


## SARS-CoV-2 mutations: the biological trackway towards viral fitness

Parinita Majumdar<sup>1</sup> and Sougata Niyogi<sup>2</sup> <sup>1</sup>Independent Researcher, Kolkata, India and <sup>2</sup>Dinabandhu Andrews Institute of Technology and Management, Block-S, 1/406A, Patuli, Kolkata, West Bengal 700094, India

## Review

**Cite this article:** Majumdar P, Niyogi S (2021). SARS-CoV-2 mutations: the biological trackway towards viral fitness. *Epidemiology and Infection* **149**, e110, 1–8. <https://doi.org/10.1017/S0950268821001060>

Received: 6 January 2021

Revised: 26 March 2021

Accepted: 27 April 2021

**Key words:**

Fitness; Mutation; SARS-CoV-2; Transmission; Virulence

**Authors for correspondence:** ParinitaMajumdar, E-mail: [parineeta.majumdar@gmail.com](mailto:parineeta.majumdar@gmail.com); Sougata Niyogi, E-mail: [sniyogi10@gmail.com](mailto:sniyogi10@gmail.com)**Abstract**

The outbreak of pneumonia-like respiratory disorder at China and its rapid transmission world-wide resulted in public health emergency, which brought lineage B betacoronaviridae SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) into spotlight. The fairly high mutation rate, frequent recombination and interspecies transmission in betacoronaviridae are largely responsible for their temporal changes in infectivity and virulence. Investigation of global SARS-CoV-2 genotypes revealed considerable mutations in structural, non-structural, accessory proteins as well as untranslated regions. Among the various types of mutations, single-nucleotide substitutions are the predominant ones. In addition, insertion, deletion and frame-shift mutations are also reported, albeit at a lower frequency. Among the structural proteins, spike glycoprotein and nucleocapsid phosphoprotein accumulated a larger number of mutations whereas envelope and membrane proteins are mostly conserved. Spike protein and RNA-dependent RNA polymerase variants, D614G and P323L in combination became dominant world-wide. Divergent genetic variants created serious challenge towards the development of therapeutics and vaccines. This review will consolidate mutations in different SARS-CoV-2 proteins and their implications on viral fitness.

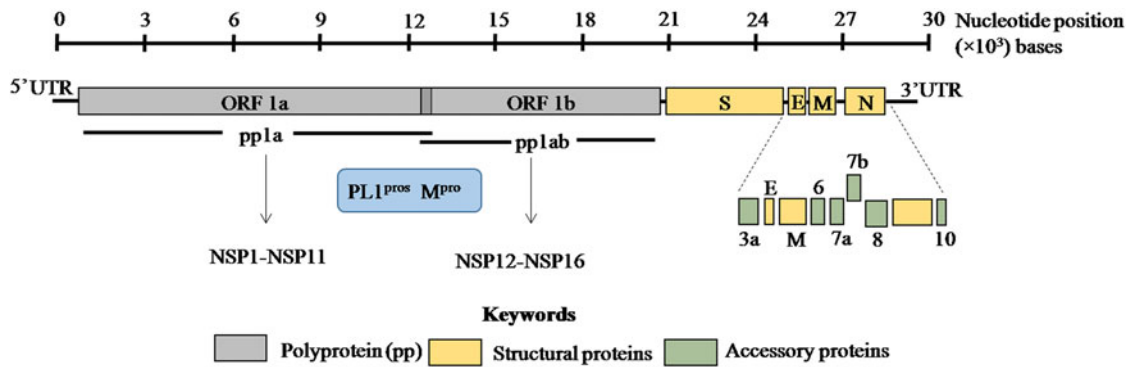
**Introduction**

The emergence of pneumonia with unknown aetiology at Wuhan province of China in December 2019, eventually led to the identification of a novel strain of human coronavirus (CoV) named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) based on its genetic relatedness with SARS-CoV, the causative agent of severe acute respiratory syndrome outbreak in 2002 [1–4]. High transmission dynamics and overwhelming infection rate of SARS-CoV-2 resulted in declaration of COVID-19 (*Coronavirus Disease 2019*) pandemic on 11th March 2020 by the World Health Organization (WHO) (<https://www.who.int>). The infectivity of SARS-CoV-2 is distinctly higher among the members of betacoronaviridae with a comparatively lower case fatality rate (CFR) of 1.4–2.1% compared to SARS-CoV (9.6%) and MERS-CoV (middle east respiratory syndrome coronavirus) (40%) [5, 6]. Several studies had highlighted the association of different lockdown strategies, viral testing capabilities and varied demographic compositions with the severity of COVID-19 pandemic [7–10]. Since 1960s with the discovery of first human CoV until date, altogether seven human CoVs are identified [1, 3, 4]. Among these seven strains, SARS-CoV, MERS-CoV and SARS-CoV-2 are associated with acute human respiratory disorder whereas the remaining four strains 229E, OC43, NL63 and HKU1 showed mild clinical symptoms including sore throat, nasal discharge, fever and cough [2, 11]. The average mutation rate of  $4 \times 10^{-4}$  nucleotide substitutions/site/year is largely, if not exclusively, responsible for the genetic diversity of betacoronaviridae [12]. In addition to mutation, frequent recombination and interspecies transmission are also common among them [11]. These factors largely account for temporal change in their infectivity and virulence. Recent studies highlighted the implication of mutations in the rapid community transmission of SARS-CoV-2- and COVID-19-associated mortality [13]. In order to understand the evolutionary trend in SARS-CoV-2, it is of utmost importance to study the mutation patterns and their effect on viral fitness. The current review aims to provide a comprehensive knowledge on SARS-CoV-2 mutations and their impact on the major viral proteins associated with viral life-cycle, pathogenicity and virulence.

