International Journal of Advanced Research and Development
www.advancedjournal.com
ISSN: 2455-4030
Received: 15-07-2021, Accepted: 30-07-2021, Published: 15-08-2021
Volume 6, Issue 4, 2021, Page No. 14-21

Genomic sequence of worldwide strains of SARS-CoV-2: Insights the role of variants in disease epidemiology

Sudheer Menon
Department of Bioinformatics, Bharathiar University, Coimbatore, Tamil Nadu, India

Abstract
In humans, the SARS-CoV-2 infection has achieved endemicity. This needs the creation of quick genetic monitoring technologies to supplement existing comprehensive, albeit slower, genome sequencing-based approaches. The study's main goal was to look into the genomic sequencing of SARS CoV-2 in order to learn more about its potential in global disease epidemiology. SARS-CoV-2 whole genome sequencing was performed on remnant samples from nasopharyngeal and oropharyngeal swabs taken from symptomatic patients. The produced sample was subjected to real-time PCR, and the resulting sequences were then symmetrized to create a dataset. This data was subsequently integrated into global datasets available on several websites, including the GISAID and ViPR. It was also intended to investigate the distribution pattern of distinct SARS-CoV-2 strains. Epidemiological study and the method of estimating reproductive numbers were used to achieve this goal. The findings revealed that genome sequencing is closely linked to the disease epidemiology of distinct SARS-CoV-2 strains. These strains tend to decrease with the passage of time and their distribution was disturbed geographically due to the invention of vaccines.

Keywords: genomic sequence, worldwide strains, SARS-CoV-2

Introduction

The SARS-CoV-2 pandemic most likely began in late 2019 in China, and by January 2021, there had been over 100 million confirmed cases and over two million deaths worldwide owing to COVID-19. The coronavirus disease of 2019 (COVID-2019) began as an outbreak in late 2019 in Wuhan, China, and shocked the world, prompting the World Health Organization (WHO) to designate it a worldwide pandemic in March 2020 (Zhu *et al*., 2020) [1]. According to WHO figures, the number of people infected with COVID-19 has topped 187 million, with a global mortality rate of 4.04 million recorded around July 14, 2021. If recent events are any indication of earlier epidemics, the rapid spread of this viral disease indicates that the antidote to this pandemic is devising techniques to overcome and reduce the disease rather than self-isolation (Rao *et al*., 2020) [2].

According to WHO figures, the number of people infected with COVID-19 has topped 187 million, with a global mortality rate of 4.04 million recorded around July 14, 2021. If recent events are any indication of earlier epidemics, the rapid spread of this viral disease indicates that the antidote to this pandemic is devising techniques to overcome and reduce the rather than self-isolation. Vaccines do not reduce or eliminate the virus; rather, they may increase the immune response to a certain viral serotype that may be ineffective against another serotype (Sariol *et al*., 2014) [3]. As a result, global scientists are tasked with devising an alternate technique that directly tackles the viral particle's function. In order to pursue this strategy, either inhibitors of SARS-CoV-2 structural proteins and enzymes must be produced (Zhou *et al*., 2020) [4], or the viral genome's functional integrity must be targeted with micro RNAs (miRNAs). SARS-CoV-2 is a single-stranded positive-sense RNA virus, according to the International Committee on Virus Taxonomy (ICTV).

Coronavirus is a single-stranded positive-sense RNA virus belonging to the Coronaviridae family, with a genomic size of 26–32 kb and 14 open reading frames (ORFs) encoding 27 polyproteins (Rao *et al*., 2020) [2]. The first ORF encodes 16 non-structural proteins (nsps), while the remaining ORFs encode accessory and structural proteins such as the spike (S) glycoprotein, envelope (E), membrane (M), and nucleocapsid (N), which are the focal target proteins for drug repurposing and therapeutics studies (Rao *et al*., 2021) [10].

Andrews and Gledhill discovered a single-stranded RNA coronavirus after screening a hepatitis virus from mice in 1951. For more than 50 years, its diseases and causes have been identified in both animals and people. The earliest coronaviruses to produce mild illnesses in humans, such as common colds, were 229E and OC43 (Isaacs *et al*., 1951) [32]. Civet cats, camels, and bats later developed severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS). Coronaviridae is the coronavirus family, which is divided into two subfamilies: Coronaviridae and Letovirinae. The initial diagnosis of pneumonia and the unknown aetiology caused a cluster of patients in China to be brought to hospitals in December 2019. (Vassilara *et al*., 2020). On February 11, 2020, WHO named the possible coronavirus outbreak coronavirus disease 19 (COVID-19) after the reports confirmed it.

Before the 1960s, coronaviruses were first found and defined, and they were thought to cause a respiratory ailment (Khan and McIntosh, 2005; Paules *et al*., 2020) [11, 16]. SARS-CoV-2 has spread among humans, wreaking havoc on health-care systems and economies. Given the higher mutation rate of RNA viruses compared to DNA viruses, it was thought that the viral genome of SARS-CoV-

2 would evolve more quickly, allowing for better tracking of the virus's dissemination (Grubaugh *et al*., 2019) [17].

