



SARS-CoV-2 Mutations and COVID-19 Clinical Outcome: Mutation Global Frequency Dynamics and Structural Modulation Hold the Key

Ranjeet Maurya^{1,2†}, Pallavi Mishra^{1†}, Aparna Swaminathan^{1†}, Varsha Ravi¹, Sheeba Saifi¹, Akshay Kanakan¹, Priyanka Mehta¹, Priti Devi^{1,2}, Shaista Praveen¹, Sandeep Budhiraja³, Bansidhar Tarai³, Shimpa Sharma⁴, Rajesh J. Khyalappa⁴, Meghnad G. Joshi⁴ and Rajesh Pandey^{1,2*}

OPEN ACCESS

Edited by:

Shashank Gupta,
Norwegian University of Life Sciences,
Norway

Reviewed by:

Gaurav Verma,
Region Skåne, Sweden
Rashmi Rana,
Sir Ganga Ram Hospital, India

*Correspondence:

Rajesh Pandey
rajeshp@igib.in

[†]These authors share first authorship

Specialty section:

This article was submitted to
Microbiome in Health and Disease,
a section of the journal
Frontiers in Cellular and
Infection Microbiology

Received: 02 February 2022

Accepted: 16 February 2022

Published: 21 March 2022

Citation:

Maurya R, Mishra P, Swaminathan A,
Ravi V, Saifi S, Kanakan A, Mehta P,
Devi P, Praveen S, Budhiraja S,
Tarai B, Sharma S, Khyalappa RJ,
Joshi MG and Pandey R (2022)
SARS-CoV-2 Mutations and COVID-
19 Clinical Outcome: Mutation Global
Frequency Dynamics and Structural
Modulation Hold the Key.
Front. Cell. Infect. Microbiol. 12:868414.
doi: 10.3389/fcimb.2022.868414

¹ INtegrative GENomics of HOst-PathogEn (INGEN-HOPE) Laboratory, Council of Scientific & Industrial Research (CSIR)-Institute of Genomics and Integrative Biology (CSIR-IGIB), Delhi, India, ² Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India, ³ Max Super Speciality Hospital (A Unit of Devki Devi Foundation), Max Healthcare, Delhi, India, ⁴ Dr. D. Y. Patil Medical College, Pune, India

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has had an enormous burden on the healthcare system worldwide as a consequence of its new emerging variants of concern (VOCs) since late 2019. Elucidating viral genome characteristics and its influence on disease severity and clinical outcome has been one of the crucial aspects toward pandemic management. Genomic surveillance holds the key to identify the spectrum of mutations *vis-à-vis* disease outcome. Here, in our study, we performed a comprehensive analysis of the mutation distribution among the coronavirus disease 2019 (COVID-19) recovered and mortality patients. In addition to the clinical data analysis, the significant mutations within the two groups were analyzed for their global presence in an effort to understand the temporal dynamics of the mutations globally in comparison with our cohort. Interestingly, we found that all the mutations within the recovered patients showed significantly low global presence, indicating the possibility of regional pool of mutations and the absence of preferential selection by the virus during the course of the pandemic. In addition, we found the mutation S194L to have the most significant occurrence in the mortality group, suggesting its role toward a severe disease progression. Also, we discovered three mutations within the mortality patients with a high cohort and global distribution, which later became a part of variants of interest (VOIs)/VOCs, suggesting its significant role in enhancing viral characteristics. To understand the possible mechanism, we performed molecular dynamics (MD) simulations of nucleocapsid mutations, S194L and S194*, from the mortality and recovered patients, respectively, to examine its impacts on protein structure and stability. Importantly, we observed the mutation S194* within the recovered to be comparatively unstable, hence showing a low global frequency, as we observed. Thus, our study provides integrative

insights about the clinical features, mutations significantly associated with the two different clinical outcomes, its global presence, and its possible effects at the structural level to understand the role of mutations in driving the COVID-19 pandemic.

Keywords: COVID-19, SARS-CoV-2, mutation analysis, disease outcome, global frequency flip, molecular dynamics simulation

INTRODUCTION

The ongoing pandemic of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that emerged in late 2019 has had devastating impacts worldwide, and global collective efforts are being undertaken to characterize the virus through genomic surveillance and *in vitro* experiments (Mercatelli and Giorgi, 2020; Zhang et al., 2020). In comparison to other RNA viruses, SARS-CoV-2 has a lower evolutionary rate, although genetic mutations are one of the major mechanisms by which the virus evolves rapidly as a result of genetic selection. The evolving mutations, combined with large population size and short generation time, allow the virus to adapt to the host environment easily (Sanjuán and Domingo-Calap, 2016). Though most of the mutations induce null changes, a proportion of mutations are observed to relatively occur in higher incidences and cause challenges toward the development of vaccines, drugs, diagnostic tools, and even the disease clinical presentations (Lai et al., 2020).

Of all the mutations in different regions of the SARS-CoV-2 genome, the spike region is of major focus, as it facilitates the virus entry to the host *via* interaction with Angiotensin Converting Enzyme 2 (ACE2) receptors (Yang et al., 2020b). Several variants of interest (VOIs) and variants of concern (VOCs) with mutations in spike regions have been characterized and monitored to track the subsequent evolution of SARS-CoV-2 (<https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>). So far, several VOCs [B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma), B.1.617.1 (Kappa), B.1.617.2 (Delta), B.1.525 (Eta), B.1.526 (Iota), C.37 (Lambda), B.1.621 (Mu), and presently B.1.1.529 and BA.1 and BA.2 (Omicron)] have been reported that modulate the viral transmission, disease severity, binding ability, and effectiveness of vaccines. One such VOC, Omicron, has more than 46 mutations in the spike protein with higher transmissibility and infectivity, leading to a global infection surge (Kannan et al., 2022). Due to the presence of an unusual number of mutations in the S protein, it might result in more efficient binding to ACE2 receptors and escape from neutralizing antibodies (Hu et al., 2022). However, a recent study demonstrated less efficient replication of Omicron due to its reduced dependence on Transmembrane Serine Protease 2 (TMPRSS2) for viral replication in comparison to Delta (Zhao et al., 2021). At the same time, data on disease severity are also steadily being reported from different parts of the world that would augment further understanding in this regard (<https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>).

The heterogeneity of clinical outcomes exhibited by SARS-CoV-2-affected individuals depends on numerous factors,

inclusive of interindividual differences is the genotype-to-phenotype association for SARS-CoV-2 pathogenesis (Mousavizadeh and Ghasemi, 2021; Pereira et al., 2021). Several studies have reported high frequency mutation hotspots in SARS-CoV-2 genome regions that include ORF1a, S, and N (Kaushal et al., 2020). Majumdar and Niyogi (2020) reported P25L mutation in the Orf3a region to be strongly associated with higher mortality rates as a result of structural modification including the acquisition of escalated antigen diversity and B-cell epitope loss. Mutations in the NSP12 region that are involved in RNA-dependent RNA polymerase (RdRp) expression have been shown to contribute toward 5.9 times increased disease severity (Voss et al., 2020). Additionally, the mutation S194L in N protein was reported to be associated with the symptomatic patients and is located in the flexible linker region of the protein N (Zeng et al., 2020; Barona-Gómez et al., 2021).