**Genome organisation of SARS-CoV-2**

The viral genome is non-segmented, single-stranded positive sense RNA, ~30 kb in size with 5' and 3' untranslated regions (UTRs) (Fig. 1) [1, 2, 14]. Genome analysis of SARS-CoV-2 revealed 79% and 50% identity with SARS-CoV and MERS-CoV, respectively [1, 2]. Moreover, 88% homology was observed with two bat coronaviruses, bat-SL-CoVZC45 and bat-SL-CoVZXC21 suggesting a plausible bat origin of SARS-CoV-2.

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Fig. 1.** ORF1a and ORF1b encode two overlapping poly-proteins pp1a and pp1ab which are proteolytically processed into 16 non-structural proteins (NSP1-NSP16) by the main protease (M<sup>pro</sup>) and papain-like proteases (PL1<sup>pros</sup>). The scale bar on the top denotes the nucleotide position of the genome.

SARS-CoV-2 genome encodes ORF1a/ORF1b (open reading frame) polyproteins and four structural proteins including S (spike), E (envelope), M (membrane) and N (nucleocapsid) with several intervening ORFs encoding accessory proteins [2, 14] (Fig. 1). Among these ORFs, ORF1a and ORF1b at the 5' terminus comprise 2/3rd of the genome and encode two overlapping poly-proteins pp1a and pp1ab [1-3] (Fig. 1). These poly-proteins undergo proteolytic cleavage by the viral main protease (M<sup>pro</sup>) which has at least 11 conserved cleavage sites and papain-like proteases (PL<sup>pros</sup>) to generate 16 non-structural proteins (Fig. 1) [15, 16]. These non-structural proteins have multi-faceted role in viral replication, transcription, morphogenesis as well as evasion of host immune response. On the contrary, accessory proteins are not crucial for viral life cycle but play important role in viral pathogenesis [17]. The biological functions of these structural, non-structural and accessory proteins in SARS-CoV-2 are discussed in Table 1.

### Mutations in SARS-CoV-2 genome

Since its emergence in 2019, SARS-CoV-2 infection had become widespread with 126 210 104 confirmed cases in more than 200 countries with a death toll of 2 769 638 as on 26th March 2021 (<https://www.who.int>). Following the sequencing of SARS-CoV-2 genome at Wuhan in December 2019, more than 10 000 genetic variants are reported [8-10]. Recently an emergent variant of SARS-CoV-2, VUI202012/01 (variant under investigation, year 2020, month December, variant 01) or VOC202012/01 (variant of concern) or B.1.1.7 in the United Kingdom with an enhanced transmissibility of 56-70% became a major concern [47, 48]. The variant strain with 14 non-synonymous mutations and three deletions transcend the existing variants at London, East and South East England [47]. The rapid spread of COVID-19 among individuals of different ages, genetic compositions and medical predispositions provides suitable mutagenic backdrop for generation of heterogeneous SARS-CoV-2 population.

### Predominant mutation clusters in SARS-CoV-2 genome

An average of  $\geq 11$  mutations per sample with the insurgence of single-nucleotide substitutions was reported for SARS-CoV-2 [8, 49]. These mutations are categorised as amino acid changing SNP (single-nucleotide polymorphism), amino acid changing triplet, 5' UTR-SNP and silent SNP. Notably, C  $\rightarrow$  T (55.1%)

transition was more common than A  $\rightarrow$  G (14.8%) transition and G  $\rightarrow$  T transversion had an occurrence of 12%. SNP variants are classified into six clusters based on the pattern of co-mutation [10]. Cluster I includes 3037C>T; NSP3:F106F (non-structural protein3:F106F) and 14408C>T; RdRp:P323L, cluster II includes 3037C>T, 14408C>T and 23403A>G; S:D614G, cluster III includes 14408C>T, cluster IV includes 3037C>T, 14408C>T, 23403A>G, 28881G>A; N:R203K, 28882G>A; N:R203K, 28883G>C; N:G204R, cluster V includes 3037C>T, 14408C>T, 23403A>G and 25563G>T; ORF3a:Q57H and cluster VI includes 8782C>T; NSP4:S76S, 28144T>C; ORF8:L84S [8, 10]. Among these six clusters, clusters III, IV and VI were predominant in Asian countries whereas clusters IV, V and VI were prevalent in the United States. In addition to SNPs, in-frame deletions and short frame-shift deletions were also observed among the genetic variants with a very low frequency of 0.6% and 0.8% respectively. However, insertion mutation was extremely rare with <0.1% among all the mutations [10].

Based on the specific mutation patterns, the genetic variants of SARS-CoV-2 are classified into three major phylogenetic clades: G, S and V. The clade G, S and V comprise variants of S: D614G (23403A>G), ORF8:L84S (8782C>T) and ORF3a:G251V (26144G>T), respectively [8] (Table 2). Clade G and V variants comprise amino acid changing SNPs whereas clade S variant include silent SNP. Clade G has two offspring, GH and GR based on the emergence of nascent mutations, in addition to the already existing one. GR clade has a combination of spike D614G and nucleocapsid RG203KR mutations, prevalent in Europe and South America while GH comprises mutations in spike D614G and ORF3a Q57H which predominates in North America.