The common cold, bronchiolitis, and pneumonia are all caused by CoVs, which are a broad and diverse family of viruses that cause a variety of respiratory, gastrointestinal, and neurologic disorders of various severity (Weiss and Leibowitz, 2011) [25]. Human CoV are now limited to the alpha (HCoV-229E and HCoV-NL63) and beta genera (HCoV-OC43, HCoV-HKU1), the latter of which includes SARS-CoV1 and the Middle East respiratory syndrome coronavirus (MERS-CoV). SARS-CoV2, a previously unknown coronavirus, was found in Wuhan (Hubei Province, China) in December 2019 and sequenced in January 2020 (Lu *et al*., 2020) [26].

Despite their species diversity, CoVs share crucial genetic components that are required for viral replication, implying that broad-spectrum therapeutic medicines could be developed to combat the current pandemic and prevent future outbreaks. The highly conserved RdRp plays a critical function in the CoV replication cycle, accelerating the production of fresh viral RNA, and is the goal of our research to find novel inhibitors (Velthuis *et al*., 2012) [31].

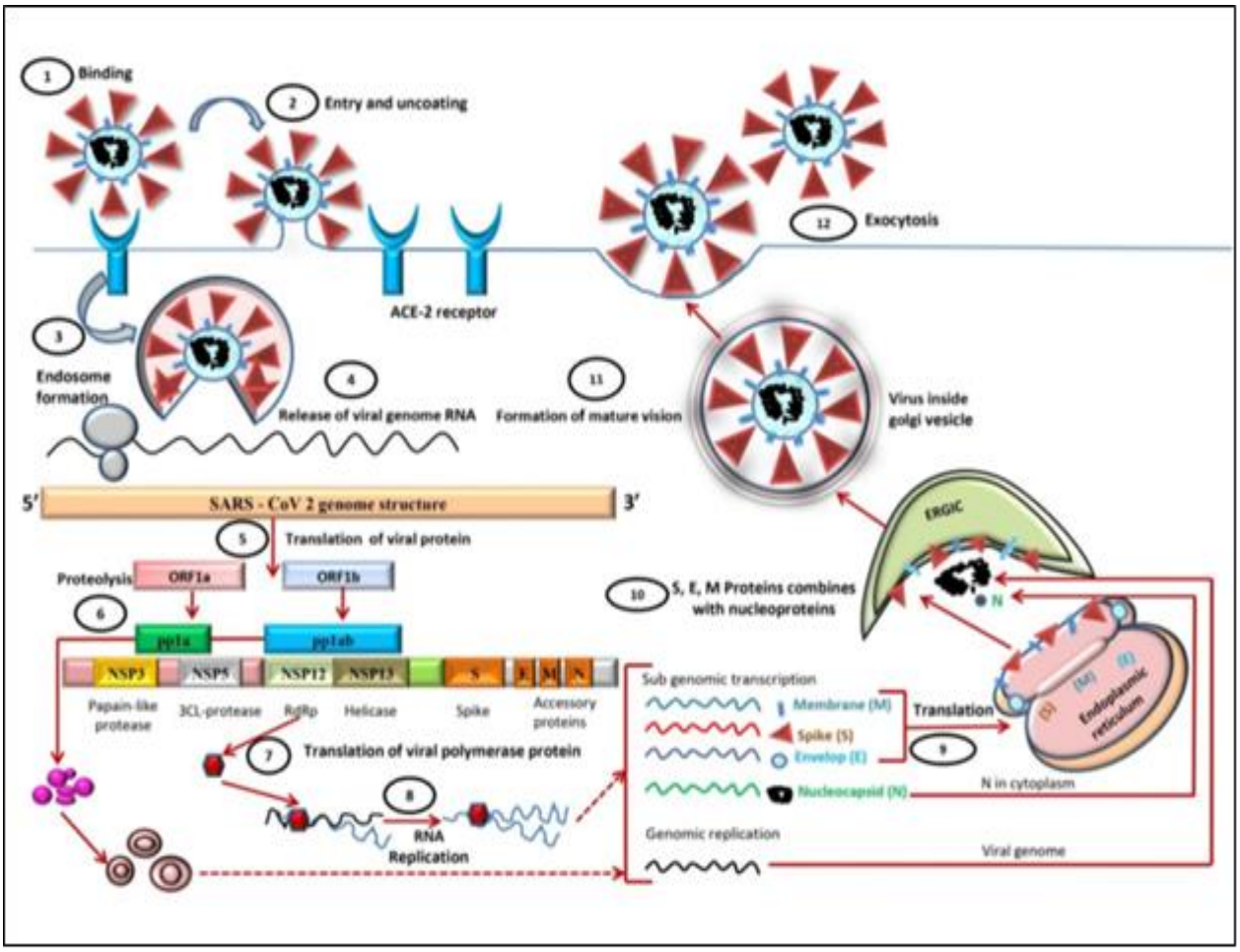

Fig 1: Life cycle of SARS CoV-2 in host cell

Modeling has been critical in understanding evolving COVID-19 epidemiology and guiding public health responses (Yang *et al*., 2020) [26], particularly with respect to vaccines and Variants of Concern, in the ever-changing epidemic landscape of COVID-19 (the disease caused by severe acute respiratory syndrome coronavirus-2; SARS-CoV-2) (VoC; viral strains hosting mutations that increase transmissibility, change immune response, or affect disease severity). Multiple effective vaccines have become available in the eighteen months since COVID-19 appeared, but outbreaks continue, in part due to poor vaccine access and developing VoC. (Chowdhury *et al*., 2020) [23].