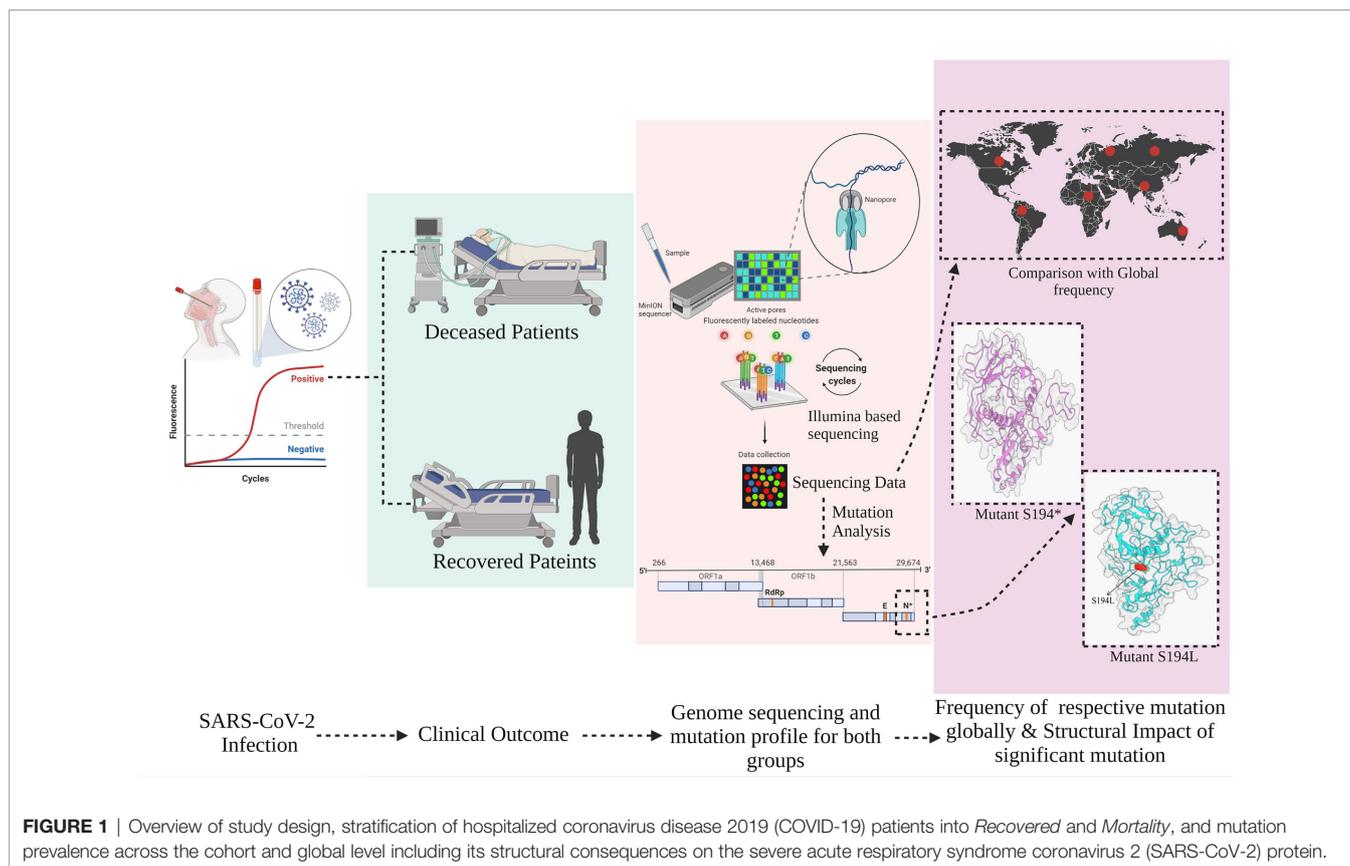
Several mutations manifest as geographic patterns alluding to the virus ability to modify/adapt itself within a distinct microenvironment (Bakhshandeh et al., 2021). To acquire comprehension of these mutation patterns, it is important to have insights into their respective mutation frequency across both local and global patterns and help preempt the emergence of VOIs and VOCs (Justo Arevalo et al., 2021; Oude Munnink et al., 2021). Mutations along with host-modulating factors have been reported as a major contributor toward disease severity in COVID-19 (Mousavizadeh and Ghasemi, 2021; Pereira et al., 2021). Substitution of an amino acid residing in the antigen-determining region can possibly repress antibody-mediated immunity, resulting in enhanced virus proliferation (Gupta et al., 2020). Hence, it is crucial to investigate the correlation between mutations and patient clinical outcome that may play a role in SARS-CoV-2 pathophysiology and progression of disease severity (Toyoshima et al., 2020).

The present study has sought to evaluate the mutational spectrum, clinical diversity, patient's severity outcomes, and their respective global frequency in hospitalized COVID-19 patients. We further analyzed the structural modification of these mutations in SARS-CoV-2 that may alter its physiochemical properties leading to amelioration of pathogenicity (**Figure 1**).

METHODS

Clinical Data and Classification

Clinical data of 246 COVID-19 patients were collected from Max Super Speciality Hospital, Delhi, and DY Patil Hospital, Maharashtra, India, and were assigned anonymous barcodes at



CSIR-IGIB. The patients included in the study were confirmed SARS-CoV-2 positive by RT-PCR. The electronic medical records including patient demography, clinical symptoms, blood profile, comorbidities recorded at the time of admission, drugs administered to the patients during the course of infection were recorded and analyzed. The patients were classified into two major groups, *Recovered* and *Mortality*, based on the clinical outcome, as captured from the clinical data.

RT-PCR

Viral RNA from VTM solutions was isolated using QIAmp viral mini kit, Qiagen, Cat. No. 52906, and SARS-CoV-2 detection and quantification were performed using TRUPCR SARS-CoV-2 kit (3B Black Bio Biotech India Ltd., Cat. No. 3B304) with a cycle threshold of 35. The detailed methodology has been published in the earlier publications from our lab.

SARS-CoV-2 Whole-Genome Sequencing

Sequencing of the SARS-CoV-2 genomes was done using a combination of Oxford Nanopore Technology (ONT) and Illumina (MiSeq).

In brief, 100 ng of total RNA was used for double-stranded cDNA synthesis. This involves first-strand cDNA synthesis using Superscript IV (Thermo Fisher Scientific, Cat. No. 18091050), followed by RNase H digestion of ssRNA and second-strand synthesis by DNA polymerase-I large (Klenow) fragments

(New England Biolabs, Cat. No. M0210S). Double-stranded DNA thus obtained was purified using 1.8× AMPure XP beads to sample ratio (Beckman Coulter, Cat. No. A63881). SARS-CoV-2 genome was then amplified from 100 ng of the purified cDNA with the ARTIC V3 primer protocol. For sequencing library preparation using Oxford Nanopore sequencing, library preparation consists of End Repair/da tailing, Native Barcode Ligation, and Adapter Ligation. This was performed with 200 ng of the multiplexed PCR amplicons according to ONT library preparation protocol-PCR tiling of COVID-19 virus (Version: PTC_9096_v109revE_06Feb2020). Sequencing in sets of 24 barcoded samples was performed on the MinION Mk1C platform.

Sequencing library preparation for Illumina was performed using 100 ng of purified ARTIC PCR product using the Illumina DNA Prep kit (Illumina, Cat. No. 20018705). The process involves tagmentation followed by post tagmentation cleanup and amplification by PCR, leading to indexed DNA fragments, which was purified prior to sequencing. The quality and quantity of the sequencing library were checked using an Agilent 2100 Bioanalyzer with high-sensitivity DNA chip and the Qubit dsDNA HS Assay kit, respectively. A loading concentration of 11 pM was prepared by denaturing and diluting the libraries in accordance with the MiSeq System Denature and Dilute Libraries Guide (Illumina, Document No. 15039740 v10). Sequencing was performed on the MiSeq system using the MiSeq Reagent Kit v3 (150 cycles) at 2 × 75 bps read length.

Sequencing Data Analysis

The ARTIC end-to-end pipeline was used for the analysis of ONT MinION raw fast5 files up to variant calling. Raw fast5 files of samples were base called and demultiplexed using Guppy basecaller that uses the basecalling algorithms of Oxford Nanopore Technologies (Nanopore Community) with phred quality cutoff score >7 on GPU-linux-accelerated computing machine. Reads having Phred quality score less than 7 were discarded to filter the low-quality reads. The resultant demultiplexed fastq was normalized by read length of 300–500 (approximate size of amplicons) for further downstream analysis and aligned to the SARS-CoV-2 reference (MN908947.3) using the aligner Minimap2 v2.17 (Li, 2018). Nanopolish (Loman et al., 2015) was used to index raw fast5 files for variant calling from the minimap output files. To create consensus fasta, bcftools v1.8 was used with normalized minimap2 output.

For the Illumina sequencing data, FASTQC was performed for all the raw fastq files generated to check the Phred quality score of all the sequences (FastQC A Quality Control tool for High Throughput Sequence Data, Babraham Bioinformatics). A phred quality score threshold of >20 was used for filtering reads from all the samples. Next, adapter trimming was performed using Trim Galore tool (–TrimGalore, Babraham Bioinformatics) and alignment of the sequences was performed using HISAT2 algorithm (Kim et al., 2019) on human data build hg38 to remove any human read contamination. BEDTools was used to generate the consensus fasta using the unaligned/filtered reads, and variant calling was performed using the high-quality reads (Quinlan and Hall, 2010).

