2020). Based on these data, studies have predicted that SARS-CoV-2 could be a chimera of different corona viruses that have undergone genetic recombination in a single host at a given time (Zhang et al., 2020).

The genome size of SARS-CoV-2 is around 30Kb and codes for 16 non-structural (nsps), four structural and accessory proteins (Gorden et. al., 2020; Chen et al., 2020b). Non-structural proteins are majorly a replicase complex involved in the replication of virus notably RdRp (RNA-dependent RNA polymerase) which is involved in the replication of positive-sense RNA, papain-like protease (PL^{Pro}), and 3-chymotrypsin-like protease (3CL^{Pro}) mediate proteolytic cleavage of ~ 800 kDa polypeptide upon transcription of the viral genome (Yin et al., 2020; Chen et al., 2020b; Tahir ul Qamar et al., 2020). The surface of enveloped SARS-CoV-2 is embedded by three structural proteins, viz., (i) membrane protein (M) and (ii) envelope protein (E) - responsible for viral assembly and fusion with host cell membrane and (iii) spike protein (S) - plays a key role in the viral pathogenesis since it mediates the entry of virus into the host cell and is also responsible for provoking the immune response in the host (Li, 2016; Chen et al., 2020a; Yin et al., 2020). Nucleocapsid (N) protein is involved in packaging of genetic material into viral capsid and is shown to be a strong immunogen among the structural proteins of the virus (Kumar et al., 2020a).

To initiate the infection, virus uses its S protein to bind with the receptor of the host and subsequently directs its genome into the host cell. This protein comes under class I fusion protein, which contains a characteristic coiled-coil structure in their C-terminals. (Belouzard,

2009; Li, 2016 Wrapp et al., 2020). There are three segments in the S protein which are responsible for viral binding and fusion. (i) A large ectodomain has two subunits S1 and S2 which are involved in binding to the receptor and membrane-fusion, respectively, (ii) a single-pass transmembrane anchor, and (iii) a short intracellular tail. They resemble a clove-shape containing a trimeric head with three S1 subunits and a trimeric S2 stalk. Further, S1 domain is divided into two major domains, (a) N-terminal domain (S1-NTD) and (b) C-terminal domain (S1-CTD) (Li, 2016; Yan et al., 2020; Wrapp et al., 2020). Both the domains have high affinity binding to sugar and protein receptors. These S1 domains binds to the receptor through receptor binding motif (RBM) of the receptor binding domain (RBD), which stabilizes the S2 domain for fusion. (Belouzard, 2009; Li, 2016; Rane, 2020; Zhang et al., 2020; Yin et al., 2020; Xu et al., 2020; Yan et al., 2020; Wrapp et al., 2020)

When the virus enters the host, the S protein is cleaved between S1 and S2 by host furin proteases (TMPRSS2) (Hoffmann et al., 2020b). Subsequently, S1 comes into action by binding through RBD to the human angiotensin converting enzyme 2 (hACE2) receptor on the host cell surface, followed by fusion of both viral and host membranes by the S2 subunits, paving way for the viral genome to enter into the host cell (Li, 2016; Hoffmann et al., 2020a; Walls et al., 2020; Wrapp et al., 2020).

The S protein of SARS-CoV-2 has undergone unique changes when compared to SARS-CoV. Interestingly, the key residues at RBM of SARS-CoV-2 are identical to Pangolin-CoV but differ from SARS-CoV, and Bat-CoV-RaTG13 (Andersen et al., 2020). SARS-CoV-2 have unique additional residues at S1/S2 furin recognition motif (P**RRAR**SV) which is similar to MERS-CoV but different from SARS-CoV, Bat-CoV-RaTG13, and Pangolin-CoV (Millet and Whittaker, 2015; Zhang et al., 2020). In addition to S1/S2 cleavage site, SARS-CoV-2 has a novel cleavage site ((L/S)**KPTKR**S) in the S2 domain (S2') which also aid in membrane fusion and virus infectivity (Belouzard et al., 2009; Hoffmann et al., 2020a)..

Since S protein plays a key role in pathogenesis of SARS-CoV-2, researchers have tried monoclonal antibodies against S protein and have shown a promising neutralization effect on viral entry in *in vitro*, which indicate its ideal candidature for vaccine to prevent COVID-19 (Thanh Le et al., 2020; Okba et al., 2020; Chen et al., 2020a; Yuan et al., 2020). There is a big global race to find viral neutralizing antibodies and vaccine against S protein of SARS-CoV-2

(Thanh Le et al., 2020; Lurie et al., 2020; Kumar et al., 2020b). Considering the short span evolution of 10 different clades of SARS-CoV-2, understanding the variation in key residues of S protein is essential to have effective vaccine coverage globally (Forster et al., 2020; Biswas and Majumder, 2020). In the present study, we have analysed the mutations of S protein from Indian SARS-CoV-2 to understand its diversity, trace the origin of imported cases to India, effect on diagnosis and vaccine coverage.

Materials and Methods

SARS-CoV-2 sequence retrieval

A total of 172 Indian SARS-CoV-2 whole genome RNA sequences were retrieved from GISAID (https://www.gisaid.org/) hCov-19 Database as on 7 May 2020 and these sequences represent strains from 19 different states of India (Shu and McCauley, 2017). Using Bio-Edit version 7.0.9 (Isis Pharmaceuticals), the sequences coding for SARS-CoV-2 S protein were trimmed appropriately by mapping to the S region of the reference genome SARS-CoV-2 Wuhan-Hu-1, China (NC_045512.2) obtained from NCBI (https://www.ncbi.nlm.nih.gov/). Mutations in S protein were analysed after filtering the sequences which had >5N but retained the sequences with degeneracy bases.

Mutation analysis

To determine the single nucleotide polymorphism (SNP), multiple sequence alignment of the Indian SARS-CoV-2 S gene was performed using CLUSTALW along with reference S gene sequences representing different countries by using MEGA (Molecular Evolutionary Genetic Analysis Platform) version X software (Kumar, et al., 2018). Mutation at both nucleotide and amino acid level were documented and further their prevalence in global strains and possible origin were analysed using global SARS-CoV-2 genomic subsampling (5040 genome) in Nextstrain/ncov (https://nextstrain.org/) (Hadfield et al., 2018). The mutation induced significant effects, such as, binding efficiency to ACE2 receptors, host protease cleavage, vaccine coverage, and diagnosis were analysed.

Phylogenetic analysis of spike gene

To uncover the vaccine coverage, the Indian S protein diversity was analysed by multiple sequence alignment of S protein using CLUSTALW and phylogenetic tree was constructed by using the maximum likelihood algorithm in MEGA version X software (Kumar, et al., 2018). Robustness of the tree topology was tested by bootstrapping with 500 replicates.

Results:

SARS-CoV-2 spike protein diversity in India

As of 7th May 2020, a total of 172 Indian SARS-CoV-2 complete and near to complete RNA genome sequences were found in the GISAID hCov-19 Database which encompasses sequences from 19 different states of India with greater part of it from capital Delhi (39), followed by Karnataka (17), Tamil Nadu (15), Gujarat (10), Madhya Pradesh (10), West Bengal (9), Maharashtra (7), Bihar (6), Ladakh (6), Uttar Pradesh (5), Telangana (4) Andhra Pradesh (2), Assam (2), Jammu & Kashmir (2), Kerala (2), Haryana (1), Odisha (1), Punjab (1), Rajasthan (1) and rest from other parts of India (32) **(Fig-1)**. Of the 166 genome analysed after filtering Ns, 31 strains did not have any mutation in S protein, and notably most of them were from Ladakh (6) Tamil Nadu (4), and Delhi (3). Based on the mutation analysis of SARS-CoV-2 S gene, we have identified 40 different SNPs with 14 synonymous and 26 non-synonymous mutations in Indian strains. Of the 40 SNPs, 87.5% of them co-occur with other mutations in S protein. Majority of mutation observed in Indian strains was from A2a and O clades. The complete details of S gene SNPs with occurrence, codon alterations, resultant amino acid changes, global prevalence and clade are shown as synonymous and non-synonymous SNPs in **Tables 1 and 2**.