### Mutation in RNA-dependent RNA polymerase

Variants of RNA-dependent RNA polymerase (RdRp) emerged early during the COVID-19 outbreak in Europe, North America, China and Asian countries and hence was considered as a mutation hotspot [10, 57]. A total of 607 mutations are reported in RdRp of which 14408C>T (P323L) mutation which lies near the interface domain of RdRp showed highest frequency (10 925 times in 15 140 genotypes) [10] (Table 2). This variant of RdRp did not alter the catalytic activity but is likely to abrogate the interaction with its cofactors and existing anti-viral drugs [57]. Crystal structure analysis revealed that RdRp (NSP12) forms a complex with NSP7 and NSP8 which provide processivity to the polymerase [26]. However, specific residues involved in

**Table 1.** Functions of various SARS-CoV-2 proteins

Non-structural proteins	Functions	Reference
NSP1	Interacts with 40S ribosome and inhibits host translation. Degrades host mRNA and facilitates viral gene expression. Evasion of host immune response	18, 19
NSP2	Viral replication	20, 21
NSP3	Proteolytic cleavage of replicase poly-protein at its N terminus. Participates in viral replication by assembly of cytoplasmic double membrane vesicle. De-ubiquitinates cellular proteins tagged with Lys48 and Lys63-linked poly-ubiquitin chain. Type I interferon mediated immune response antagonist. Blocks NF-kappa- $\beta$ signal transduction	16, 20, 22
NSP4	Assembly of cytoplasmic double membrane vesicle and helps in viral replication	23, 24
NSP5	Proteolytic cleavage of replicase poly-protein at its C terminus	15
NSP6	Triggers autophagosome formation from host endoplasmic reticulum	Swiss-model repository ( <a href="https://swissmodel.expasy.org/">https://swissmodel.expasy.org/</a> ) 25
NSP7	Cofactor of RdRp	16, 26
NSP8	Cofactor of RdRp	16, 26
NSP9	Binds single-stranded RNA and participates in viral replication	27
NSP10	Methylates the 5' cap structure of viral mRNA	28, 29
NSP11	Not identified	
NSP12	Replication and transcription	26
NSP13	Helicase, nucleoside triphosphatase, have 5' RNA triphosphatase activity and potent interferon antagonising activity	30–32
NSP14	Cleaves single-stranded and double-stranded RNA from 3' to 5' end and has N7-guanine methyl-transferase activity. Exoribonuclease activity, interferon antagonising activity	32–34
NSP15	Harbours endo-ribonuclease activity, interferon antagonising activity	32, 35
NSP16	Possesses nucleoside-2' O-methyl-transferase activity	36
Structural proteins		
S	Binds to ACE2 host cell receptor and mediates viral entry within the host cell	2, 37
E	Maturation of virion, forms viroporin on host membrane and facilitate ion transport	38
M	Maintains spherical membrane curvature of the virus, stabilises nucleocapsid and facilitates viral assembly, antagonises type I and III interferon responses	39
N	Encapsidates viral nucleic acid	40
Accessory proteins		
ORF3a	Induces apoptosis, helps viral entry, blocks STAT1 and inhibits IFN activity	41, 42
ORF6	Antagonises interferon signalling by blocking nuclear entry of STAT1 via Rae1 and Nup98	43
ORF8	Immune evasion by down-regulating the surface expression of MHC I	44
ORF10	Ubiquitin ligase, interacts with CUL2 and degrades host proteins	45
ORF7a	Interacts with CD14+ monocytes and triggers aberrant inflammatory responses, inhibits STAT2 and antagonises IFN	46
ORF7b	Inhibits both STAT1, STAT2 and blocks IFN stimulated gene expression	42

their interaction remain unresolved. Unlike RNA viruses, RdRp of CoVs has proof reading activity, a characteristic of *Nidovirales*, which is conferred by 3'  $\rightarrow$  5' exonuclease ExoN/NSP14 [60]. An *in vitro* biochemical assays could detect interactions between NSP12-NSP7-NSP8 and ExoN/NSP14. Such an interaction is necessary for the excision of wrongly incorporated bases from nascent RNA.

The 14408C>T (P323L) mutation was found to be associated with increasing point mutations in viral isolates in Europe during the early phase of COVID-19 outbreak. Thus, it is possible that

mutations in RdRp might alter the interaction of RdRp with these cofactors which could render the proofreading activity less effective leading to the emergence of numerous SARS-CoV-2 variants [57]. *In silico* analysis predicted the docking site of anti-viral drugs within a hydrophobic cleft located near the 14408C>T mutation site [57]. This mutation was predicted to diminish the affinity of RdRp for existing anti-viral drugs. Mutation in the catalytic domain of RdRp, D484Y resulted in remdesivir resistance, the first anti-viral drug used in the United States [61]. Thus, the emergence of RdRp genetic variants in SARS-CoV-2 posed

**Table 2.** Different mutations in SARS-CoV-2 proteins

Protein name	Non-synonymous amino acid mutations	Reference
Spike protein	P323L, A97V, T141I, A449V, D63Y, Q239K, V341I, A435S, K458R, I472V, H519P, A831V, S943T, N439K, L452R, A475V, V483A, F490L, Y508H, V1176F, S4777N, F32I, H49Y, S247R, N354D D614G + V341I, D614G + K458R, D614G + I472V, D614G + A435S N501Y, P681H, A570D, T716I, S982A, D1118H, Δ69/Δ70, Δ144/145	10, 47, 50–54
Nucleocapsid protein	R203K, G204R, P13L, S188L, S202N, D103Y, I292T, S194L, S197L, T339I, T148I, P344S	10, 54, 55
Membrane protein	T175M, D3G, C64Y, S4F, R158C, I52T, I76F, T7I, F193L, G78C	10, 56
Envelop protein	T9I, V24M, V58F, L73F	10, 56
RdRp	P323L, A97V, T141I, A449V, D63Y	10, 57
ORF3a	G251V, W128L, L127I, Q57H, W131C, L129F, D173Y, H93Y, P25L, T175I, L94F, K16N, W149L	56, 58
ORF6	P57L, T21I	56
ORF7a	E92D, M1R, L5F, A8S, Y20N, R78H, A105S, A106S	56
ORF8	Q91K, Q72H, P36S, I9T, P30S, R52T, E106Q, A65V, F120L, I121L, R101L, G66S, Q72H, L84S	56, 59
ORF10	D31Y	56

tremendous challenge towards the efficacy of anti-viral therapeutics.

