The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has become the major sarbecovirus of the 21st century that has been the most widespread and deadliest, causing worldwide epidemic and pandemic over a year and still ongoing. The evolutionary history of SARS-CoV-2 has shown it to be the fastest spreading infection over the globe [2,3] compared to previously reported coronavirus such as severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), both in terms of the number of people infected and the spread of epidemic areas. The primary reason behind the widespread ability of SARS-CoV-2 across the globe are, its fast pace to adapt to the different global geo-climatic environment and host-immune system. The evolutionary history of SARS-CoV-2 compared to other rapidly evolving RNA viruses has been highly recombinogenic. Additionally, different part of its genome follows a different evolutionary path, violating the assumption of the standard phylogenetic lines [4]. Therefore, SARS-CoV-2 causing coronavirus disease 19 (COVID-19) has become an extraordinary threat to global health [5,6].

Recent studies have shown mutations in the regions of the structural protein as well as in nonstructural protein (NSP), and in the open reading frame (ORF) regions of the SARS-CoV-2 RNA genome has significantly promoted virus infectivity and pathogenicity [7-9]. Morphologically, it is composed of four structural protein types, i.e., the spike (S) glycoprotein (S-protein), the envelope (E) protein, the membrane (M) protein, and the nucleocapsid (N) protein (Figure 1). Whereas, the nonstructural proteins, such as NSP 12 (RNA-dependent RNA polymerase, RdRp), NSP13 (helicase), papain-like protease, and 3-chymotrypsin-like protease, are encoded by the ORF region of the RNA.
genome \[10\]. Evolution of SARS-CoV-2 evidently showing two directional mutational trends, i.e., mutations towards stability and mutations towards variability. Based on above evolutionary aspect, the sections below focus on the most prominent reported viral variants along with their predominated mutational regions.

Along the evolutionary trend of mutations towards stability direction, the most prominent mutations of SARS-CoV-2 took an evolutionary path promoting stability to the virus, i.e., favoring its enhanced fitness and/or infectivity.

Among non-synonymous mutations, the S-protein has succumbed major mutational variations. The S-protein that facilitates viral particle attachment and entry into the host cell (Figure 1) has demonstrated the highest variations directed towards its stability, to adapt and improve its compatibility with the host. Of particular interest, mutagenesis of S-protein was mainly observed at the receptor-binding domain (RBD). One of the most dominant and widespread SARS-CoV-2 variants that has emerged during the Covid-19 pandemics is the substitution of aspartic acid (D) to glycine (G) at 614 position (D614G) in the RBD region \[47\]. The D614G mutation provides a stable linkage between the S1 and S2 subunits at the RBD site, prompting a more “open” easy to be cleaved conformation than the ancestral D614 form (Figure 1), that improved the ability to bind to the host cells \[11\]. This single residue substitution mutation has improved the local interaction energy due to significant change in free energy (i.e., -2.6 kcal/mol of closed conformation changed to -2.0 kcal/mol for open conformation) and thus reducing free energy, has provided more thermodynamic stability to the S1 within the S1/S2 complex \[48,49\]. Overall, the D614G mutation favored S-protein structural stability and enhanced “open” confirmation has enhanced improving viral infectivity.

The global initiative on sharing all influenza data (GIASID) review shows that the D614G mutation initially emerged in Europe and spread worldwide \[11,12\]. Even though the D614G SARS-CoV-2 variant has shown to be 1.22 times faster at spreading than the conventional wild type, statistically, its significance was found to be low, presumably due to the influence of other parallel co-mutation in other vital proteins.

Another non-synonymous mutational hot-spot was observed in RdRp, the central subunit for RNA-synthesize needed for transcription and replication of essential viral proteins (Figure 1). As mentioned above, a parallel co-evolution was observed in the RdRp (NSP12), where a single substitution mutation in RdRp, i.e., the substitution of proline (P) to leucine (L) at 323 (P323L). The P323L RdRp mutation with the D614G S-protein mutation (P323L/D614G) significantly predominates and presently is the most widespread SARS-CoV-2 variant \[13\]. An artificial intelligent modeling study indicates that the altered conformational structural change in S-protein due to D614G mutation has become pervasive only when complemented with P323L RdRp mutation \[14\]. Therefore, the P323L RdRp mutation complementing the D614G mutation could be the primary reason behind their co-evolution and simultaneous worldwide spread.

The exact role of the P323L mutation of RdRp has not been evaluated yet. However, it is assumed that this mutation may have the potential to interfere with potential drug binding as this substitution is close to the major drug docking site on RdRp \[13\]. Further, structural study reveals P323L mutation gave rigidity to RdRp structure and thus, the positive free energy change (ΔΔG: 0.908 kcal/mol) \[18\], suggesting P323L mutation may give more stability to the RdRp structure. The improved intramolecular interactions and stability of the RdRp due to P323L may have some functional consequences, yet to be determined. However, a study published supports that the P323L mutation may have promoted the viral replication speed \[14\]. However, the 3D structural position of the P323L mutation, being far away from the RdRp catalytic site, doesn’t potently support this hypothesis \[17\].