Synonymous mutation in Indian SARS-CoV-2 spike protein

A total of 14 synonymous SNPs were identified in Indian strains, of which seven mutations were uniquely found to be present only in Indian strains. Remarkably, a C2367T (Y789Y) mutation, in 37% (62) Indian strains and globally one in USA strains, of the subset was analysed. Of these 62 (C2367T) mutations, nine strains also co-occur with other mutations in S protein particularly in association with K77M (3) and A243S (2). Interestingly, four synonymous mutation were also observed at RBD (I410I, G431G, V433V, L518L), and uniquely, two SNPs in Indian strains
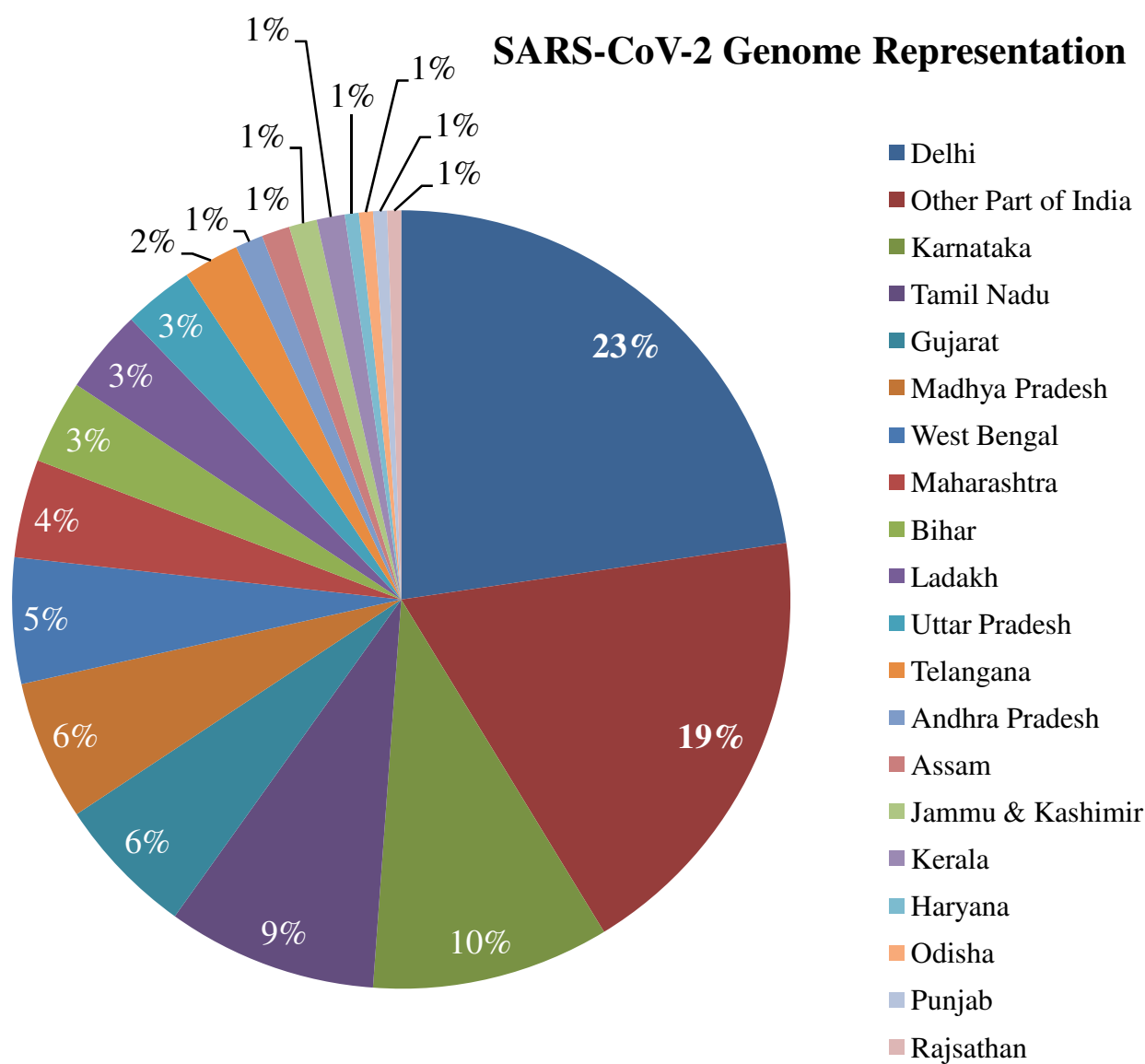
Figure 1. Distribution of SARS-CoV-2 genome from Indian states: Pie chart shows the percentage of Indian SARS-CoV-2 genome in each states from a total of 166 strains.

Table 1. Details of synonymous SNPs in Spike Protein From Indian SARS-CoV-2 strains

Mutations	No of Strains	No of Co-Occurrence Mutation	Codon Alteration	AA Change	Global Observation	SARS-CoV-2 Phylogenetic Clade
C33A	1	1	GTC-GTA	V11V	Only seen in Indian Strain. The same site is mutated in South African strain as C33T	O
C882T	6	6	GAC-GAT	D294D	Also seen in Canada, USA and Turkey	A2a
C1230A	1	-	ATC-ATA	I410I	Only seen in Indian Strain. The same site is mutated in Canadian strain as C1230T	A3
C1293A	1	1	GGC-GGA	G431G	Only seen in Indian Strain	A2a
T1299A	1	1	GTT-GTA	V433V	Only seen in Indian Strain	A2a
A1554T	1	1	CTA-CTT	L518L	Also seen in Belgium, Denmark	A2a
C2073T	1	1	TCC-TCT	S691S	Also seen in USA, Philippines, and Finland	A2a, A3
C2076T	1	1	ATC-ATT	I692I	Also seen Netherlands	A2a
C2367T	62	9	TAC-TAT	Y789Y	Only seen in Indian Strain and also one in USA strain	A2a
C2568T	1	1	AAC-AAT	N856N	Also seen in Australia, Canada, USA, and Argentina	A2a, and A1a
A3132T	1	1	GGA-GGT	G1044G	Also seen in USA, Canada, Mexico, Argentina, Australia, and Singapore	B1
C3342A	1	1	ATC-ATA	I1114I	Only seen in Indian Strain	O
C3342T	1	1	ATC-ATT	I1114I	Also seen in USA, Argentina, Australia, and Saudi Arabia	A2a, A3, B4
C3756T	1	1	TCC-TCT	S1252S	Only seen in Indian Strain	A2a

Table shows the entire data of nucleotide changes, codon alteration and resulted amino acid changes, no of strains harbor similar SNPs, co-mutations, similar observations in global strains and possible origin and the cluster in which given SNP is observed.

Table 2. Details of Non-synonymous SNPs in Spike Protein From Indian SARS-CoV-2 strains

Mutations	No of Strains	No of Co-Occurrence Mutation	Codon Alteration	AA Change	Global Observation	SARS-CoV-2 Phylogenetic Clade
C13T	1	-	CTT-TTT	L5F	Majorly seen in USA, might be associated with travel history	Seen in all clades except A7, A3 and B4
C26A	1	1	CCA-CAA	P9Q	Only seen in Indian Strain. The same site is mutated in one USA strain P9L	O
C52T	1	1	CTT-TTT	L18F	Also seen in Canada and Iceland, might be associated with travel history	B1 and A1a

SNP			Codon	AA change	Description	Cluster
T82C	1	1	TAC-CAC	T82C	Only seen in Indian Strain	O
A230T	3	3	AAG-ATG	K77M	Only seen in Indian Strain. The same site is mutated in Belgium strain K77N (A2a)	O
A671G	1	1	GAA-GGA	E224G	Only seen in Indian Strain	O
G727T	2	2	GCT-TCT	A243S	Only seen in Indian Strain	O
A812G	2	2	CAA-CGA	Q271R	Only seen in Indian Strain	A2a
C821T	1	-	ACT-ATT	T274I	Only seen in Indian Strain	O
G868T	1	1	GAC-TAC	D290Y	Only seen in Indian Strain	O
C896T	1	1	ACA-ATA	T299I	Only seen in Indian Strain	A2a
G1223T	1	1	AGA-ATA	R408I	Only seen in Indian Strain, RBD	O
C1296A	1	1	TGC-TGA	C432*	Only seen in Indian Strain, RBD	A2a
T1301A	1	1	ATA-AAA	I434K	Only seen in Indian Strain, RBD	A2a
C1313T	2	2	TCT-TTT	S438F	Only seen in Indian Strain, RBM	O
T1480C	1	1	TCA-CCA	S494P	Only seen in Indian Strain, RBM	O
G1546T	1	1	GAA-TAA	E516*	Only seen in Indian Strain, RBD	O
A1841G	63	20	GAT-GGT	D614G	Majorly seen in USA and Europe Strains, might be associated with travel history	Majorly seen in A2a, A1a and few in B, B1, B2, and O
G1864A	1	-	GTT-ATT	V622I	Also observed in Canada, might be associated with travel history, in one Indonesian Strain (V622F)	A3 (Canada, India), A6 (Indonesia)
C2168T	1	1	ACC-ATC	T723I	Only seen in Indian Strain	A2a
C2312T	1	1	GCT-GTT	A771V	Also seen one in Belgium, two in Turkey	A2a and A3
C2789T	1	-	GCT-GTT	A930V	Only seen in Indian Strain	B
G2828C	1	1	AGT-ACT	S943T	S943T muatation observed in both Belgium and Indian Strain and a same site found to have unique S943P mutation in 11 Belgium strains	Majorly seen in A2a and one in A2 and Indian strain in A6
G3371T	2	2	GGT-GTT	G1124V	Only observed in five Australian strain, possible origin could be from Australia	Majorly seen in A2a, one in A1a and one in B
G3749T	1	1	TGT-TTT	C1250F	Observed in two strains, one each in India and Vietnam	A2a (Indian) and A1a (Vietnam)
C3788T	1	1	CCA-CTA	P1263L	Majorly in UK, also seen in USA, Sweden, Australia, Poland, Scotland and one in Hong Kong	A2a