### Mutation in spike protein

Spike glycoprotein mediates viral entry within the host cell by interacting with the membrane-bound angiotensin-converting enzyme 2 (ACE2) and plays a remarkable role in SARS-CoV-2 infectivity and transmissibility [37, 62]. A 1273 amino acid containing spike protein can be divided into S1 and S2 subunits [63]. The C terminal domain of S1 in SARS-CoV-2 harbours the receptor binding domain (RBD) and the residues 442–487 are crucial for interaction with the host cell receptor [62]. S2 subunit is crucial for mediating host–viral membrane fusion [37, 63]. Mutations are continuously being reported for S gene having 1004 unique mutations among 15 140 genotypes and found out to be the second most non-conserved protein in SARS-CoV-2 after nucleocapsid protein [10]. Notably, mutations are more frequent in S1 unit and in past few months, almost half of the amino acid residues in RBD had been mutated creating a major challenge for vaccine development. Mutations in S protein have multiple consequences including altered protein stability, receptor affinity and sensitivity to neutralising monoclonal antibody (mAb) as well as convalescent serum [50, 51, 64]. R408I mutation stabilising S protein was reported in an Indian strain [64]. Among all S protein variants, D614G increased at an alarming rate which was observed 10 969 times in 15 140 genome isolates, suggesting a positive selection of this variant during the course of viral evolution [10]. D614G variant was highly transmissible and became predominant in Europe, Canada, Australia and United States [65]. Moreover, this particular variant of SARS-CoV-2 was more infectious and found to be associated with enhanced mortality across the world [13]. Structural analysis revealed D614G mutation favours open conformation of S protein which facilitates binding with the host receptor thereby enhances its infectivity [66]. Two new variants, V1176F and S4777N are also associated with higher mortality and found to spread rapidly across the world [50]. V1176F arose independently and also co-occurred with D614G. *In silico* analysis predicts V1176F variant could facilitate the interaction with ACE2 by stabilising spike protein

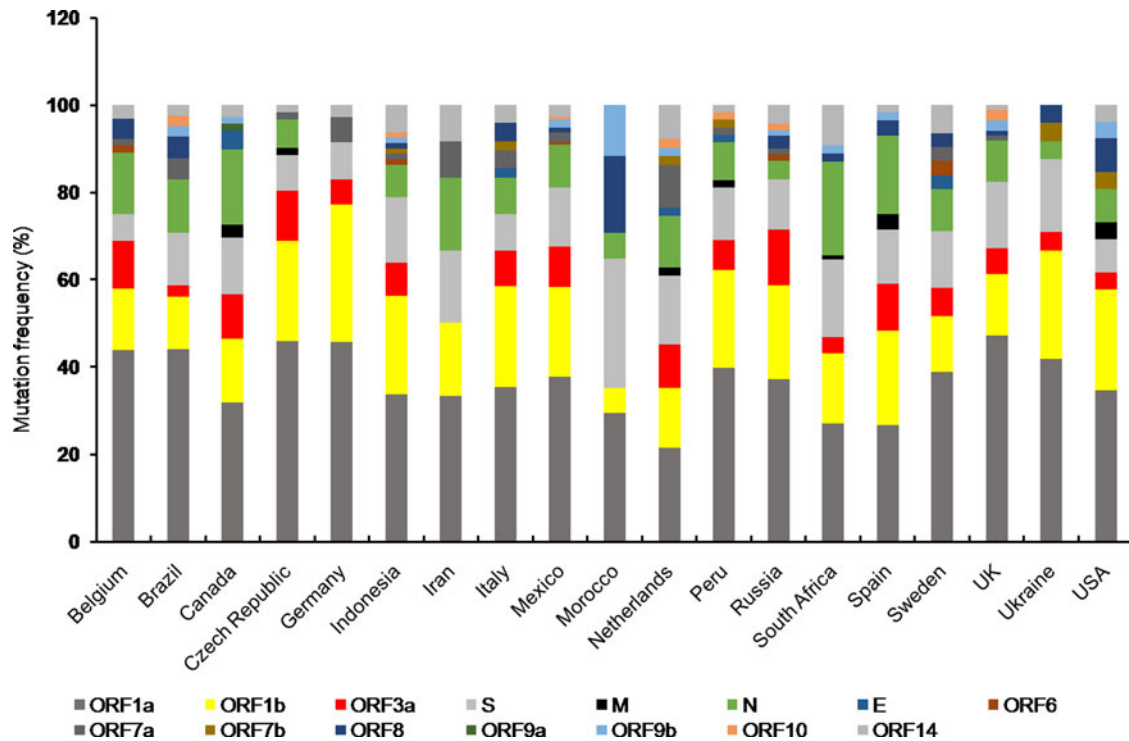
trimeric complex. The co-mutations D614G + V341I, D614G + K458R and D614G + I472V fall within the RBD of S protein and enhance the infectivity of virus by favouring binding with the host receptor [51].

VUI202012/01 had eight mutations in S protein of which N501Y, P681H, Δ69 and Δ70 have potential implications on viral infectivity [47, 52]. N501 is one of the six key residues mediating contact with the host cell receptor [37]. N501Y falls within the RBD and had been shown to enhance the binding affinity of S protein with human ACE2 [47]. Deletion of two amino acids at positions 69 and 70 of S protein is likely to be associated with host immune evasion and increased infectivity [52]. The furin cleavage site near S1/S2 is a unique feature of SARS-CoV-2 and is linked with viral infectivity [63]. P681H mutation lies near the furin cleavage site and might interfere with viral infectivity and transmission [47]. In addition to these mutations, A570D (RBD), Δ144/145 (S1 subunit), T716I, S982A and D1118H (S2 subunit) are also reported in VUI202012/01 [52]. The precise role of these mutations in viral life cycle and pathogenesis is currently under investigation.