Phylogenetic Analysis

The Wuhan reference genome for SARS-CoV-2 (NC_045512.2) was used to perform multiple sequence alignment using MAFFT (v7.475) (Katoh et al., 2002). The alignment was manually trimmed, and a phylogenetic tree was constructed using the IQ-tree (Nguyen et al., 2015). Lineage classification was performed using PANGOLIN. The phylogenetic analysis was visualized using FIGTREE software (<http://tree.bio.ed.ac.uk/software/figtree/>).

Mutation Analysis

The vcf files were used to get the frequency of the mutations in the samples for the mutational spectra analysis within the recovered and mortality patients, and snpEff v5.0 (Cingolani et al., 2012) has been used to perform the variant annotation information such as the variant definition and effect of the variants. SnpEff database has been created by “SnpEff build” using the Wuhan reference NC_045512.2. Furthermore, global frequency of mutations was checked against a global dataset available at 2019 Novel Coronavirus Resource (2019nCoV), China National Center for Bioinformation (CNCB) (Song et al., 2020). The lineages defining mutations for different VOCs/VOIs were curated from two public databases (<https://covariants.org/shared-mutations>, <https://github.com/cov-lineages/constellations>). A comparative study of mutation frequency in recovered and mortality patients with the respective cumulative

global frequency of mutations plotted as the lollipop plot representing the mutations was generated in R v4.1.0 using g3viz (Guo et al., 2020), rtracklayer (Lawrence et al., 2009), and trackViewer (Ou and Zhu, 2019) packages followed by data visualization using the ggplot2 (R package ggplot2 v3.3.5) package. Variant position along the SARS-CoV-2 genome is indicated in the plot, which is used to show the frequent mutations with the global frequency, which is later selected as VOC/VOI.

Quantification and Statistical Analysis

The statistical analysis was performed by R software. The nonparametric Fisher exact test of significance for independence between two categorical variables was performed to check the difference of mutation profile for the mortality vs. recovered patients. All p -values were calculated from 2-sided tests using 0.05 as the significance level. The direction of the association between mutation and group was calculated using phi-coefficient correlation (r_ϕ) by measuring the strength of association.

Generation of Wild-Type and Mutant N-Protein Models

MD was performed on N protein (uniprot id-P0DTC9), with fasta sequence retrieved from the UniProt database and the protein being modeled using Phyre (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>). The mutation for recovered (S194*) and mortality (S194L) was incorporated using PyMOL (PyMOL | pymol.org). The protein complex simulation is done with 200 ns of timescale for understanding the dynamic behavior and the interaction pattern (Shafreen et al., 2013; Selvaraj et al., 2018). Structures are analyzed using the CHARM-AA force field solvated by the SPC16 water model within a periodic boundary box of distance 1.0 nm, fixed between the protein and cubic box (Nguyen et al., 2014; Bandaru et al., 2017). For neutralizing the whole system, the accurate concentration of (Na^+/Cl^-) ions is added based on the rebalancing charges. Initial energy minimization is appended with the prepared complex systems for 1,000 steps of steepest descent algorithm *via* a tolerance of 10 kJ/mol/nm to avoid the steric clashes (Kumar et al., 2021). Thermostat coupling is set with a reference temperature of 300K using Berendsen thermostat and pressure coupling with 1.0 bar reference pressure using Parrinello-Rahman along with periodic boundary conditions with cutoffs for Lennard-Jones and Coulomb interactions. The particle mesh Ewald method is used for calculating the long-range interactions for biomolecular systems (Wohlert and Edholm, 2004).

After the initial minimization step, the whole system is again well equilibrated at 2,000 ps at 300K and 1 bar pressure in Number of particles system volume and temperature (NVT) and Number of particles system pressure and temperature (NPT) ensembles (Childers and Daggett, 2018). Final MD simulation step is processed for three protein complexes (wild-type, recovered, and mortality N-proteins) with respect to the timescale of 200 ns. For MD simulation analysis, the results of root mean square deviation (RMSD) and root mean square

fluctuation (RMSF) are performed using the XMGRACE v5.1.2 and GROMACS v.5.1.2 tools.

RESULTS

Patient Clinical Demographics and Classification

Our initial dataset was curated to exclude patients with any missing data for important parameters, resulting in a total sample size of 246 patients. These patients were classified based on the clinical outcome, with 228 *Recovered* and 18 *Mortality* patients. Regression analysis using the chi-square test and Kruskal–Wallis test highlighted a significant distribution of specific clinical parameters. Symptoms such as shortness of breath (p -value = 0.037) and abdominal pain (p -value = 0.001) were observed to be significantly associated with the mortality patients (**Figure 2A**). Among the patients with comorbidities, hypertension (p -value = 0.01) and hypothyroidism (p -value = 0.01) were significantly correlated with the mortality patients (**Figure 2C**). Notably, most of the patients who recovered did

not significantly present any comorbidities (p -value = 0.001). Also, a difference in age was seen among the recovered and mortality patients (p -value = 0.01) (**Figure 2B**). All the mortality patients were on respiratory support, whereas only 28% those recovered required respiratory support. The clinical parameters and analysis are shown in **Table 1**.

SARS-CoV-2 Phylogenetic Analysis and Distribution

Phylogenetic analysis of 246 SARS-CoV-2 genomes using PANGOLIN (lineage) and Nextstrain (clades) showed the dominant occurrence of B.6 (29.2%), B.1.36 (9.82%), B.1 (10.56%), B.1.1 (4.47%), B.1.1.306 (4.47%), and B.6.6 (4.47%), whereas clade distribution highlighted the dominant prevalence of 19A (52.84%) followed by 20A (34.55%), as mentioned in **Supplementary Table S1**. We observed that the least number of patients belonged to the 20B (13%) clade, whereas 19A has the highest number of patients. Delving further, within the recovered samples, major distribution for the lineage was B.6 (31.57%), B.1.36 (12.28%), and B.1 (9.64%), whereas 19A was the dominant clade (45.17%); and within mortality, major distribution for the