Table shows the entire data of nucleotide changes, codon alteration and resulted amino acid changes, no of strains harbor similar SNPs, co-mutations, similar observations in global strains and possible origin and the cluster in which given SNP is observed.
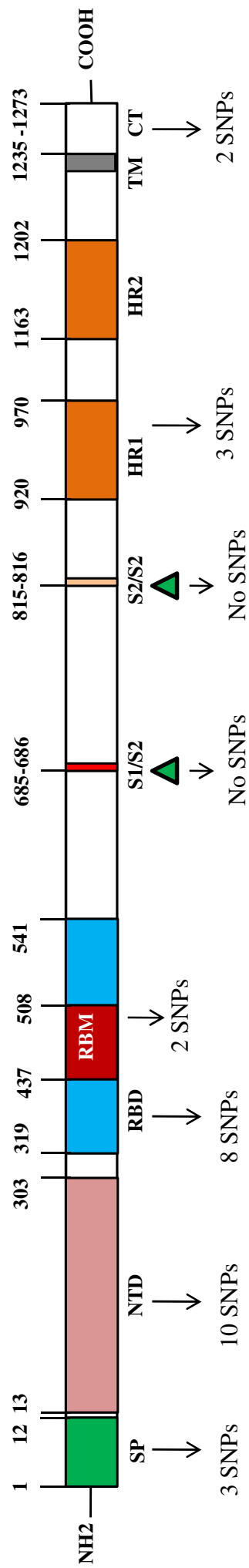
Figure 2: Structure of SARS-CoV-2 Spike protein: Schematic representation of full length spike protein indicating the positions of each domains, motifs, and host proteases cleavage sites in the proteins and corresponding amino acid positions (Based on Uniport annotation). The down arrows indicate the number of mutations accrued in Indian SARS-CoV-2 Spike protein (Both synonymous and non-synonymous).

(C33A [V11V] and C1230A [I410I]) were also identically observed in strains of other countries, South Africa (C33T [V11V]) and Canada (C1230T [I410I]), but with different nucleotide degeneracy. An extra seven synonymous mutations were also observed in other countries like USA, Canada, Turkey, Belgium, Denmark, Philippines, Finland, Mexico, Argentina, Australia, Singapore, Saudi Arabia and Netherlands (**Table 1**). Majorly all synonymous SNPs are prevalent in A2a cluster, and few others in A1a, A2, B1, and B4 clusters.

Non-synonymous SNPs in Indian spike protein

A total of 26 non-synonymous SNPs were identified in Indian strains, of which 17 mutations were novel uniquely found only in Indian strains (**Table 2**). Most of non-synonymous SNPs were observed at RBD (10), NTD (10), SP (3), CT (3), and HR1(2) and interestingly no mutations were observed at S1/S2 or S2' host protease cleavage site (**Fig-2**). Among 26 SNPs, two (D614G and S494P) of them are with unique features. Non-synonymous mutations were widely distributed among different cluster of SARS-CoV2 but majorly seen in A2a, followed by A1a and A3.

 (i) **Signature D614G mutation**

Based on Nextstrain analysis, a unique D614G signature mutation was observed in 38% (64) of Indian strains which is a characteristic SNP majorly found in A2a cluster comprising mostly USA and European strains with limited Asian and Asia Pacific strains. Of these 63 (D614G) mutations, 20 strains also co-occur with other mutations in S protein especially in association with D294D (6), Q271R (2) S438F (2), and G1124V (2). Interestingly, D614G mutation was not observed in early Indian genome sequences associated with travel from Wuhan or other Asian countries. Notably more than 50% of this mutation was observed in northern and 14% in southern part of India but not in Kerala and Tamil Nadu strains. The **figure 3** shows how D614G is confined to A2a cluster and its global spread from a subset of global genome data in Nextstrain/ncov. Interestingly, the region containing D614 (608-VAVLYQ**D**VNCT-618) was found to be highly conserved domain in both human- (SARS-CoV-1 and SARS-CoV-2) and animal-associated (Bat-CoV and Pangolin-CoV) corona virus (**Fig-4**).
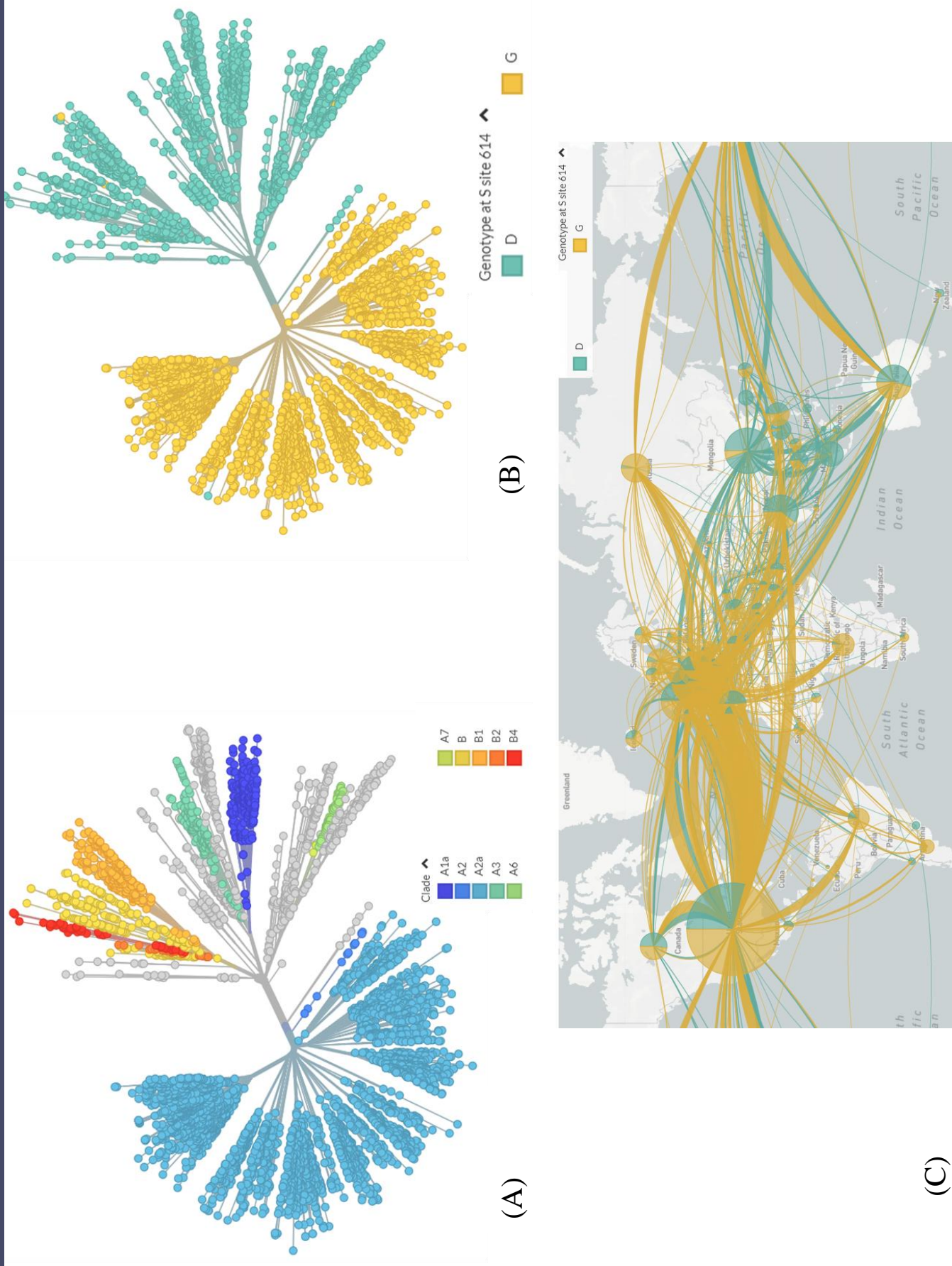
Figure 3. Details of D614G Mutation: (A) Shows 10 different clades of SARS-CoV-2 (B) Shows D614G SNP majorly confined to A2a clade (D614-Green and G614-Yellow). (C) Shows the global spread of D614G by a subset of global genome data in Nextstrain/ncov. (The data is obtained from Nextstrain/ncov. [https://nextstrain.org/])