The S2 unit comprises of fusion peptide (FP), heptad repeat 1 (HR1), HR2, transmembrane domain and cytoplasmic domain [63]. The insertion of four amino acids upstream of HR1 at positions 681–684 increases the length and flexibility of the connecting region between the FP and HR1 [67]. This favours viral entry within the host and also serves as a genetic determinant of SARS-CoV-2 pathogenicity. Several mutations including A475V, N439K, L452R, F490L, V483A and Y508H in S protein resulted in decreased sensitivity to mAb [51–53, 65–67]. The antigenic properties of S protein had already been exploited in vaccine development. Thus, it is crucial to understand the evolution of S protein antigenicity by studying their mutation patterns and subsequent implications on viral pathogenesis.

### Genetic determinant of SARS-CoV-2 virulence and N protein mutation

Nucleocapsid phosphoprotein has multi-faceted role in SARS-CoV-2 life cycle including replication of viral genome, assembly of mature virions and encapsidation of viral nucleic



**Fig. 2.** Stacked bar chart shows frequency distribution of mutations at various SARS-CoV-2 ORFs from indicated countries as of 29th December 2020. Mutations in SARS-CoV-2 proteins for respective countries were obtained from NextStrain open source project (<https://nextstrain.org/ncov>). Mutation frequency was calculated by dividing the number of mutations for a particular protein with total number of mutations corresponding to all the proteins for a given country, multiplied by 100.

acid [68]. The positively charged amino acid residues in the N terminal domain of nucleocapsid protein (46–176 amino acids) and serine/arginine-rich linker region (184–204 amino acids) are important for interaction with viral RNA [69, 70]. The C terminal dimerisation domain also facilitates RNA binding. Moreover, N protein helps to unwind viral RNA following infection through phosphorylation of specific amino acid residues involved in such RNA–protein interaction. Any mutation affecting the phosphorylation sites of N protein is likely to interfere with viral life cycle. R203K, G204R, P13L, D128D, L139L, S188L, S202N, D103Y and I292T mutations are more frequently observed in N protein [10] (Table 2). However, the biological implications of these mutations warrant further investigation.

An enrichment of positively charged amino acid within the NLS (nuclear localisation signal) of nucleocapsid proteins compared to the less harmful CoVs including HKU1, NL63, OC43 and 229E is considered as one of the genetic determinants of SARS-CoV-2 pathogenicity [67]. Such enrichment is also present in SARS-CoV and MERS-CoV nucleocapsid proteins indicating convergent evolution. The abundance of positively charged residues is expected to strengthen the nuclear localisation of N protein and thereby facilitates its interaction with viral as well as host proteins [67]. Thus, mutations strengthening the NLS of N protein could affect its subcellular localisation and subsequent interaction with host proteins.

#### Co-mutations in SARS-CoV-2

SARS-CoV-2 variants with certain co-mutations became prevalent world-wide compared to single mutation suggesting their

fitness [66]. NSP3:F106F (3037C>T) mutation co-evolved with RdRp:P323L, S:D614G, N:R203K, N:G204R and ORF3a:Q57H mutations and these strains with co-mutations were predominant in Russia, United States and Europe [10, 71]. Since 3037C>T mutation is silent and does not have major impact on NSP3 protein *per se*, it may change codon usage and thereby might affect the translation efficiency of NSP3 [8]. Mutations in NSP3 had been linked with positive selection of viruses leading to evolution in betacoronaviruses [72]. Interestingly, 3037C>T, 14408C>T and 23403A>G co-mutations had the highest number of descendants world-wide indicating positive selection of this epidemiologically dominant SARS-CoV-2 variants. In addition to this co-mutation, a novel non-synonymous mutation NSP3:S1515F (4809C>T) was observed only in Indian strains early in March 2020 [71]. NSP3 interacts with nucleocapsid protein and tethers the nascently translated replicase–transcriptase complex to the viral genome during the early stages of infection in SARS-CoV [73]. *In silico* analysis predicts this mutation as a stabilising one and it is intriguing to address whether this mutation strengthens the interaction of N protein with the replicase–transcriptase complex favouring viral infection.

#### Mutations in accessory proteins

Mutations are found in all the accessory proteins of SARS-CoV-2 with varying frequency (Fig. 2). Among the accessory proteins, ORF3a and ORF8 are brought into limelight due to the rapid spread of cluster V (NSP3:F106F, RdRp:P323L, S:D614G and ORF3a:Q57H) and VI (NSP4:S76S and ORF8:L84S) [10]. Mutation in ORF3a was associated with a higher CFR in the

COVID-19 pandemic [56]. Among 51 non-synonymous mutations in ORF3a, Q57H (17.4%) and G251V (9.7%) were predominant ones [58] of which Q57H mutation was found to cause disease severity in hospitalised [74]. Moreover, Q57H mutation co-occurred with either of W131C, L129F and D173Y second site mutations [58]. ORF3a is the largest accessory protein (~30 kDa) in SARS-CoV-2 which elicits host inflammatory responses through activating innate immune receptor NLRP3 (NOD, LRR and pyrin domain containing 3) inflammasome [75]. This results in uncontrolled release of pro-inflammatory cytokines and other inflammatory mediators including tumour necrosis factor, interleukin-6, leukotrienes and prostaglandins, leading to cytokine storm, the clinical characteristic of SARS-CoV-2 pathogenesis [75, 76]. Mutations in ORF3a are predicted to cause loss of B cell epitopes thereby affects antigenicity of ORF3a [56]. Since ORF3a was predicted to interact with the host signalling pathways including JAK- STAT, chemokine and cytokine-related pathways, it is possible that ORF3a variants could aggravate host immune response leading to the varied severity of COVID-19 among infected individuals.