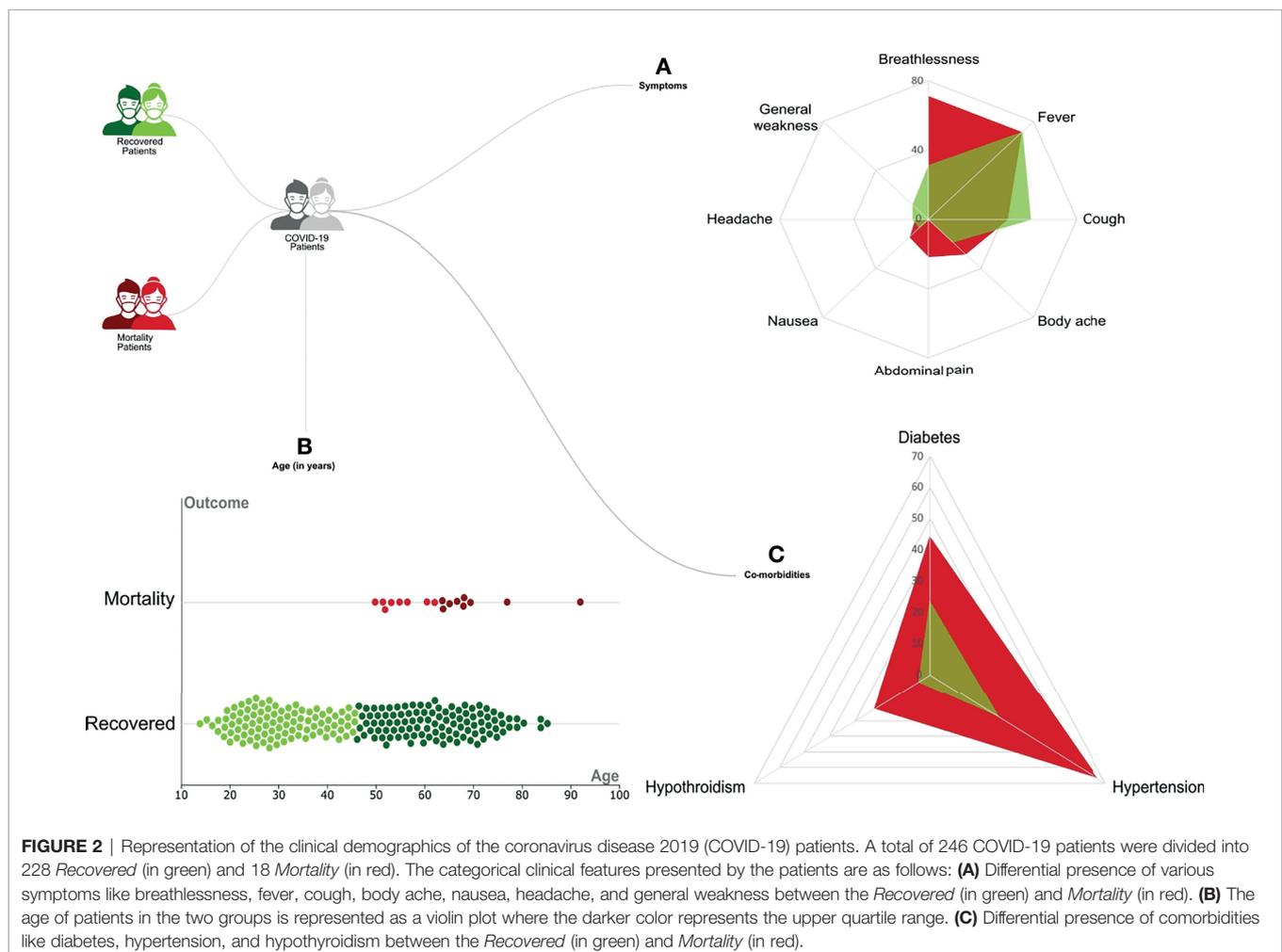


TABLE 1 | Clinical summary of the COVID-19 patients highlighting clinical parameters across *recovered* and *mortality*.

Groups	Total (n = 246)	Mortality (n = 18)	Recovered (n = 228)	p-value
Gender F M	79 166	3 14*	76 152	0.18 ^b
Age	48 (30–63)	63 (55–68)	46 (30–63)	<0.001^a
Respiratory Support	83 (33.73%)	18 (100%)	65 (28.50%)	–
E gene	24.19 (21.21–27.51)	22.48 (19.62–27.69)	24.34 (21.56–27.51)	0.21 ^a
RdRp gene	25.11 (21.51–28.91)	23.67 (20.59–28)	25.24 (21.57–28.94)	0.26 ^a
Symptoms				
Shortness of Breath	82 (33.33%)	10 (55.56%)	72 (31.57%)	0.037^b
Fever	174 (70.73%)	10 (55.56%)	164 (71.92%)	0.141 ^b
Sore Throat	132 (53.65%)	6 (33.33%)	126 (55.26%)	0.72 ^b
Body ache	47 (19.10%)	4 (22.22%)	43 (18.85%)	0.72 ^b
Abdominal pain	4 (1.62%)	3 (16.67%)	1 (0.43%)	<0.001^b
Nausea	18 (7.31%)	2 (11.11%)	16 (7.01%)	0.52 ^b
Headache	20 (8.13%)	1 (5.56%)	19 (8.33%)	0.64 ^b
General weakness	28 (11.38%)	0 (0)	28 (12.28%)	–
No Co-morbidities	141 (57.31%)	4 (22.22%)	137 (60.08%)	<0.001^b
Comorbidities				
Diabetes	62 (25.20%)	14 (77.77%)	91 (39.92%)	0.05 ^b
Hypertension	74 (30.08%)	8 (44.44%)	54 (23.68%)	<0.001^b
Hypothyroidism	14 (5.69%)	12 (66.66%)	62 (27.19%)	<0.001^b
Asthma	4 (1.62%)	4 (22.22%)	10 (4.38%)	–
Hospital stay	11 (5–14)	0 (0%)	4 (1.75%)	–
		13 (6–18)	11 (5–14)	0.34 ^a

Data are shown as median [interquartile range (IQR)] or n (%).

^aKruskal–Wallis test.

^bChi-square test.

*Missing data.

Values of significance are highlighted in bold.

lineages was B.1.36 (33.33%) followed by B.1 (22.22%) and 20A clade (83.33%) (**Figure 3**).

SARS-CoV-2 Mutation Profile Within Different Clinical Outcomes and Its Frequency Flip

Individual mutation analysis across our sample cohort revealed a total of 3,449 SARS-CoV-2 genome mutations. Importantly, we observed that within the mortality cases in our cohort, there was a higher number of mutations per sample, 28.94, than the recovered group, which had 13.60 mutations/sample. This observation made us explore the possible reasons for higher mutations per sample within mortality and their role in disease outcome. The Fisher exact test that was performed across our cohort revealed a total of 31 mutations to be statistically significant, with a *p*-value <0.005 associated with the clinical outcome (**Supplementary Table S2, Figure 4A**). Further correlation analysis of these 31 mutations resulted in the stratification of five mutations into the recovered and 26 mutations with the mortality patients. When compared with their global frequency, 7/31 significant mutations showed a low global presence indicating a frequency flip from their high distribution observed in our samples. The segregation of these seven mutations exhibited frequency flips with five mutations within the recovered group and two mutations in the mortality group, showing a significant difference in their association with the respective groups (*p*-value = 0.0001). Notably, all the mutations associated with the recovered group (S2015R, Y789Y, T2016K, S194*, A4489V) showed a frequency flip, whereas two mutations associated with the mortality group

showed a shift in their frequency from our cohort to global level. The recovered group displaying a frequency flip for all the mutations could indicate that these mutations are highly likely to be deselected by the virus during the course of its evolution. On the other hand, among the rest of the mutations in the mortality group that did not exhibit any frequency flip (24 mutations), we observed five mutations (C241T, F924F, P4715L, D614G, Q75H) with a high cohort and global distribution. Interestingly, all these five mutations subsequently became a part of VOIs/VOCs such as B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma), B.1.617.1 (Kappa), B.1.617.2 (Delta), B.1.525 (Eta), B.1.526 (Iota), C.37 (Lambda), B.1.621 (Mu), and presently being seen in BA.1 and BA.2 (Omicron) (**Figure 4B**). Hence, the presence of these mutations at high frequency throughout the pandemic and its selection in VOIs/VOCs could possibly suggest the propensity of the virus to retain these mutations, thereby enhancing the functional characteristics of SARS-CoV-2.

Analysis of Molecular Dynamics Trajectories for Effect of Mutation on the Protein Structure

Among the significant mutations, we found S194L in the N protein to be highly significant (*p*-value = 2.35481E-13) and exclusive to the mortality patients. Similarly, a mutation at the same position, S194*, in the N protein was observed in the recovered as well, which was unique and significant with a *p*-value of 0.0045. Evidently, N protein is reported to be more conserved in comparison to other proteins such as Spike and membrane glycoprotein, which plays an essential role in viral genome packaging (Cubuk et al., 2021). Henceforth, we