Figure 4. Conserved region of D614 and RS S1/S2 cleavage site: Amino acid sequence alignment of Spike protein with conserved residues at D614 found in human and animal origin CoV (608-VAVLYQ**D**VNCT-618) is shown in green box and mutated Indian strains D614G is indicated by arrow. Both upstream and down stream of S1/S2 cleavage site (Shown in Red Box) has conserved residues which indicate the specific function of these motifs highlighted in Black Bold letter on top.
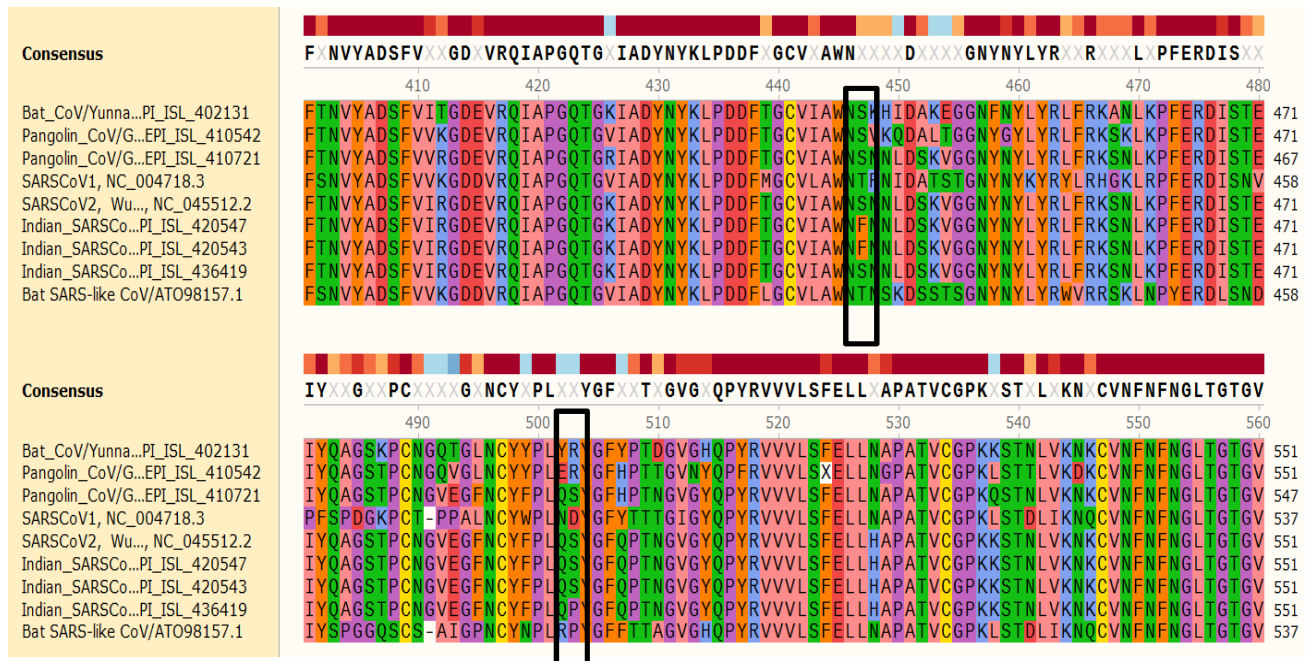
.

Figure 5. Mutation in RBM of Indian Spike protein: Amino acid sequence alignment of spike protein from SARS-CoV, Bat-CoVs, Pangolin-CoVs and Indian SARS-CoV-2 mutated and unmutated strains. Two mutations at RBM (S438F, and S494P) are shown in black vertical boxes. One Indian strain had S494P SNP and this residue was found to be similar to Bat-SARS-like CoV.
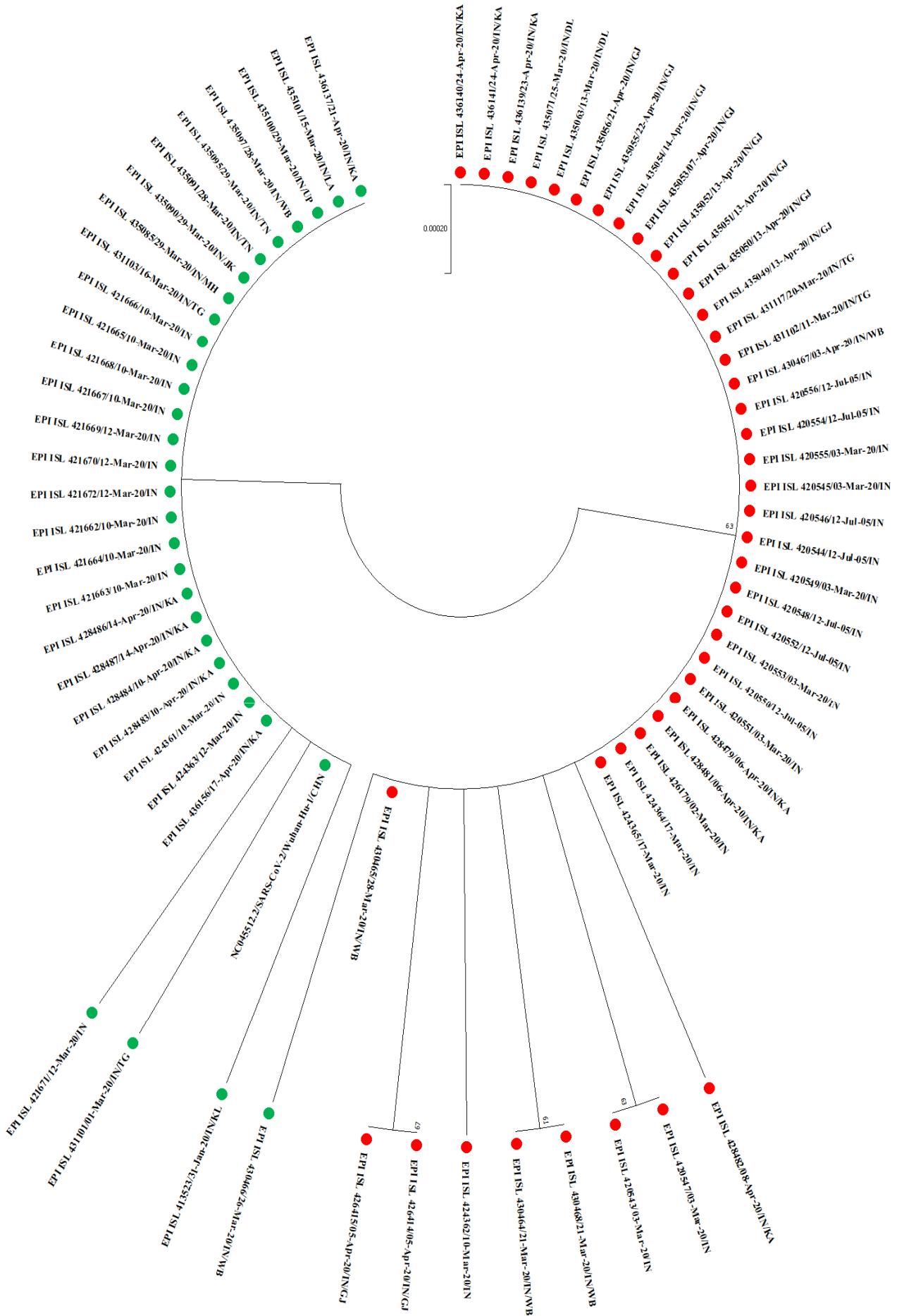
Figure 6. Indian spike protein cluster: Cluster generated using MEGA X, showing two different clusters of Indian spike proteins D614 (Green) and G614 (Red).