ORF8 is most divergent in SARS-CoV-2 with no paralogues or orthologues outside lineage B betacoronaviruses [59]. This suggests that ORF8 might play an important role in lineage specific adaptation of betacoronaviruses within the host [17]. SARS-CoV-2 ORF8 down-regulates MHCI expression on the surface of antigen-presenting cells which facilitates viral infection by evasion of host immune response [44, 77]. Mutational analysis revealed ORF8 locus is subjected to point mutations, non-sense mutation generating stop codon and deletion mutations [59]. Among the point mutations, L84S is the predominant one and associated with mild disease symptoms among the hospitalised individuals [59, 74]. Three deletion mutations of ORF8 are reported world-wide of which 382 nucleotide deletions resulted in complete loss of ORF8 and the terminal part of ORF7b. This variant was originated in Wuhan and traced to Taiwan and Singapore [59]. Notably, deletion of this locus was associated with milder infection due to reduced systemic release of cytokines and a better immune response to SARS-CoV-2 [78]. In addition to deletions, several non-synonymous amino acid substitutions in ORF8 are reported world-wide indicating positive natural selection of those variants [50].

A 27 amino acid in-frame deletion is reported for ORF7a locus [46]. Structural analysis revealed loss of putative signal peptide and first two beta strands from ORF7a, the orthologue of SARS-CoV ORF7a. However, the implication of such mutation on viral fitness needs further investigation.

## Conclusion

The unusually larger genome of CoVs among RNA viruses is primarily responsible for their daunting genome plasticity due to frequent mutation and recombination [1]. In addition to this, presence of error prone replication machinery in RNA viruses largely contributes to their genetic diversity with varying outcomes including shift in their biological properties, interspecies transmission and altered transmissibility [11, 79]. The overall outcome of mutations is reflected at the species level either by making it stronger or weaker. Any mutation which provides survival advantage is positively selected by nature and thus mutational studies are essential to understand the evolutionary trend at the organismal level [80]. Frequency distribution of mutations in different proteins of SARS-CoV-2 variants from countries with total

infection >2 lakhs showed almost all the protein coding ORFs harboured mutations to a varying extent (Fig. 2). Furthermore, mutations in ORF1a, ORF1b, N and S proteins were present in almost all the countries of which Canada, South Africa and Spain showed comparatively higher number of mutations in N protein. However, Morocco had highest number of S protein mutations (Fig. 2).

Among the structural proteins, M and E had least number of variants indicating these are conserved proteins [10] (Fig. 2). The emergence of numerous genetic variants has brought SARS-CoV-2 into spotlight due to its enhanced transmissibility and infectivity compared to the original Wuhan strain [13]. Moreover, mutations in structural (spike) and accessory proteins (ORF3a) of SARS-CoV-2 are associated with a higher CFR of COVID-19 pandemic [13, 56, 65]. The nucleocapsid phosphoprotein and spike glycoprotein are among the most non-conserved proteins in SARS-CoV-2 posing a major challenge towards vaccine development [10]. Moreover, S protein variants are highly infectious due to effective binding with the host cell receptor. On the contrary, the other structural proteins including membrane and envelope were relatively more conserved suggesting perturbation within these genes are not encouraged which otherwise might affect viral integrity and life cycle [10]. Among the SARS-CoV-2 non-structural protein variants, deletion at position Asp268 of NSP2 spread rapidly in Europe [81]. Deletion of three amino acids, KSF towards the 3' end of NSP1 at positions 241–243 was found in viral isolates from different geographical locations, suggesting their rapid spread [82]. Whether such mutations have any effect on viral pathogenicity needs to be explored.

There had been considerable advancements in the field of vaccines, therapeutic antibodies and anti-viral therapy to combat COVID-19 [51, 61]. However, the emergent genetic variants might undermine the effectiveness of those therapeutic interventions. With the outbreak of COVID-19 pandemic, there has been an explosive deposition of SARS-CoV-2 genome sequences in the repositories which made detailed analysis of SARS-CoV-2 genetic variants much easier. As COVID-19 pandemic progresses, closer investigation of those evolving strains of SARS-CoV-2 is crucial to understand the biological significance of the mutations on viral fitness.

## Author contributions.

SN and PM conceptualised the idea, retrieved and analysed the data. PM wrote the manuscript, and SN emended and approved the final version.

**Conflict of interest.** The authors declare no potential conflicts.

**Data availability statement.** The data presented in this review paper would be available from the corresponding author upon request. The mutation data on SARS-CoV-2 variants are freely accessible from NextStrain open source project (<https://nextstrain.org/ncov>).

## References

1. Lu R *et al.* (2020) Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* **395**, 565–574.
2. Zhou P *et al.* (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273.
3. Zhu N *et al.* (2020) A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine* **382**, 727–733.
4. Kahn JS and McIntosh K (2005) History and recent advances in coronavirus discovery. *The Pediatric Infectious Disease Journal* **24**, S223–S227.