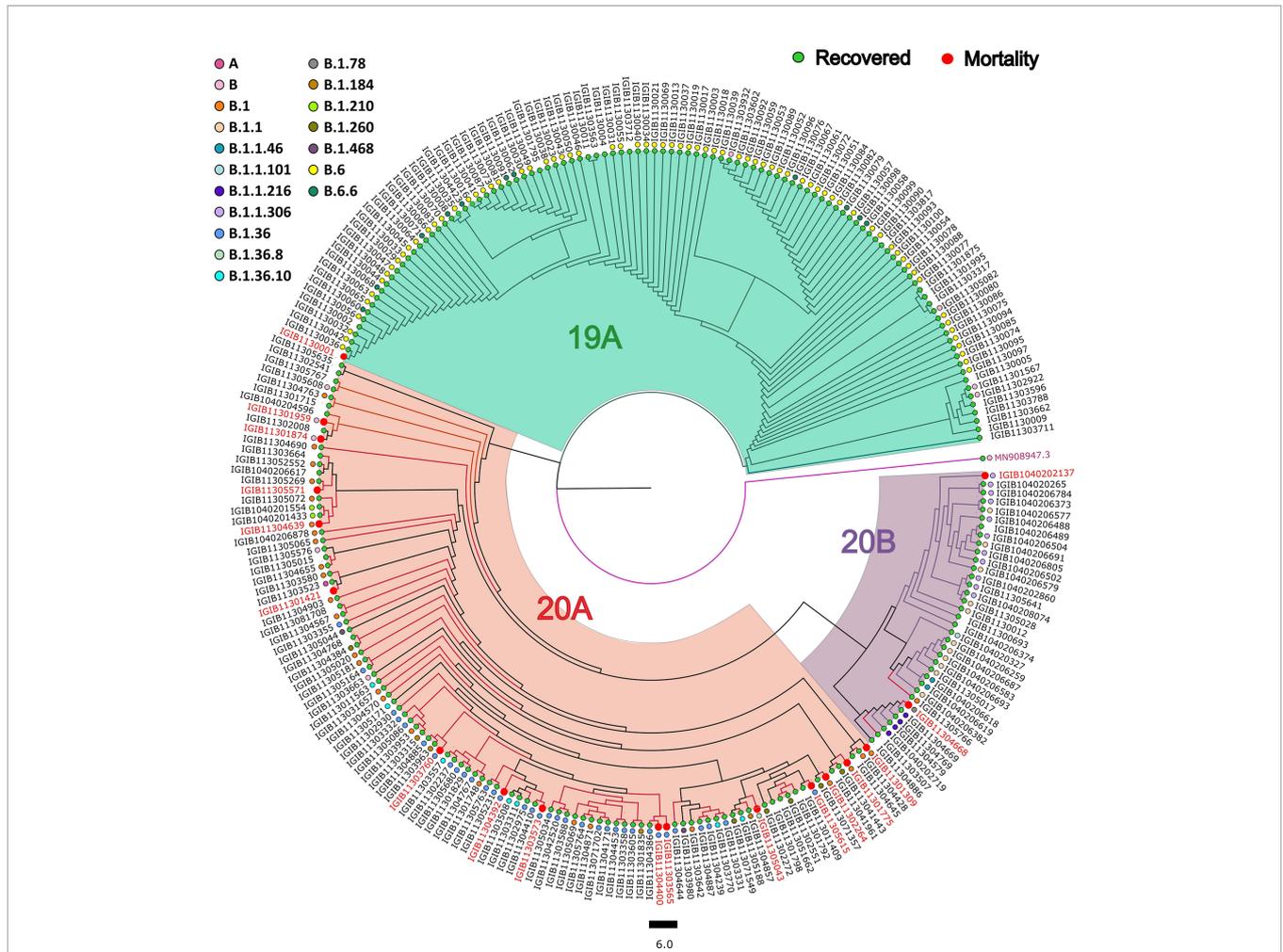


FIGURE 3 | Phylogenetic analysis of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genomes. Showing the distribution of 246 SARS-CoV-2 clades and lineages among COVID-19 patients compared with the wild-type strain.

performed MD simulations to study and analyze the consequences of these two mutations compared to the wild-type N protein, in an endeavor to understand its association with the disease outcome. The RMSD was calculated for wild-type and mutants to elucidate changes in the overall stability of the protein that was considered the primary criterion for calculating the overall convergence of the system. The deviations for all the three systems (wild-type, S194*, and S194L) were found to be between 0.1 and 1.1 nm (**Supplementary Table S3**). A state of equilibrium was achieved for both the mutant proteins at 115 ns; whereas for the wild-type, it was at 150 ns. Subsequently, we calculated the fluctuations of the residues using RMSF (**Figure 5A**). There were local changes observed, with higher peaks of fluctuations seen at 10–18 and 179th positions in the case of S194*, reaching up to 0.8–1.2 nm and 0.7 nm, respectively (**Supplementary Table S4**). Fluctuation was seen more in N protein with S194* as compared to the S194L mutant N protein (**Figure 5B**). It is indicative of the recovered patients exhibiting a comparatively unstable protein *vis-à-vis* the mortality patients.

DISCUSSION

There is a significant amount of SARS-CoV-2 genome sequencing data being deposited in public databases in real time. However, our comprehension of different patient clinical presentations and their correlation with the SARS-CoV-2 mutations remains limited (Esper et al., 2021). Toward that, we made an effort to analyze and correlate the SARS-CoV-2 mutations within the recovered and mortality patients along with their impacts on the protein structure. Clinically, similar to other studies from around the globe, advanced age and underlying conditions such as hypertension and hypothyroidism were significantly associated with the mortality of COVID-19 patients (Wang et al., 2020; Zhou et al., 2020; Huang et al., 2021; Li et al., 2021). With reference to presenting symptoms as potential predictors for disease severity, a rare symptom, abdominal pain, was also observed to be significantly associated with the mortality group as reported by Li et al. (2021).

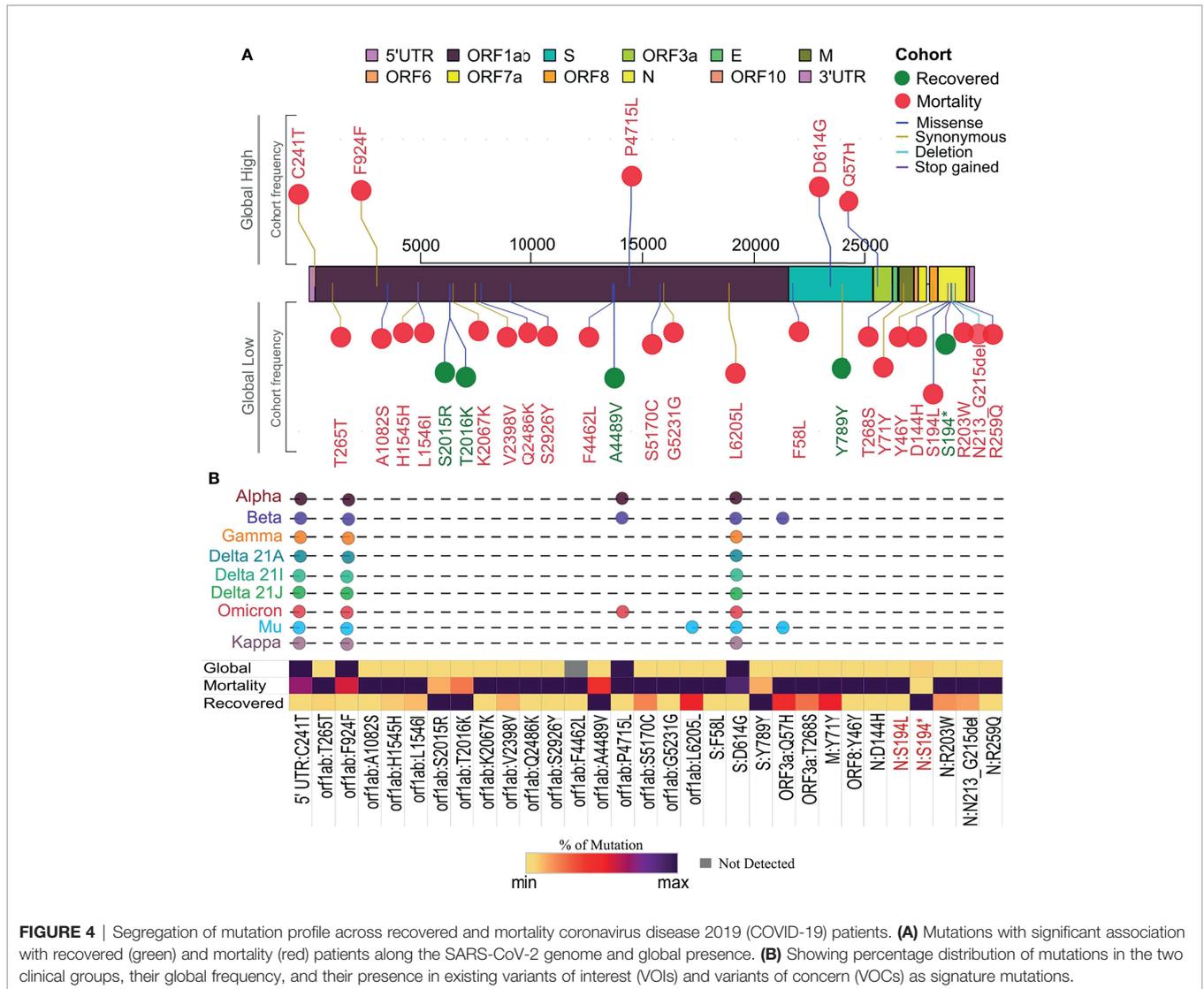


FIGURE 4 | Segregation of mutation profile across recovered and mortality coronavirus disease 2019 (COVID-19) patients. **(A)** Mutations with significant association with recovered (green) and mortality (red) patients along the SARS-CoV-2 genome and global presence. **(B)** Showing percentage distribution of mutations in the two clinical groups, their global frequency, and their presence in existing variants of interest (VOIs) and variants of concern (VOCs) as signature mutations.