(ii) Novel SNPs in RBD of the spike protein from Indian SARS-CoV-2

Alarmingly, we have observed six SNPs at RBD (R408I, C432*, I434K, S438F, S494P, and E516*), of which two was observed at RBM (S438F and S494P). Of the six SNP, three (C432*, I434K and S438F) mutations were observed in single strain EPI_ISL_424362. The change in S438F and S494P at RBM is shown in **figure 5** along with other CoV-related S proteins. The impact of other RBD mutations is needed to study further in order to unravel the crucial role of these sites for interacting with host ACE2.

Wuhan-Hu-1 genotype and other interesting SNPs in India

Interestingly, 19% (31) of the India strains were found to be similar to Wuhan genotype, among them 41% of them were from Tamil Nadu (13) and 19% from Ladakh (6). Based on 72 Indian strains S protein analysis, we found it was diverged into two different cluster D614 and G614 with 11 different variants as shown in **figure 6**. Of the global subset analysed, the S943P was found to be a unique mutation site (11 strains) in Belgium strains in A2a clade and the same site was mutated to S943T one each in Indian (A6 clade) and Belgium strain. The P9Q, K77M, and V622I mutation sites in Indian strains were also mutated in USA (P9L), and Belgium (K77N) Indonesia (V622F) strains, respectively, with different amino acid substitutions.

Discussion

In the current study based on the mutation analysis in 166 Indian SARS-CoV-2 genome from GISAID hCov-19 Database (Shu and McCauley, 2017), it is revealed that S protein have 40 different SNPs with 14 synonymous and 26 non-synonymous SNPs which help us to understand the possible origin of imported SARS-CoV-2, unique strains in circulation, associated impact on Real-Time PCR diagnosis and vaccine coverage for Indian strains. Among the synonymous SNPs, seven were confined to India and interestingly C2367T (Y789Y) mutation was observed predominantly (37%) in Indian strains and one in USA strain (EPI ISL 436898) of the subset. Although C2367T SNP did not alter the amino acid (Y789Y), it made a signature SNP which aided us to understand the strain circulation within the community. Notably, C2367T (Y789Y) SNP was found in association with K77M in 3 strains, each one from Delhi, Bihar, and Tamil Nadu and A243S in two strains each one from Delhi, and Bihar, this probably indicate possible COVID-19 infection in association with travel (Eden et al., 2020). Receptor binding domain

expressed four synonymous SNPs (I410I, G431G, V433V, L518L) which indicate that these site are prone to mutate further and may lead to alteration in the binding property of S protein to ACE2 receptor. Seven synonymous mutations were uniquely observed in some of the western countries majorly from A2a cluster and, remarkably six Indian strains had C882T (D294D) SNPs which were majorly seen in Canada, USA, and Turkey, which probably indicate the imported SARS-CoV-2 cases to India (Eden et al., 2020). From the present study, we could understand that same synonymous mutation sites, (C33A [V11V] and C1230A [I410I]), in Indian strains differ from South African (C33T [V11V]) and Canadian strains (C1230T [I410I]) with degeneracy. Hence, there is a possibility of misdiagnosis by Real-Time PCR due to inefficient primer binding in certain SARS-CoV-2 strains which were imported from specific countries (Brufsky, 2020).

Among 26 non-synonymous SNPs, 17 mutations were novel, uniquely found only in Indian strains. It has been shown that D614G characteristic SNP was majorly observed in A2a cluster which mostly comprised USA and European strains and is believed to have an impact on COVID-19 severity (Phan, 2020; Brufsky, 2020; Stefanelli et al., 2020; Bhattacharyya et al., 2020; Biswas and Majumder, 2020). Interestingly, D614G signature mutation was observed in 38% (64) of Indian strains and was found to be high in states with more prevalence and CFR of COVID-19. Korber et al. (2020) showed high number of D614G mutation in ICU patients. Similarly, the association of D614G SNP and disease severity is well correlated in Delhi, Gujarat, Madhya Pradesh, and West Bengal, where CFR is high. But in contrast, this association was not observed in Maharashtra, which encountered the high death rate, and this may be due to the limited genome analysis from the states (**Fig-7**). D614G mutation was not observed in early Indian genome sequence associated with travel from Wuhan or other Asian countries. More than 50% of this mutation was observed in northern and 14% in southern part of India but not in Kerala and Tamil Nadu strains. It was observed that D614G SNP had signature co-mutation association with specific states and this may probably indicate a possibility of COVID-19 infection in association with travel and also community circulation of unique strain (Stefanelli et al., 2020; Eden et al., 2020). This signature SNP (D614G) site is located between the RBD and S1/S2 polybasic protease cleavage site of S protein and their effect on ACE2 receptor binding or host-mediated furin cleavage are need to be analysed to understand their role in enhanced
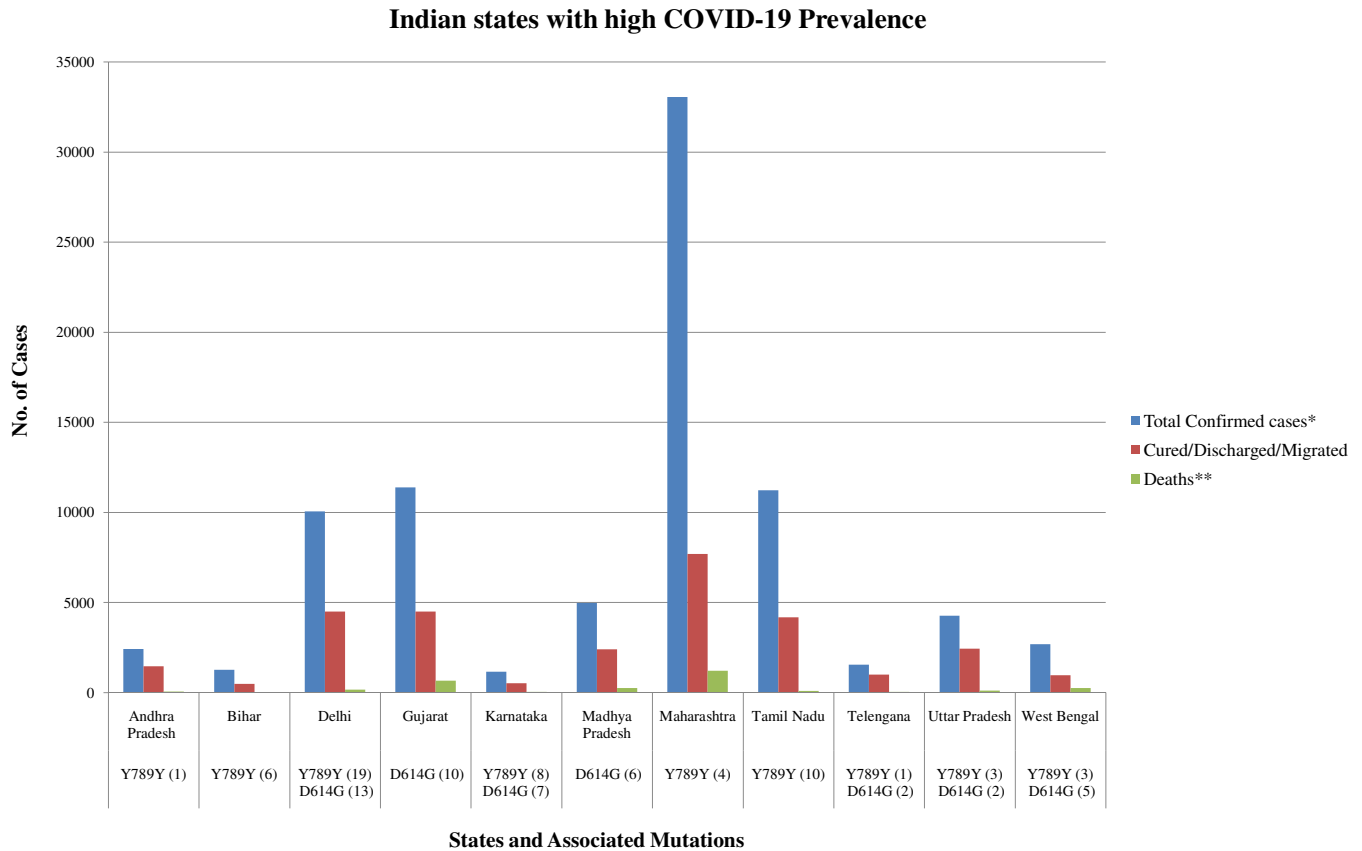
Figure 7. Graph of COVID-19 in India: Graph showing Indian states with total number of confirmed (*Including foreign Nationals), and discharged cases with deaths (**more than 70% cases due to co-morbidities). Below each states mentioned is associated SNPs in Spike protein. (INDIA COVID-19 Data acquired from Ministry of Health and Family Welfare, GOI, as on 18 May 2020, 08:00 IST [https://www.mohfw.gov.in/])