5. Guan WJ *et al.* (2020) Clinical characteristics of coronavirus disease 2019 in China. *The New England Journal of Medicine* **382**, 1708–1720.
6. Peng X *et al.* (2020) Transmission routes of 2019-nCoV and controls in dental practice. *International Journal of Oral Science* **12**, 1–6.
7. Pachetti M *et al.* (2020) Impact of lockdown on COVID-19 case fatality rate and viral mutations spread in 7 countries in Europe and North America. *Journal of Translational Medicine* **18**, 1–7.
8. Mercatelli D and Giorgi FM (2020) Geographic and genomic distribution of SARS-CoV-2 mutations. *Frontiers in Microbiology* **11**, 1800.
9. Toyoshima Y *et al.* (2020) SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *Journal of Human Genetics* **65**, 1075–1082.
10. Wang R *et al.* (2020) Decoding SARS-CoV-2 transmission, evolution and ramification on COVID-19 diagnosis, vaccine, and medicine. *The Journal of Physical Chemistry Letters* **11**, 10007–10015.
11. Su S *et al.* (2016) Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends in Microbiology* **24**, 490–502.
12. Salemi M *et al.* (2004) Severe acute respiratory syndrome coronavirus sequence characteristics and evolutionary rate estimate from maximum likelihood analysis. *Journal of Virology* **78**, 1602–1603.
13. Becerra Flores M and Cardozo T (2020) SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. *International Journal of Clinical Practice* **74**, e13525.
14. Wu F *et al.* (2020) A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269.
15. Jin Z *et al.* (2020) Structure of M pro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **582**, 289–293.
16. Perlman S and Netland J (2009) Coronaviruses post-SARS: update on replication and pathogenesis. *Nature Reviews Microbiology* **7**, 439–450.
17. Michel CJ *et al.* (2020) Characterization of accessory genes in coronavirus genomes. *Virology Journal* **17**, 1–13.
18. Jauregui AR *et al.* (2013) Identification of residues of SARS-CoV nsp1 that differentially affect inhibition of gene expression and antiviral signaling. *PLoS One* **8**, e62416.
19. Schubert K *et al.* (2020) SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. *Nature Structural and Molecular Biology* **27**, 959–966.
20. Angeletti S *et al.* (2020) COVID-2019: the role of the nsp2 and nsp3 in its pathogenesis. *Journal of Medical Virology* **92**, 584–588.
21. Graham RL *et al.* (2006) The nsp2 proteins of mouse hepatitis virus and SARS coronavirus are dispensable for viral replication. *The Nidoviruses* **581**, 67–72.
22. Frieman M *et al.* (2009) Severe acute respiratory syndrome coronavirus papain-like protease ubiquitin-like domain and catalytic domain regulate antagonism of IRF3 and NF- $\kappa$ B signaling. *Journal of Virology* **83**, 6689–6705.
23. Sakai Y *et al.* (2017) Two-amino acids change in the nsp4 of SARS coronavirus abolishes viral replication. *Virology* **510**, 165–174.
24. V'kovski P *et al.* (2020) Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews Microbiology* **19**, 155–170.
25. Benvenuto D *et al.* (2020) Evolutionary analysis of SARS-CoV-2: how mutation of non-structural protein 6 (NSP6) could affect viral autophagy. *Journal of Infection* **81**, e24–e27.
26. Gao Y *et al.* (2020) Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science (New York, N.Y.)* **368**, 779–782.
27. Littler DR *et al.* (2020) Crystal structure of the SARS-CoV-2 non-structural protein 9, Nsp9. *iScience* **23**, 101258.
28. Decroly E *et al.* (2011) Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex. *PLoS Pathogen* **7**, e1002059.
29. Lin S *et al.* (2020) Crystal structure of SARS-CoV-2 nsp10/nsp16 2'-O-methylase and its implication on antiviral drug design. *Signal Transduction and Targeted Therapy* **5**, 1–4.
30. Ivanov KA and Ziebuhr J (2004) Human coronavirus 229E nonstructural protein 13: characterization of duplex-unwinding, nucleoside triphosphatase, and RNA 5'-triphosphatase activities. *Journal of Virology* **78**, 7833–7838.
31. Yuen CK *et al.* (2020) SARS-CoV-2 nsp13, nsp14, nsp15 and orf6 function as potent interferon antagonists. *Emerging Microbes and Infections* **9**, 1–29.
32. Yan L *et al.* (2020) Architecture of a SARS-CoV-2 mini replication and transcription complex. *Nature Communications* **11**, 1–6.
33. Chen Y *et al.* (2009) Functional screen reveals SARS coronavirus non-structural protein nsp14 as a novel cap N7 methyltransferase. *Proceedings of the National Academy of Sciences* **106**, 3484–3489.
34. Ogando NS *et al.* (2020) The enzymatic activity of the nsp14 exoribonuclease is critical for replication of MERS-CoV and SARS-CoV-2. *Journal of Virology* **94**, e01246-20.
35. Kim Y *et al.* (2020) Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV-2. *Protein Science* **29**, 1596–1605.
36. Viswanathan T *et al.* (2020) Structural basis of RNA cap modification by SARS-CoV-2. *Nature Communications* **11**, 1–7.
37. Walls AC *et al.* (2020) Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **181**, 281–292.
38. Kamau A *et al.* (2020) Functional pangenome analysis provides insights into the origin, function and pathways to therapy of SARS-CoV-2 coronavirus. *bioRxiv*.
39. Wang PH *et al.* (2020) Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) membrane (M) protein inhibits type I and III interferon production by targeting RIG-I/MDA-5 signaling. *bioRxiv*.
40. Le Bert N *et al.* (2020) SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature* **584**, 457–462.
41. Ren Y *et al.* (2020) The ORF3a protein of SARS-CoV-2 induces apoptosis in cells. *Cellular and Molecular Immunology* **17**, 881–883.
42. Konno Y *et al.* (2020) SARS-CoV-2 ORF3b is a potent interferon antagonist whose activity is increased by a naturally occurring elongation variant. *Cell Reports* **32**, 108185.
43. Miorin L *et al.* (2020) SARS-CoV-2 Orf6 hijacks Nup98 to block STAT nuclear import and antagonize interferon signaling. *Proceedings of the National Academy of Sciences* **117**, 28344–28354.
44. Park MD (2020) Immune evasion via SARS-CoV-2 ORF8 protein? *Nature Reviews Immunology* **20**, 408.
45. Gordon DE *et al.* (2020) A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468.
46. Holland LA *et al.* (2020) An 81 nucleotide deletion in SARS-CoV-2 ORF7a identified from sentinel surveillance in Arizona (Jan–Mar 2020). *Journal of Virology* **94**, e00711-20.
47. Davies NG *et al.* (2020) Estimated transmissibility and severity of novel SARS-CoV-2 variant of concern 202012/01 in England. *medRxiv*.
48. Mahase E (2020) COVID-19: what have we learnt about the new variant in the UK? *The British Medical Journal* **371**:m4944, 1–2.
49. Tushir S *et al.* (2021) Proteo-genomic analysis of SARS-CoV-2: a clinical landscape of single-nucleotide polymorphisms, COVID-19 proteome, and host responses. *Journal of Proteome Research* **20**, 1591–1601.
50. Farkas C *et al.* (2020) Large-scale population analysis of SARS-CoV-2 whole genome sequences reveals host-mediated viral evolution with emergence of mutations in the viral Spike protein associated with elevated mortality rates. *medRxiv*.
51. Li Q *et al.* (2020) The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell* **182**, 1284–1294.
52. Kemp S *et al.* (2020) Recurrent emergence and transmission of a SARS-CoV-2 spike deletion  $\Delta$ H69/V70. *bioRxiv*.
53. Islam MR *et al.* (2020) Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity. *Scientific Reports* **10**, 1–9.
54. Phan T (2020) Genetic diversity and evolution of SARS-CoV-2. *Infection, Genetics and Evolution* **81**, 104260.
55. Rahman MS *et al.* (2020) Evolutionary dynamics of SARS-CoV-2 nucleocapsid protein and its consequences. *Journal of Medical Virology* **93**, 2177–2195.
56. Majumdar P and Niyogi S (2020) ORF3a Mutation associated with higher mortality rate in SARS-CoV-2 infection. *Epidemiology and Infection* **148**, e262, 1–6.
57. Pachetti M *et al.* (2020) Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *Journal of Translational Medicine* **18**, 1–9.
58. Issa E *et al.* (2020) SARS-CoV-2 and ORF3a: nonsynonymous mutations, functional domains, and viral pathogenesis. *Msystems* **5**, 1–7.
59. Pereira F (2020) Evolutionary dynamics of the SARS-CoV-2 ORF8 accessory gene. *Infection, Genetics and Evolution* **85**, 104525.