Genomic analysis of 246 samples demonstrated that the lineage B.6 had a dominant occurrence followed by B.1.36, B.1, B.1.1, B.1.1.306, B.6.6, and 20A for the clade, respectively, which is consistent with the lineage distribution observed during the first wave in India (Chong et al., 2020; Jha et al., 2021). Moreover, a significant frequency flip (p -value = 0.0001) was seen for the five mutations, S2015R, Y789Y, T2016K, S194*, and A4489V, in the recovered with a low global presence, indicating that the mutations associated with COVID-19 recovery are deselected by the virus during its evolution. Intriguingly, 5/26 significant mutations (C241T, Q57H, F924F, P4715L, D614G) in the mortality patients did not exhibit any frequency flip with a high frequency occurrence observed in the cohort as well as the global level. This could be indicative that the mutations associated with severe outcome are preferably selected and present globally. A study by Mercatelli and Giorgi (2020) also demonstrated the mutations C241T, F924F, and P4715L to have a similar frequency as D614G in their cohort as well, which shows their co-occurrence in the viral genome.

It is interesting that these five mutations are the characteristic mutations of Type VI SARS-CoV-2 strain and show a high pairwise allelic association. Moreover, the mutation P4715L in RdRp protein affects the viral replication (Pachetti et al., 2020), and the spike mutation D614G influences the viral interaction with the ACE2 receptor, thereby enhancing the overall fitness of the virus (Korber et al., 2020; Toyoshima et al., 2020). Also, the Type VI strain is observed to be present in high frequency in most of the countries throughout the pandemic that could possibly imply that these mutations provide stability and positively influence the virulence by enhancing the viral transmission and pathogenicity (Yang et al., 2020a). A study by Bianchi et al. (2021) reported that the mutation Q57H, which is observed in our mortality patients, is consistently present in high frequency at all the time points. Another study describes the loss of function of orf3b protein, an accessory protein, due to the truncation of orf3a protein by Q57H mutation. Furthermore, they also found that the loss of orf3b coincides with the emergence of D614G spike mutation that strikingly correlates with our observation of

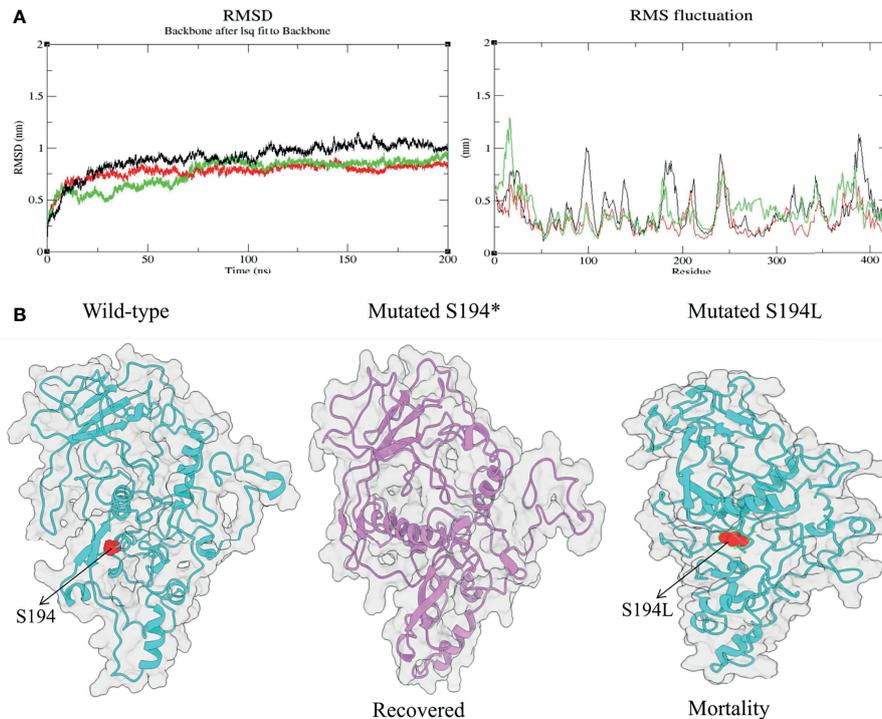


FIGURE 5 | Structural changes during molecular dynamic simulation: **(A)** Superimposed root mean square deviation (RMSD) and root mean square fluctuation (RMSF) spectrum of wild-type (black) and mutant proteins [S194* (recovered) shown in green and S194L (mortality) shown in red] during 200 ns of molecular dynamics simulation period. **(B)** The obtained simulated structures of mutated N protein with wild type.

D614G being present at high frequency in our mortality patients (Lam et al., 2020). Also, the mutation S194L found in the Nucleocapsid gene was a highly significant mutation observed across our cohort and is also associated with mortality. Several other studies have also reported that this mutation is deleterious and correlated with fatal outcomes (Limaye et al., 2021; Mehta et al., 2021; SeyedAlinaghi et al., 2021). This could possibly indicate their crucial role in elevating the pathogenicity of the virus, thereby contributing to increased disease severity and clinical outcome. Such significant observation can aid us in tracking and prioritizing the upcoming SARS-CoV-2 variants as well as monitoring the disease progression for early interventions to possibly manage fatal outcomes.

To understand the correlation of mutations with severity outcome, one of the important aspects is its functional consequences at the protein structural level. The MD simulations performed for the mutations S194* and S194L from the recovered and mortality group, respectively, revealed RMSD values, which was evaluated for both the mutations and the wild-type, implying a state of equilibrium achieved for all the three systems. Although mutation S194L is previously reported to be associated with mortality as seen in our cohort, its effect on the protein structure has not been explored and elucidated (Joshi et al., 2021). A comparative analysis of the parameter RMSF value demonstrated that the recovered mutant protein S194* showed more fluctuations in comparison with the wild type and

the protein mutation at the same position associated with the mortality group. This suggests the instability of the mutation S194* in the recovered group, which is exactly correlating with its low frequency occurrence at the global level as compared to S194L in the mortality group.

In summary, the findings from our study highlight the significance of the mutations in modulating COVID-19 presentations and outcomes. Assessing the clinical symptoms of COVID-19 patients is largely important, since patients manifest a wide spectrum of symptoms, from asymptomatic to symptomatic and a subset leading to fatal outcomes. Disease symptoms and its progression are modulated by several factors that include viral characteristics, environmental factors, host immunity, and comorbidities. An integrative analysis of these disease modulators can aid us in devising strategies for the disease symptom management and treatment, if possible. The co-analysis of mutations and the clinical outcome would be extremely beneficial for diagnosing mutations susceptible to increased disease severity and also to evaluate the effects of control measures implemented during the course of the pandemic. This approach can assist us in tracking the SARS-CoV-2 evolution and its upcoming variants, along with providing insights to its association with the disease severity or outcome for the amelioration of the disease. Thus, we can identify the subgroups of patients predicted to have a poor prognosis and can be assisted with required treatments prior to any detrimental outcomes. Hence, this approach can lead to a better

utilization of the healthcare management system and could aid us in a sustainable pandemic preparedness.

DATA AVAILABILITY STATEMENT

The clinical dataset collected and analyzed as part of this study is available in the **Supplementary Material** as **Data Sheet 1**. The SARS-CoV-2 genomic data have been uploaded to GISAID.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by CSIR-IGIB's Human Ethics Committee Clearance (Ref No: CSIR-IGIB/IHEC/2020-21/01). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RM, PM, VR, AK and PrM performed the analysis. RM, PM, AS, VR, SS, AK, PrM, PD, SP and RP wrote the article. RP designed, conceptualized, implemented and coordinated the study. SB, BT, ShS, RJK and MGJ shared the clinical samples and clinical data. All authors contributed to the article and approved the final submission.