pathogenicity in USA and Europe patients. It was found that the region contain D614 (608-VAVLYQ**D**VNCT-618) in upstream and few downstream regions of S1/S2 showed highly conserved motif among corona virus from different species which may indicate specific key role in efficient interaction with host proteases in viral pathogenesis (Walls et al., 2020; Brufsky, 2020). The study by Bhattacharyya et al. (2020) showed that D614G SNP in this motif resulted in additional host-mediated cleavage of S1 by Cathepsin-L and elastase. Belouzard et al. (2009) have shown that cleavage of addition proteolytic site by cathepsin L in SARS-CoV enhances the membrane fusion. This probably indicate the possible mechanism of D614G SNP association with disease severity of COVID-19. However, we need further studies to clearly elucidate the impact of D614G SNP associated with disease severity.

The present study clearly traces the possible origin of imported Indian SARS-CoV-2 by unique SNPs confined to specific geographic location (Tables 1 and 2), which represented K77M SNPs only for Indian Strain, and same site is mutated in Belgium strain K77N (A2a). Similarly, a unique S943P mutation was observed only in Belgium strains and the same was also observed one in each India and Belgium strain, with different amino acid (S943T) in the later strain (Korber et al., 2020). These finding unraveled the travel associated origin of SARS-CoV-2 to India (Eden et al., 2020).

Alarmingly, we have observed six novel SNPs at RBD (R408I, C432*, I434K, S438F, S494P, and E516*) of the S protein from Indian SARS-CoV-2, from which two was at RBM (S438F and S494P) and majority of them belong to O clade. The residues at RBM are crucial for the interaction of hACE2 receptor. Based on the S protein alignment of SARS-CoV-2, SARS-CoV, BatCoV RaTG13, Pangolin-CoV and Bat–SARS-like CoV, it was found that S494 is a conserved key amino acid in RBM of SARS-CoV-2 to interact with ACE2 and it is identical with Pangolin-CoV (Andersen et, al., 2020). A novel mutation in Indian strain at RBM (S494P) was found to be similar to that of Bat–SARS-like CoV and studies in SARS-CoV on equivalent position variation have shown reduced ACE2 binding affinity and antigenicity (Chakraborti et al., 2005; Watabe and Kishino, 2010; Rockx et al., 2010; Wu et al., 2012). Similarly, the SNPs at R408I and N439K (adjacent to S438F) in S protein shown to have weak interaction with receptors (Jia et al., 2020). The impact of other RBD mutations is needed to study further in order to unravel the crucial role of these sites for interacting with host ACE2.

Effect of synonymous SNPs on real-time PCR diagnosis

Considering the fact that synonymous mutation does not show any effect on conformational change and function of S protein, but will adversely affect the diagnosis of SARS-CoV-2 by Real-Time PCR if any of the primers' 3' prime binding sites fall on the mutational sites (Tahamtan and Ardebili, 2020; Xi et al., 2020). This clearly indicate the reason for some reported cases of Real-Time PCR negative with Chest X-ray positive (Lei et al., 2020). The primer mismatching due to SNPs may lead to false negative results, especially for the confirmatory test of SARS-CoV-2 since primary screening of target E gene is highly conserved, and thus, suspected patients will not be quarantined which may result in spreading of the disease unknowingly in the community (Zhou et al., 2020; Corman et al., 2020). So, for better detection by Real-Time PCR, more conserved region must be considered in three or four target genes or different regions with in the same gene which would be ideal to reduce the false negative results due to their increased efficiency in detection (Zhou et al., 2020; Corman et al., 2020; Tahamtan and Ardebili, 2020; Wu et al., 2020c). In such case, even if mutation occurs in any one of the target region or gene, other two or three will be positive, so COVID-19 patients will be less likely to be misdiagnosed. Also combination of real-time RT-PCR and clinical features by Chest X-ray or CT will facilitate enhancement in SARS-CoV-2 diagnosis (Xi et al., 2020; Tahamtan and Ardebili, 2020; Wang et al., 2020).

Impact on global vaccine coverage and convalescent plasma therapy

Even though different strategies have been tried on the global race of SARS-CoV-2 vaccine discovery, a majority of them were targeting S protein (Korber et al., 2020). In such case, non-synonymous SNPs will adversely affect the global vaccine coverage for SARS-CoV-2 (Wrapp et al., 2020). So, SNPs studies from different geographical regions are crucial to bring in very effective pan vaccine for COVID-19 (Watabe and Kishino, 2010; Jia et al., 2020; Yin, 2020). More studies on the S protein SNPs will also throw a light on requirement of multivalent like Pneumococcal vaccines (Daniels et al., 2016) or seasonal vaccine like Influenza (Fiore et al., 2009). Some studies have shown that convalescent plasma therapy is ineffective in COVID-19 (Zeng et al., 2020; Brown and McCullough, 2020), and this may be due to the variants of S protein which resulted in poor viral neutralization (Marano et al., 2016; Rockx et al., 2010;

Wrapp et al., 2020). Considering this, it would be a better strategy if we do a strain matching by S protein sequencing between donor and recipients before plasma therapy, like HLA matching in organ transplantation (Sheldon and Poulton, 2006).

Early indication of community spread

In the present study, we could observe that two mutations C2367T (Y789Y, 37%) and D614G (38%) majorly from A2a clusters were predominant in India compared to other mutations. The D614G was found to be in association with D294D synonymous SNP in six strains, interestingly five of them were from Delhi and one from Gujarat (Gandhinagar), similarly two SNPs both from Gujarat (Q271R), and West Bengal (Kolkata) (G1124V) uniquely co-occur with D614G. More than 50% of the D614G SNP was found to be prevalent in northern part of India majorly from Delhi, Gujarat, Madhya Pradesh and West Bengal, which indicated distinctive strains circulation in the local community. A unique C2367T (Y789Y) mutation was confined only to India and was majorly observed in states with high prevalence of COVID-19, such as, Delhi, Tamil Nadu, Karnataka, Bihar and Maharashtra, which may be an early indication of emergence of community transmission in India (Stefanelli et al., 2020; Eden et al., 2020).

Conclusion

The present study clearly shows that Indian strains have mutation at RBD and RBM majorly linked to O clade, also an A2a clade associated D614G SNP similar to USA and Europe. The impact of these mutations on pathogenesis of SARS-CoV-2 is needed to elucidate with further studies either by site-directed mutagenesis or viral strains harboring these mutations. Certain characteristic SNPs on S protein clearly unraveled the possible origin of Indian strains. Circulation of these signature SNPs in certain states may be an early indication of emergence of community spread in India. More SARS-CoV-2 genome sequencing from India and other part of the world would aid in unraveling the mutations in the different region of genome and its impact on ACE2 receptor binding, S1/S2 cleavage by host proteases, disease severity associated with specific strain, drug resistance, Real-Time PCR diagnosis and global vaccine coverage. Understanding these factors would greatly help us to effectively combat these new super spreading invaders (SARS-Cov-2) into human community.

Author contributions:

NKV conceptualized and designed the study. SV and NKV obtained and processed the sequence data from GISAID and NCBI database. SV and NKV performed the analysis and interpretation of mutation in Indian strains and compared with subset from Nextstrain/ncov database. SV and NKV drafted, read and approved the manuscript.

Conflicts of Interest:

No conflict of interest.

REFERENCES

Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., & Garry, R. F. (2020). The proximal origin of SARS-CoV-2. Nature Medicine, 26(4), 450–452. https://doi.org/10.1038/s41591-020-0820-9

Bauch CT, Lloyd-Smith JO, Coffee MP, Galvani AP. Dynamically modeling SARS and other newly emerging respiratory illnesses: past, present, and future. Epidemiology. 2005;16(6):791-801. doi:10.1097/01.ede.0000181633.80269.4c

Belouzard, S., Chu, V. C., & Whittaker, G. R. (2009). Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. Proceedings of the National Academy of Sciences of the United States of America, 106(14), 5871–5876. https://doi.org/10.1073/pnas.0809524106

Bethany L. Brown, Jeffrey McCullough, Treatment for emerging viruses: Convalescent plasma and COVID-19, Transfusion and Apheresis Science, 2020, https://doi.org/10.1016/j.transci.2020.102790.