60. **Robson F et al.** (2020) Coronavirus RNA proofreading: molecular basis and therapeutic targeting. *Molecular Cell* **79**, 710–727.
61. **Martinot M et al.** (2020) Remdesivir failure with SARS-CoV-2 RNA-dependent RNA-polymerase mutation in a B-cell immuno deficient patient with protracted COVID-19. *Clinical Infectious Diseases* **1474**, 1–14.
62. **Hoffmann M et al.** (2020) SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280.
63. **Coutard B et al.** (2020) The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Research* **176**, 104742.
64. **Khan MI et al.** (2020) Comparative genome analysis of novel coronavirus (SARS-CoV-2) from different geographical locations and the effect of mutations on major target proteins: an *in silico* insight. *PLoS One* **15**, e0238344.
65. **Callaway E** (2020) Making sense of coronavirus mutations. *Nature* **585**, 174–177.
66. **Ilmjärvi S et al.** (2020) Epidemiologically most successful SARS-CoV-2 variant: concurrent mutations in RNA-dependent RNA polymerase and spike protein. medRxiv.
67. **Gussow AB et al.** (2020) Genomic determinants of pathogenicity in SARS-CoV-2 and other human coronaviruses. *Proceedings of the National Academy of Sciences* **117**, 15193–15199.
68. **McBride R et al.** (2014) The coronavirus nucleocapsid is a multifunctional protein. *Viruses* **6**, 2991–3018.
69. **Dinesh DC et al.** (2020) Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *PLoS Pathogens* **16**, e1009100.
70. **Kang S et al.** (2020) Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharmaceutica Sinica B* **10**, 1228–1238.
71. **Joshi A and Paul S** (2020) Phylogenetic analysis of the novel coronavirus reveals important variants in Indian strains. BioRxiv.
72. **Forni D et al.** (2016) Extensive positive selection drives the evolution of nonstructural proteins in lineage C betacoronaviruses. *Journal of Virology* **90**, 3627–3639.
73. **Hurst KR et al.** (2013) Characterization of a critical interaction between the coronavirus nucleocapsid protein and nonstructural protein 3 of the viral replicase-transcriptase complex. *Journal of Virology* **87**, 9159–9172.
74. **Nagy Á et al.** (2020) Different mutations in SARS-CoV-2 associate with severe and mild outcome. *International Journal of Antimicrobial Agents* **57**, 106272.
75. **Shah A** (2020) Novel coronavirus-induced NLRP3 inflammasome activation: a potential drug target in the treatment of COVID-19. *Frontiers in Immunology* **11**, 1–5.
76. **Mehta P et al.** (2020) COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet (London, England)* **395**, 1033.
77. **Zhang Y et al.** (2020) The ORF8 protein of SARS-CoV-2 mediates immune evasion through potently downregulating MHC-I. bioRxiv.
78. **Young BE et al.** (2020) Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *The Lancet* **396**, 603–611.
79. **Sanjuán R et al.** (2010) Viral mutation rates. *Journal of Virology* **84**, 9733–9748.
80. **Domingo EJJH and Holland JJ** (1997) RNA Virus mutations and fitness for survival. *Annual Review of Microbiology* **51**, 151–178.
81. **Bal A et al.** (2020) Molecular characterization of SARS-CoV-2 in the first COVID-19 cluster in France reveals an amino acid deletion in nsp2 (Asp268del). *Clinical Microbiology and Infection* **26**, 960–962.
82. **Benedetti F et al.** (2020) Emerging of a SARS-CoV-2 viral strain with a deletion in nsp1. *Journal of Translational Medicine* **18**, 1–6.