FUNDING

This research was funded by the Bill and Melinda Gates Foundation (BMGF) [Project code : INV-033578], Indo-US

REFERENCES

- Bakhshandeh, B., Jahanafrooz, Z., Abbasi, A., Goli, M. B., Sadeghi, M., Mottaqi, M. S., et al. (2021). Mutations in SARS-CoV-2; Consequences in Structure, Function, and Pathogenicity of the Virus. *Microb. Pathog.* 154, 104831. doi: 10.1016/j.micpath.2021.104831
- Bandaru, S., Alvala, M., Nayarisseri, A., Sharda, S., Goud, H., Mundluru, H. P., et al. (2017). Molecular Dynamic Simulations Reveal Suboptimal Binding of Salbutamol in T164I Variant of β_2 Adrenergic Receptor. *PLoS One* 12, e0186666. doi: 10.1371/journal.pone.0186666
- Barona-Gómez, F., Delaye, L., Díaz-Valenzuela, E., Plisson, F., Cruz-Pérez, A., Díaz-Sánchez, M., et al. (2021). Phylogenomics and Population Genomics of SARS-CoV-2 in Mexico During the Pre-Vaccination Stage Reveals Variants of Interest B.1.1.28.4 and B.1.1.222 or B.1.1.519 and the Nucleocapsid Mutation S194L Associated With Symptoms. *Microb. Genom.* 7. doi: 10.1099/mgen.0.000684
- Bianchi, M., Borsetti, A., Ciccozzi, M., and Pascarella, S. (2021). SARS-Cov-2 ORF3a: Mutability and Function. *Int. J. Biol. Macromol.* 170, 820–826. doi: 10.1016/j.ijbiomac.2020.12.142
- Childers, M. C., and Daggett, V. (2018). Validating Molecular Dynamics Simulations Against Experimental Observables in Light of Underlying Conformational Ensembles. *J. Phys. Chem. B* 122, 6673–6689. doi: 10.1021/acs.jpcc.8b02144
- Chong, Y. M., Sam, I.-C., Chong, J., Kahar Bador, M., Ponnampalavanar, S., Syed Omar, S. F., et al. (2020). SARS-CoV-2 Lineage B.6 Was the Major Contributor to Early Pandemic Transmission in Malaysia. *PLoS Negl. Trop. Dis.* 14, e0008744. doi: 10.1371/journal.pntd.0008744
- Science and Technology Forum (IUSSTF) [Project code : CLP-0033], Foundation for Innovative New Diagnostics (FIND) [Project code : CLP-0038], and AIDS Healthcare Foundation (AHF) [Project code : CLP-0043].

ACKNOWLEDGMENTS

We acknowledge the role of Dr. Aradhita Baral as research manager and coordinator for the funders. The authors also acknowledge the support of Anil Kumar and Nisha Rawat toward SARS-CoV-2 sample transport and sample management.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcimb.2022.868414/full#supplementary-material>

Supplementary Table 1 | Distribution of clade and lineages across 246 cohort samples.

Supplementary Table 2 | Individual mutations across our sample cohort with its statistical significance (*p-value*) and phi-coefficient correlation (*r ϕ*) value.

Supplementary Table 3 | Results of Root mean square deviation (RMSD) during simulation analysis with respect to the timescale of 200ns.

Supplementary Table 4 | Results of Root mean square fluctuation (RMSF) during simulation analysis with respect to the timescale of 200ns.

Supplementary Data Sheet 1 | GISAID Accession numbers associated with samples used in the Cohort.

- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff: SNPs in the Genome of *Drosophila Melanogaster* Strain W1118; Iso-2; Iso-3. *Fly (Austin)* 6, 80–92. doi: 10.4161/fly.19695
- Cubuk, J., Alston, J. J., Incicco, J. J., Singh, S., Stuchell-Brereton, M. D., Ward, M. D., et al. (2021). The SARS-CoV-2 Nucleocapsid Protein Is Dynamic, Disordered, and Phase Separates With RNA. *Nat. Commun.* 12, 1936. doi: 10.1038/s41467-021-21953-3
- Esper, F. P., Cheng, Y.-W., Adhikari, T. M., Tu, Z. J., Li, D., Li, E. A., et al. (2021). Genomic Epidemiology of SARS-CoV-2 Infection During the Initial Pandemic Wave and Association With Disease Severity. *JAMA Netw. Open* 4, e217746. doi: 10.1001/jamanetworkopen.2021.7746
- Guo, X., Zhang, B., Zeng, W., Zhao, S., and Ge, D. (2020). G3viz: An R Package to Interactively Visualize Genetic Mutation Data Using a Lollipop-Diagram. *Bioinformatics* 36, 928–929. doi: 10.1093/bioinformatics/btz631
- Gupta, A. M., Chakrabarti, J., and Mandal, S. (2020). Non-Synonymous Mutations of SARS-CoV-2 Leads Epitope Loss and Segregates Its Variants. *Microbes Infect.* 22, 598–607. doi: 10.1016/j.micinf.2020.10.004
- Huang, C., Soleimani, J., Herasevich, S., Pinevich, Y., Pennington, K. M., Dong, Y., et al. (2021). Clinical Characteristics, Treatment, and Outcomes of Critically Ill Patients With COVID-19: A Scoping Review. *Mayo Clin. Proc.* 96, 183–202. doi: 10.1016/j.mayocp.2020.10.022
- Hu, J., Peng, P., Cao, X., Wu, K., Chen, J., Wang, K., et al. (2022). Increased Immune Escape of the New SARS-CoV-2 Variant of Concern Omicron. *Cell. Mol. Immunol.* 9, 267–275. doi: 10.1038/s41423-021-00836-z
- Jha, N., Hall, D., Kanakan, A., Mehta, P., Maurya, R., Mir, Q., et al. (2021). Geographical Landscape and Transmission Dynamics of SARS-CoV-2