Bhattacharyya, C., Das, C., Ghosh, A., & Singh, A. K. (2020). Global Spread of SARS-CoV-2 Subtype with Spike Protein Mutation D614G is Shaped by Human Genomic Variations that Regulate Expression of TMPRSS2 and MX1 Genes.

N. K. Biswas, P. P. Majumder, Analysis of RNA Sequences of 3636 SARS-CoV-2 Collected from 55 Countries Reveals Selective Sweep of One Virus Type. Indian J. Med. Res, Epub ahead of print DOI: 10.4103/ijmr.IJMR_1125_20 (2020).

Brufsky, A. (2020) Distinct Viral Clades of SARS-CoV-2: Implications for Modeling of Viral Spread. 0–1. https://doi.org/10.1002/jmv.25902

Cauchemez S, Nouvellet P, Cori A, Jombart T, Garske T, Clapham H, Moore S, Mills HL, Salje H, Collins C, Rodriquez-Barraquer I, Riley S, Truelove S, Algarni H, Alhakeem R, AlHarbi K, Turkistani A, Aguas RJ, Cummings DA, Van Kerkhove MD, Donnelly CA, Lessler J, Fraser C, Al-Barrak A, Ferguson NM. Unraveling the drivers of MERS-CoV transmission. Proc Natl Acad Sci U S A. 2016 Aug 9;113(32):9081-6. doi: 10.1073/pnas.1519235113. Epub 2016 Jul 25. PMID: 27457935; PMCID: PMC4987807

Chakraborti, S., Prabakaran, P., Xiao, X., & Dimitrov, D. S. (2005). The SARS coronavirus S glycoprotein receptor binding domain: Fine mapping and functional characterization. Virology Journal, 2, 1–10. https://doi.org/10.1186/1743-422X-2-73.

Chen, W.H., Hotez, P.J., and Bottazzi, M.E. (2020a). Potential for developing a SARS-CoV receptor-binding domain (RBD) recombinant protein as a heterologous human vaccine against coronavirus infectious disease (COVID)-19. Hum Vaccin Immunother, 1-4.

Chen, Y., Liu, Q., & Guo, D. (2020b). Emerging coronaviruses: Genome structure, replication, and pathogenesis. Journal of Medical Virology, 92(4), 418–423. https://doi.org/10.1002/jmv.25681

Corman, V. M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D. K. W., Bleicker, T., Brünink, S., Schneider, J., Schmidt, M. L., Mulders, D. G. J. C., Haagmans, B. L., Van Der Veer, B., Van Den Brink, S., Wijsman, L., Goderski, G., Romette, J. L., Ellis, J., Zambon, M., … Drosten, C. (2020). Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Eurosurveillance, 25(3). https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045

Cui, J., Li, F., and Shi, Z.-L. (2019). Origin and evolution of pathogenic coronaviruses.Nat. Rev. Microbiol. 17, 181–192.

Daniels CC, Rogers PD, Shelton CM. A Review of Pneumococcal Vaccines: Current Polysaccharide Vaccine Recommendations and Future Protein Antigens. J Pediatr Pharmacol Ther. 2016;21(1):27-35. doi:10.5863/1551-6776-21.1.27

Eden, J., Rockett, R., Carter, I., Ligt, J. De, Hadfield, J., Storey, M., Ren, X., Tulloch, R., Basile, K., Wells, J., Byun, R., Gilroy, N., Sullivan, M. V. O., Sintchenko, V., Chen, S. C., Sorrell, T. C., Holmes, E. C., Dwyer, D. E., Kok, J., & Group, S. (2020). An emergent clade of SARS-CoV-2 linked to returned travellers from Iran. 6(1), 1–4. https://doi.org/10.1093/ve/veaa027

F. Wu, S. Zhao, B. Yu, et al. A new coronavirus associated with human respiratory disease in China Nature, 579 (2020b), pp. 265-269

Fiore AE, Bridges CB, Cox NJ. Seasonal influenza vaccines. Curr Top Microbiol Immunol. 2009;333:43-82. doi:10.1007/978-3-540-92165-3_3

Forster, P., Forster, L., Renfrew, C., & Forster, M. (2020). Phylogenetic network analysis of SARS-CoV-2 genomes. Proceedings of the National Academy of Sciences of the United States of America, 117(17), 9241–9243. https://doi.org/10.1073/pnas.2004999117

Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, et al. A SARS-CoV-2-human protein-protein interaction map reveals drug targets and potential drug repurposing. bioRxiv 2020. https://doi.org/10.1101/2020.03.22.002386.

Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, Neher RA. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics. 2018 Dec 1;34(23):4121-4123. doi: 10.1093/bioinformatics/bty407. PMID: 29790939; PMCID: PMC6247931.

Hoffmann, M., Kleine-Weber, H., & Pöhlmann, S. (2020a). A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. Molecular Cell, 78(4), 779-784.e5. https://doi.org/10.1016/j.molcel.2020.04.022

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.-H., Nitsche, A., et al. (2020b). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell 181, 271-280.e278.

J. Sun, W. T. He, L. Wang, A. Lai, X. Ji, X. Zhai, G. Li, M. A. Suchard, J. Tian, J. Zhou, M. Veit, S. Su, COVID-19: Epidemiology, Evolution, and Cross-Disciplinary Perspectives. Trends Mol. Med. (2020), doi:10.1016/j.molmed.2020.02.008.

Jia, Y., Shen, G., Zhang, Y., Huang, K.-S., Ho, H.-Y., Hor, W.-S., Yang, C.-H., Li, C., & Wang, W.-L. (2020). Analysis of the mutation dynamics of SARS-CoV-2 reveals the spread history and emergence of RBD mutant with lower ACE2 binding affinity. BioRxiv, 2020.04.09.034942. https://doi.org/10.1101/2020.04.09.034942

Jonathan M. Read, Jessica R.E. Bridgen, Derek A.T. Cummings, Antonia Ho, C. P. J. (2020). Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. Antimicrobial Agents and Chemotherapy, 58(12), 7250–7257. https://doi.org/10.1101/2020.01.23.20018549

Korber, B., Fischer, W., Gnanakaran, S. G., Yoon, H., Theiler, J., Abfalterer, W., Foley, B., Giorgi, E. E., Bhattacharya, T., Parker, M. D., Partridge, D. G., Evans, C. M., Silva, T. de, LaBranche, C. C., Montefiori, D. C., & Group, S. C.-19 G. (2020). Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. BioRxiv, 2020.04.29.069054. https://doi.org/10.1101/2020.04.29.069054

Kumar S, Stecher G, Li M, Knyaz C, and Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution 35:1547-1549

Kumar, A., Parveen, A., Kumar, N., Bairy, S., & Kaushik, V. (2020b). Characterization of nucleocapsid (N) protein from novel coronavirus SARS-CoV-2. May. https://doi.org/10.20944/preprints202005.0413.v1

Kumar, G.V., Jeyanthi, V., and Ramakrishnan, S. (2020a). A short review on antibody therapy for COVID-19. New Microbes New Infect, 100682.

Lei, P., Fan, B., Mao, J., & Wang, P. (2020). Multiple parameters required for diagnosis of COVID-19 in clinical practice. Journal of Infection, 80(6), e27–e28. https://doi.org/10.1016/j.jinf.2020.03.016

Li, F. (2016). Structure, Function, and Evolution of Coronavirus Spike Proteins. Annual Review of Virology, 3(1), 237–261. https://doi.org/10.1146/annurev-virology-110615-042301

Lipsitch, M. et al. (2003) 'Transmission dynamics and control of severe acute respiratory syndrome', Science, 300(5627), pp. 1966–1970.

Lurie N, Saville M, Hatchett R, Halton J. Developing Covid-19 Vaccines at Pandemic Speed. N Engl J Med. 2020;382(21):1969-1973. doi:10.1056/NEJMp2005630.