According to GISAID, the D614G/P323L mutant variant is the most dominant SARS-CoV-2 variant worldwide. The genealogy and success model of SARS-CoV-2 haplotypes indicated the haplotype that harbored the initial D614G mutation alone was unsuccessful, however, the addition of the P323L mutation in nsp12 may have introduced a compensatory adaptive change in the haplotype that has successfully enhanced viral fitness and its spreading potency.

One of the most predominant synonymous mutation was observed in NSP3. NSP3, a large protein with a papain-like protease domain and essential for the viral polyprotein cleavage. Although infrequently several sites in NSP3 protein have undergone more than one mutational change. Among various mutations, one predominant mutation is the nucleotide mutation leading to phenylalanine (F) to F synonymous mutation at 106 position (F106F). The SARS-CoV-2 haplotypes comparison model study and wavelet plot \[14\] indicated the F106F synonymous mutation in NSP3 is also linked to D614G and have co-existed, indicating its possible co-evolved along with D614G mutation.

Although synonymous mutations are neutral, however, the codon usage of F106F mutation SARS-CoV-2 variant compared to pangolin lineage (predicted to be evolved from the pangolin host) has shown a significant increase in this codon frequency (TTT from TTC, \[p < 0.01\] \[14,18\]. This suggests that the change in the codon has helped the
The viroporin proteins another set of proteins supporting towards variability evolutionary trend. Viroporin proteins are small hydrophobic ion-channel proteins that promote virus release, replication, and virulence by altering cellular membrane potential (Figure 1). Among 3 Coronavirus associate viroporins (i.e., E-protein and accessory proteins 3a and 8), E-protein and viroporin 3a protein is associated with viral viability and virulence [23]. The viroporin 3a protein, the largest among all, is associated with cell lysis and virus release by dissipating membrane potential, hence, playing an essential role in the viral spread. The major mutations in viroporin 3a protein were found at their N-terminal transmembrane domain (TD) and a C-terminal cytosolic domain (CM) region [21]. Based on mapping the intrinsic disorder and gain-loss of binding energy, the viroporin 3a protein has been categorized as “moderately disordered”. This is because at the mutational entropy level the high-entropy substitution with glutamine (Q) replacing histidine (H) at position 57 (Q57H) at the pore of the TD region has been compensated by the presence of low-entropy mutation of G by valine (V) at 197 position (G197V) of the CM domain (lesser intrinsic disorder) leading to the entropic return mode. Whereas the low entropy mutation at position 13 (V13L) and high-entropy mutation 251 (G251V), located at terminal amino acids region, follows entropy expansion and return mode, respectively. Therefore, mutation at N-terminal and pore (i.e., V13L and Q57H) doesn’t alter disorder or binding affinity of N-protein, but mutations in the CD loop region (G197V) and C-terminal (G251V) which significantly do decrease disorder and increased binding potential. Further, with time these mutations are growing in magnitude (G251V mutation expanded 13.8% in all proteomes). Hence, the increasing mutation magnitude and entropy indicates in the future that these mutations could significantly contribute to the viral evolution.

Another evolutionary trend very common in infectious RNA viruses like influenza and SARS-CoV-2, is sporadic clustered mutations promoting fast viral evolution. Viruses are known to mutate continually over time. It is expected that the accumulation of these mutations may lead to new viral variants. In general, the molecular clock estimates to introduce two substitutions per genome per month [24]. SARS-CoV-2 is also mutating but relatively slowly [50, 51].
However, recent reported sporadic fast multivariant mutations accumulation in particular geographical areas has led to the generation of new phylogenetic clusters establishing new dominant SARS-CoV-2 variants. The onset of viral lineages due to fast multiple clustered mutations accumulation was originated due to prolonged viral association in an immunocompromised host [25]. The section below discuss various SARS-CoV-2 variants with clustered mutations that are presently circulating globally and are the major cause of Covid-19 second pandemic wave [24].

The phylogenetic cluster, B.1.1.7 lineage classified under “variants of concern” [52] (also known as Variant Under Investigation (VUI) – 202012/01), rapidly emerged in the United Kingdom (UK), in early December 2020, with an estimated 70% increase in transmissibility [27, 28]. The B.1.1.7 variant emerged due to a sudden accumulation of an unusually large number of genetic mutations at an unprecedented rate. It has 23 mutational differences from the original Wuhan strain, including 13 non-synonymous mutations, 4 deletions, and 6 synonymous mutations [29]. The multiple mutations in the S-protein (deletion 69-70, deletion 144, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H) along with mutations in other regions, reportedly could be the reason behind its increased transmission potential.