- Variants Across India: A Longitudinal Perspective. *Front. Genet.* 12, 753648. doi: 10.3389/fgene.2021.753648
- Joshi, M., Puvar, A., Kumar, D., Ansari, A., Pandya, M., Raval, J., et al. (2021). Genomic Variations in SARS-CoV-2 Genomes From Gujarat: Underlying Role of Variants in Disease Epidemiology. *Front. Genet.* 12, 586569. doi: 10.3389/fgene.2021.586569
- Justo Arevalo, S., Zapata Sifuentes, D., Huallpa, C. J., Landa Bianchi, G., Castillo Chávez, A., Garavito-Salini Casas, R., et al. (2021). Global Geographic and Temporal Analysis of SARS-CoV-2 Haplotypes Normalized by COVID-19 Cases During the Pandemic. *Front. Microbiol.* 12, 612432. doi: 10.3389/fmicb.2021.612432
- Kannan, S. R., Spratt, A. N., Sharma, K., Chand, H. S., Byrareddy, S. N., and Singh, K. (2022). Omicron SARS-CoV-2 Variant: Unique Features and Their Impact on Pre-Existing Antibodies. *J. Autoimmun.* 126, 102779. doi: 10.1016/j.jaut.2021.102779
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436
- Kaushal, N., Gupta, Y., Goyal, M., Khaiboullina, S. F., Baranwal, M., and Verma, S. C. (2020). Mutational Frequencies of SARS-CoV-2 Genome During the Beginning Months of the Outbreak in USA. *Pathogens* 9, 565. doi: 10.3390/pathogens9070565
- Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-Based Genome Alignment and Genotyping With HISAT2 and HISAT-Genotype. *Nat. Biotechnol.* 37, 907–915. doi: 10.1038/s41587-019-0201-4
- Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., et al. (2020). Tracking Changes in SARS-CoV-2 Spike: Evidence That D614G Increases Infectivity of the COVID-19 Virus. *Cell* 182, 812–827.e19. doi: 10.1016/j.cell.2020.06.043
- Kumar, D., Kumari, K., Jayaraj, A., Kumar, V., Kumar, R. V., Dass, S. K., et al. (2021). Understanding the Binding Affinity of Nospapines With Protease of SARS-CoV-2 for COVID-19 Using MD Simulations at Different Temperatures. *J. Biomol. Struct. Dyn.* 39, 2659–2672. doi: 10.1080/07391102.2020.1752310
- Lai, C.-C., Shih, T.-P., Ko, W.-C., Tang, H.-J., and Hsueh, P.-R. (2020). Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and Coronavirus Disease-2019 (COVID-19): The Epidemic and the Challenges. *Int. J. Antimicrob. Agents* 55, 105924. doi: 10.1016/j.ijantimicag.2020.105924
- Lam, J.-Y., Yuen, C.-K., Ip, J. D., Wong, W.-M., To, K. K.-W., Yuen, K.-Y., et al. (2020). Loss of Orf3b in the Circulating SARS-CoV-2 Strains. *Emerg. Microbes Infect.* 9, 2685–2696. doi: 10.1080/22221751.2020.1852892
- Lawrence, M., Gentleman, R., and Carey, V. (2009). Rtracklayer: An R Package for Interfacing With Genome Browsers. *Bioinformatics* 25, 1841–1842. doi: 10.1093/bioinformatics/btp328
- Li, H. (2018). Minimap2: Pairwise Alignment for Nucleotide Sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191
- Li, J., Huang, D. Q., Zou, B., Yang, H., Hui, W. Z., Rui, F., et al. (2021). Epidemiology of COVID-19: A Systematic Review and Meta-Analysis of Clinical Characteristics, Risk Factors, and Outcomes. *J. Med. Virol.* 93, 1449–1458. doi: 10.1002/jmv.26424
- Limaye, S., Kasibhatla, S. M., Ramtirthkar, M., Kinikar, M., Kale, M. M., and Kulkarni-Kale, U. (2021). Circulation and Evolution of SARS-CoV-2 in India: Let the Data Speak. *Viruses* 13, 2238. doi: 10.3390/v13112238
- Loman, N. J., Quick, J., and Simpson, J. T. (2015). A Complete Bacterial Genome Assembled *De Novo* Using Only Nanopore Sequencing Data. *Nat. Methods* 12, 733–735. doi: 10.1038/nmeth.3444
- Majumdar, P., and Niyogi, S. (2020). ORF3a Mutation Associated With Higher Mortality Rate in SARS-CoV-2 Infection. *Epidemiol. Infect.* 148, e262. doi: 10.1017/S0950268820002599
- Mehta, P., Alle, S., Chaturvedi, A., Swaminathan, A., Saifi, S., Maurya, R., et al. (2021). Clinico-Genomic Analysis Reveals Mutations Associated With COVID-19 Disease Severity: Possible Modulation by RNA Structure. *Pathogens* 10, 1109. doi: 10.3390/pathogens10091109
- Mercatelli, D., and Giorgi, F. M. (2020). Geographic and Genomic Distribution of SARS-CoV-2 Mutations. *Front. Microbiol.* 11, 1800. doi: 10.3389/fmicb.2020.01800
- Mousavizadeh, L., and Ghasemi, S. (2021). Genotype and Phenotype of COVID-19: Their Roles in Pathogenesis. *J. Microbiol. Immunol. Infect.* 54, 159–163. doi: 10.1016/j.jmii.2020.03.022
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Nguyen, T. T., Viet, M. H., and Li, M. S. (2014). Effects of Water Models on Binding Affinity: Evidence From All-Atom Simulation of Binding of Tamiflu to A/H5N1 Neuraminidase. *ScientificWorldJournal* 2014, 536084. doi: 10.1155/2014/536084
- Oude Munnink, B. B., Worp, N., Nieuwenhuijse, D. F., Sikkema, R. S., Haagmans, B., Fouchier, R. A. M., et al. (2021). The Next Phase of SARS-CoV-2 Surveillance: Real-Time Molecular Epidemiology. *Nat. Med.* 27, 1518–1524. doi: 10.1038/s41591-021-01472-w
- Ou, J., and Zhu, L. J. (2019). Trackviewer: A Bioconductor Package for Interactive and Integrative Visualization of Multi-Omics Data. *Nat. Methods* 16, 453–454. doi: 10.1038/s41591-019-0430-y
- Pachetti, M., Marini, B., Benedetti, F., Giudici, F., Mauro, E., Storici, P., et al. (2020). Emerging SARS-CoV-2 Mutation Hot Spots Include a Novel RNA-Dependent-RNA Polymerase Variant. *J. Transl. Med.* 18, 179. doi: 10.1186/s12967-020-02344-6
- Pereira, N. L., Ahmad, F., Byku, M., Cummins, N. W., Morris, A. A., Owens, A., et al. (2021). COVID-19: Understanding Inter-Individual Variability and Implications for Precision Medicine. *Mayo Clin. Proc.* 96, 446–463. doi: 10.1016/j.mayocp.2020.11.024
- PyMOL | pymol.org. Available at: <https://pymol.org/2/> (Accessed January 8, 2022).
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033
- Sanjuán, R., and Domingo-Calap, P. (2016). Mechanisms of Viral Mutation. *Cell. Mol. Life Sci.* 73, 4433–4448. doi: 10.1007/s00018-016-2299-6
- Selvaraj, C., Sakkiah, S., Tong, W., and Hong, H. (2018). Molecular Dynamics Simulations and Applications in Computational Toxicology and Nanotoxicology. *Food Chem. Toxicol.* 112, 495–506. doi: 10.1016/j.fct.2017.08.028
- SeyedAlinaghi, S., Mirzapour, P., Dadras, O., Pashaei, Z., Karimi, A., MohseniPour, M., et al. (2021). Characterization of SARS-CoV-2 Different Variants and Related Morbidity and Mortality: A Systematic Review. *Eur. J. Med. Res.* 26, 51. doi: 10.1186/s40001-021-00524-8
- Shafreen, R. M. B., Selvaraj, C., Singh, S. K., and Pandian, S. K. (2013). Exploration of Fluoroquinolone Resistance in Streptococcus Pyogenes: Comparative Structure Analysis of Wild-Type and Mutant DNA Gyrase. *J. Mol. Recognit.* 26, 276–285. doi: 10.1002/jmr.2270
- Song, S., Ma, L., Zou, D., Tian, D., Li, C., Zhu, J., et al. (2020). The Global Landscape of SARS-CoV-2 Genomes, Variants, and Haplotypes in 2019ncovr. *Genomics Proteomics Bioinf.* 18, 749–759. doi: 10.1016/j.gpb.2020.09.001
- Toyoshima, Y., Nemoto, K., Matsumoto, S., Nakamura, Y., and Kiyotani, K. (2020). SARS-CoV-2 Genomic Variations Associated With Mortality Rate of COVID-19. *J. Hum. Genet.* 65, 1075–1082. doi: 10.1038/s10038-020-0808-9
- Voss, J. D., Skarzynski, M., McAuley, E. M., Maier, E. J., Gibbons, T., Fries, A. C., et al. (2020). Variants in SARS-CoV-2 Associated With Mild or Severe Outcome. *medRxiv.* 9, 267–275. doi: 10.1101/2020.12.01.20242149
- Wang, Y., Lu, X., Li, Y., Chen, H., Chen, T., Su, N., et al. (2020). Clinical Course and Outcomes of 344 Intensive Care Patients With COVID-19. *Am. J. Respir. Crit. Care Med.* 201, 1430–1434. doi: 10.1164/rccm.202003-0736LE
- Wohlert, J., and Edholm, O. (2004). The Range and Shielding of Dipole-Dipole Interactions in Phospholipid Bilayers. *Biophys. J.* 87, 2433–2445. doi: 10.1529/biophysj.104.044222
- Yang, H.-C., Chen, C.-H., Wang, J.-H., Liao, H.-C., Yang, C.-T., Chen, C.-W., et al. (2020a). Analysis of Genomic Distributions of SARS-CoV-2 Reveals a Dominant Strain Type With Strong Allelic Associations. *Proc. Natl. Acad. Sci. U.S.A.* 117, 30679–30686. doi: 10.1073/pnas.2007840117
- Yang, J., Petitjean, S. J. L., Koehler, M., Zhang, Q., Dumitru, A. C., Chen, W., et al. (2020b). Molecular Interaction and Inhibition of SARS-CoV-2 Binding to the ACE2 Receptor. *Nat. Commun.* 11, 4541. doi: 10.1038/s41467-020-18319-6
- Zeng, W., Liu, G., Ma, H., Zhao, D., Yang, Y., Liu, M., et al. (2020). Biochemical Characterization of SARS-CoV-2 Nucleocapsid Protein. *Biochem. Biophys. Res. Commun.* 527, 618–623. doi: 10.1016/j.bbrc.2020.04.136
- Zhang, X., Tan, Y., Ling, Y., Lu, G., Liu, F., Yi, Z., et al. (2020). Viral and Host Factors Related to the Clinical Outcome of COVID-19. *Nature* 583, 437–440. doi: 10.1038/s41586-020-2355-0

- Zhao, H., Lu, L., Peng, Z., Chen, L.-L., Meng, X., Zhang, C., et al. (2021). SARS-CoV-2 Omicron Variant Shows Less Efficient Replication and Fusion Activity When Compared With Delta Variant in TMPRSS2-Expressed Cells. *Emerg. Microbes Infect.* 11, 1–18. doi: 10.1080/22221751.2021.2023329
- Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., et al. (2020). Clinical Course and Risk Factors for Mortality of Adult Inpatients With COVID-19 in Wuhan, China: A Retrospective Cohort Study. *Lancet* 395, 1054–1062. doi: 10.1016/S0140-6736(20)30566-3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Maurya, Mishra, Swaminathan, Ravi, Saiji, Kankan, Mehta, Devi, Praveen, Budhiraja, Tarai, Sharma, Khyalappa, Joshi and Pandey. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.