Majumder, M., & Mandl, K. D. (2020). Early Transmissibility Assessment of a Novel Coronavirus in Wuhan, China. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3524675

Marano G, Vaglio S, Pupella S, et al. Convalescent plasma: new evidence for an old therapeutic tool?. Blood Transfus. 2016;14(2):152-157. doi:10.2450/2015.0131-15

Millet, J. K., & Whittaker, G. R. (2015). Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. Virus Research, 202, 120–134. https://doi.org/10.1016/j.virusres.2014.11.021

Okba, N. M. A., Muller, M. A., Li, W., Wang, C., GeurtsvanKessel, C. H., Corman, V. M., Lamers, M. M., Sikkema, R. S., Bruin, E. de, Chandler, F. D., Yazdanpanah, Y., Hingrat, Q. Le, Descamps, D., Houhou-Fidouh, N., Reusken, C. B. E. M., Bosch, B.-J., Drosten, C., Koopmans, M. P. G., & Haagmans, B. L. (2020). SARS-CoV-2 specific antibody responses in COVID-19 patients. MedRxiv, 2020.03.18.20038059. https://doi.org/10.1101/2020.03.18.20038059

Phan T. Genetic diversity and evolution of SARS-CoV-2. Infect Genet Evol. 2020;81:104260. https://doi.org/10.1016/j. meegid.2020.104260 PMID: 32092483

Qing-Lei Zeng, Zu-Jiang Yu, Jian-Jun Gou, Guang-Ming Li, Shu-Huan Ma, Guo-Fan Zhang, Jiang-Hai Xu, Wan-Bao Lin, Guang-Lin Cui, Min-Min Zhang, Cheng Li, Ze-Shuai Wang, Zhi-Hao Zhang, Zhang-Suo Liu. 2020. Effect of Convalescent Plasma Therapy on Viral Shedding

and Survival in Patients With Coronavirus Disease 2019, The Journal of Infectious Diseases, , jiaa228, https://doi.org/10.1093/infdis/jiaa228

Rane, J. S., Chatterjee, A., Kumar, A., & Ray, S. (2020). Targeting SARS-CoV-2 Spike Protein of COVID-19 with Naturally Occurring Phytochemicals : An in Silco Study for Drug Development. April. https://doi.org/10.26434/chemrxiv.12094203.v1

Rockx, B., Donaldson, E., Frieman, M., Sheahan, T., Corti, D., Lanzavecchia, A., & Baric, R. S. (2010). Escape from Human Monoclonal Antibody Neutralization Affects In Vitro and In Vivo Fitness of Severe Acute Respiratory Syndrome Coronavirus. The Journal of Infectious Diseases, 201(6), 946–955. https://doi.org/10.1086/651022

Sheldon S, Poulton K. HLA typing and its influence on organ transplantation. Methods Mol Biol. 2006;333:157-174. doi:10.1385/1-59745-049-9:157

Stefanelli, P., Faggioni, G., Lo Presti, A., Fiore, S., Marchi, A., Benedetti, E., Fabiani, C., Anselmo, A., Ciammaruconi, A., Fortunato, A., de Santis, R., Fillo, S., Capobianchi, M. R., Gismondo, M. R., Ciervo, A., Rezza, G., Castrucci, M. R., Lista, F., Di Martino, A., … Riccardo, F. (2020). Whole genome and phylogenetic analysis of two SARSCoV-2 strains isolated in Italy in January and February 2020: Additional clues on multiple introductions and further circulation in Europe. Eurosurveillance, 25(13), 1–5. https://doi.org/10.2807/1560-7917.ES.2020.25.13.2000305

Tahamtan A, Ardebili A. Real-time RT-PCR in COVID-19 detection: issues affecting the results. Expert Rev Mol Diagn. 2020;20(5):453-454. doi:10.1080/14737159.2020.1757437

Tahir ul Qamar, M., Alqahtani, S. M., Alamri, M. A., & Chen, L. L. (2020). Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants. Journal of Pharmaceutical Analysis. https://doi.org/10.1016/j.jpha.2020.03.009

Thanh Le, T., Andreadakis, Z., Kumar, A., Gómez Román, R., Tollefsen, S., Saville, M., & Mayhew, S. (2020). The COVID-19 vaccine development landscape. Nature Reviews. Drug Discovery, 19(5), 305–306. https://doi.org/10.1038/d41573-020-00073-5

Walls, A.C., Park, Y.-J., Tortorici, M.A., Wall, A., McGuire, A.T., and Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 181, 281-292.e286.

Wang, Y., Kang, H., Liu, X., & Tong, Z. (2020). Combination of RT-qPCR Testing and Clinical Features For Diagnosis of COVID-19 facilitates management of SARS-CoV-2 Outbreak. J Med Virol, PG-10.1002/jmv.25721, 10.1002/jmv.25721. https://doi.org/10.1002/jmv.25721

Watabe, T., & Kishino, H. (2010). Structural considerations in the fitness landscape of a virus. Molecular Biology and Evolution, 27(8), 1782–1791. https://doi.org/10.1093/molbev/msq056

WHO. (2020a) WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020 (2020), (available at https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks- at-the-media-briefing-on-covid-19---11-march-2020).

WHO. (2020b) Coronavirus Disease 2019 (COVID-19) Situation Report-126. May 25 2020. WHO Situat. Rep. 2019 (2020), pp. 1–17.

WHO. (2020c) Statement on the meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV) - WHO, January 23, 2020

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., Graham, B. S., & McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science, 367(6483), 1260–1263. https://doi.org/10.1126/science.aax0902

Wu, J., Liu, J., Li, S., Peng, Z., Xiao, Z., Wang, X., Yan, R., & Luo, J. (2020c). Detection and analysis of nucleic acid in various biological samples of COVID-19 patients. Travel Medicine and Infectious Disease, 101673. https://doi.org/10.1016/j.tmaid.2020.101673

Wu, K., Peng, G., Wilken, M., Geraghty, R. J., & Li, F. (2012). Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. Journal of Biological Chemistry, 287(12), 8904–8911. https://doi.org/10.1074/jbc.M111.325803

Wu, Y., Ho, W., Huang, Y., Jin, D. Y., Li, S., Liu, S. L., Liu, X., Qiu, J., Sang, Y., Wang, Q., Yuen, K. Y., & Zheng, Z. M. (2020a). SARS-CoV-2 is an appropriate name for the new coronavirus. The Lancet, 395(10228), 949–950. https://doi.org/10.1016/S0140-6736(20)30557-2

Xi M, Wei Q, Qihua F, et al. Understanding the influence factors in viral nucleic acid test of 2019 novel coronavirus (2019-nCoV). Chin J Lab Med. 2020;43:E002–E002.

Xu, X., Chen, P., Wang, J., Feng, J., Zhou, H., Li, X., Zhong, W., & Hao, P. (2020). Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. Science China Life Sciences, 63(3), 457–460. https://doi.org/10.1007/s11427-020-1637-5

Y. Shu, J.J.E. McCauley. GISAID: Global initiative on sharing all influenza data–from vision to reality Euro Surveill., 22 (2017), 10.2807/1560-7917.ES.2017.22.13.30494

Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., & Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science, 367(6485), 1444–1448. https://doi.org/10.1126/science.abb2762

Yin, C. (2020). Genotyping coronavirus SARS-CoV-2: methods and implications. Genomics, 19, 1–12. https://doi.org/10.1016/j.ygeno.2020.04.016

Yuan, M., Wu, N.C., Zhu, X., Lee, C.-C.D., So, R.T.Y., Lv, H., Mok, C.K.P., and Wilson, I.A. (2020). A highly conserved cryptic epitope in the receptor-binding domains of SARS-CoV-2 and SARS-CoV. Science, eabb7269.

Z.-L. Shi, P. Zhou, X.-L. Yang, et al. Discovery of a Novel Coronavirus Associated with the Recent Pneumonia Outbreak in Humans and its Potential Bat Origin (2020), 10.1101/2020.01.22.914952

Zhang, T., Wu, Q., & Zhang, Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. Current Biology, (2020) 30, 1346–1351. https://doi.org/10. 1016/j.cub .2020.03.022.

Zhao, S., Lin, Q., Ran, J., Musa, S. S., Yang, G., Wang, W., Lou, Y., Gao, D., Yang, L., He, D., & Wang, M. H. (2020). Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. International Journal of Infectious Diseases, 92, 214–217. https://doi.org/10.1016/j.ijid.2020.01.050

Zhou, yunying, Pei, F., Wang, L., Zhao, H., Li, H., Ji, M., Yang, W., Wang, Q., Zhao, Q., & Wang, Y. (2020). Sensitivity evaluation of 2019 novel coronavirus (SARS-CoV-2) RT-PCR detection kits and strategy to reduce false negative. MedRxiv, 105, 2020.04.28.20083956. https://doi.org/10.1101/2020.04.28.20083956