Among multiple mutations in the S-protein, 3 mutations are assumed to have a potential biological effect, i.e., asparagine (N) to tyrosine (Y) substitution at position 501 in RBD (N501Y), deletion 69-70 (69-70del) in S1 subunit, and P to H substitution at position 68 (P681H) in S2 subunit. A study suggested N501Y mutation increases angiotensin-converting enzyme 2 (ACE2) receptor affinity of RBD [30, 31], and the P681H mutation within the furin cleavage site between S1 and S2 region [32] which improved viral entry and transmission in animal models [33, 34]. The presence of the P681H mutation at the furin cleavage (absent in other related coronaviruses) may also have contributed to viral ability to escape the innate immunity by allowing viral infection independent of endosomal entry [33]. The 69-70del mutation has a deletion of 2 amino acids in RBD of S1 subunit, but the significance of this mutation is still not well understood. However, experts predict 69-70del and N501Y mutation together induces significant RBD changes resulting in enhanced transmissibility.

B.1.1.7 lineage is a dominant, rapidly growing lineage in the UK and Ireland [34] which has spread to over 50 countries, including the USA and Canada (until 30th Jan. 2021). These mutations could evade the immune system or could increase illness severity is less likely [53]. However, a recent study illustrated that recurrent deletion regions (RDRs) like 69-70del, near antibody epitope sites, promoting SARS-CoV-2 antigenic evolution, have potential to confer resistance to neutralizing antibodies [54]. In contrast another study reflects novel potential therapeutics are under development with low chance of virus resistance [53]. Therefore, more studies are under investigation to evaluate if these mutations could reduce the efficacy of vaccine-induced immunity.

Another, is the SARS-CoV-2 B.1.351 lineage, classified under “variants of concern” [52] also known as 501Y.V2. Reportedly, emerged in South Africa, almost parallelly to the B.1.1.7 UK variant in December 2020 and has been reported to be detected in 70 countries worldwide (until 30th Jan. 2021) [38]. Similar to the B.1.1.7 variant, a sporadic accumulation of an unusually large number of mutations increased B.1.351 variant transmissibility. Even though B.1.351 and B.1.1.7 variants share the N501Y mutation, both variants are phylogenetically different and have emerged independently. Additionally, unlike B.1.1.7, it lacks a 69-70del mutation.

In addition to N501Y mutation in the RBD region, B.1.351 also reported having two more mutations at the RBD site: K to substitution N at 417 position (K417N) and glutamic acid (E) to K substitution at 484 position (E484K). The presence of secondary mutations and the N501Y mutation (K417N, E484K, N501Y) may have arisen to compensate the fitness penalty to support changes in S protein structure [39]. From the study, it is evident that the 501Y.V2 variant is acting under “diversifying” positive selection pressure.

Altogether, the 501Y.V2 variant with better fitness and improved ACE2 binding affinity [31], has resulted in higher transmissibility than the UK B.1.1.7 variant. The fact that, the 501Y.V2 variant infected population predominance in the South African population within few weeks somewhat supports this hypothesis [40]. There are some real concerns associated with this variant because the conformation changes in the S protein due to these mutations (K417N, E484K, N501Y) may impact the effectiveness of the COVID-19 vaccines. Studies have already indicated the E484K mutation has reduced sera neutralization efficacy by >10-fold [41, 42]. But presently, both variants are under further thorough investigation.

The good news comes from a recent study published by Moderna Inc. which has shown that Moderna COVID-19 Vaccine has been found to be active against both UK B.1.1.7 variant and South African 501Y.V2 strain [43]. Another piece of positive news is that the South African 501Y.V2 variant has shown no indication of increased illness severity.

The P.1 lineage, another “variants of concern” SARS-CoV-2 variant, originated in Brazil’s Amazon region in December 2020, with 17 non-synonymous mutations, 3 deletions, 4
synonymous mutations, and one 4nt insertion [44]. As of now (28th Jan. 2021), only 8 countries have reported having cases with the P.1 variant [48]. Most mutations of which there are 10 are in the S-protein, including K417T, N501Y, and E484K. It is a descendant of the B.1.1.28 lineage, however, mutations at the RBD site resemble with the B.1.1.7 variant by having N501Y mutation and with the B.1.351 variant due to presences of K417N/T and E484K mutation. As with the UK B.1.1.7 and the South African B.1.351 variant, Brazil P.1 variant also harbor the orf1b deletion del11288-11296.

The P.1 variant is more worrisome because, in addition to higher transmissibility and immune invading ability [42], this variant has shown to have reinfection ability (re-infect who had already been infected with COVID-19) [44]. However, there is no evidence of its increased illness severity.

The evolutionary path of SARS-CoV-2 indicates, the virus initiated its evolutionary journey by increasing the stability of major functional proteins, and now it is shifting towards variability to improve transmissibility, immune-invading ability, and infectivity.

REFERENCES

6. Q. Han, et al., J Infect. Path, 80, 373 (2020).
28. N.V.I.M. team., Public health england, imperial college, the university of edinburgh , the welcome sanger institute 2020.
38. A. O’Toole, et al., Virolological.org2021, pp. https://virolological.org/t/ tracking-the-
international-spread-of-sars-cov-2-lineages-b-
1-1-7-and-b-1-351-501-y-v502/592.
selection-analysis-of-the-south-african-v501-
42. Y. Weisblum, et al., Elle, 9 (2020).
52. CDC, in: CDC (Ed.), NCIR2021