Modeling has identified important epidemiological characteristics and helped identify when and what type of restrictions should be applied to gain epidemic control since the beginning of the pandemic. Models have explicitly recorded social interactions to estimate the effects of reopening schools, increasing the number of on-site workers (Karatayev *et al*., 2020) [24], and resuming international travel, for example.

The cause of the ailment was unknown at first, but because of its quick spread, it was suspected of being an infectious agent. A viral agent belonging to the CoV group but with completely distinct genetic traits than previous CoVs was identified from infected individuals, according to findings released by the US Centers for Disease Control and Prevention (CDC) and the World Health Organization (WHO) (Arguin *et al*., 2004) [39]. The outbreak was finally brought under control by the end of 2003. (Lau *et al*., 2004) [40]. Given that the disease is a newly emerging pandemic and is likely to recur, understanding suspected clinical and laboratory symptoms, the treatment process, its transmission, some strategies for controlling the infection, and essential recommendations before traveling to and from infected areas were shown to be necessary.

Since the outbreak began, a tremendous amount of

sequencing data has been collected, and this data is shared on the GISAID database, which is a valuable resource for virus surveillance. Based on spatial distribution, three general groupings, G, V, and S, were defined from these data. The G group has a competitive advantage due to a substitution in the spike protein (D614G), which allowed the G group to overpower the other viruses in less than two months, albeit it did not entirely displace the other viruses, which are still circulating in various parts of the world.

One theory is that the G viruses are More fit and have a higher transmissibility rate, however this is currently being researched. G, GR, and GH are now the three subgroups of the G group. Another group of substitutions in regions near to the receptor binding site, of which the S477 is especially noteworthy (S477I/S477 N), has recently been identified at the protein level that interacts with the ACE2 receptor, with a probable improved capacity to attach to the receptor (Salje *et al*., 2020) [42].

Looking at the overall history of the virus based on SARS-CoV-2 genomic monitoring in France, virus sequences from the early stages of the epidemic, March to June, show virtually little evolution. This is due to the fact that the receptor binding site does not tolerate many sequence variations. Changes in these viruses occur at a rate of two to three nucleotides each month at most. Furthermore, there are just a few hundred single nucleotide polymorphism (SNP) locations that can be modified, indicating that the virus is quite stable and thus not comparable to influenza (Cauchemez *et al*., 2020) [43].

As of August 11, GISAID has more than 52,600 full and high-coverage genomes. Systematically tracking alterations in SARS-CoV-2 genomes is crucial because it enables for national and worldwide surveillance of the molecular epidemiology of circulating viral sequences. The nucleotide variation landscape of a large group of internationally generated SARS-CoV-2 genomes was studied and important mutation events were identified. As a result of this research, we were able to create a first-generation genetic classifier, or 'barcode,' that defined the major clades of the virus that were circulating up to May 28, 2020. This barcode, in particular, made it possible to trace the spatial distribution and prevalence of these viral clades across time. While most non-synonymous mutations appear to have little effect on protein function or stability, we did find indications of changes in the spike protein that could modify SARS-CoV-2-human host interactions.

The spike (S) protein, the envelope (E) protein, the membrane (M) protein, and the nucleocapsid (N) protein are the four key structural proteins in SARS-CoV-2 (Chen *et al*., 2020). Furthermore, there are 16 non-structural proteins (NSPs) that could be used in medication development. S protein is made up of two parts: S1 and S2. The receptor-binding domain (RBD) of S protein is involved in viral entrance and binding to the host receptor protein ACE2, while S2 is involved in viral fusion (Hatmal *et al*., 2020) [45]. The M and E proteins play a key role in viral assembly. N protein is an RNA-binding multivalent protein that is required for viral replication and genome packaging. Because mutations in structural and non-structural proteins are so important in virulence and virus propagation, they should be taken into account while developing treatments and vaccines (Premkumar *et al*., 2020) [46]. SARS-CoV-2 contains major genetic alterations that are completely unknown to the human innate immune system (Tomaszewski *et al*., 2020) [47]. On comparing the epidemiology and pathobiology of SARS-CoV-2 with SARS and MERS, Ganesh *et al*. collated findings that can be used while treatments are being developed (Ganesh 2021) [48].

AIMS

In humans, the SARS-CoV-2 infection has achieved endemicity. This needs the development of quick genetic monitoring technologies to supplement existing comprehensive, genome-based approaches. Peptide research has transformed the discovery of new treatments for a variety of infectious diseases. The continued possibility of a COVID-19 pandemic necessitates the development of potential therapeutic alternatives. Due to the possibility of repeated infection by existing strains or pandemic outbreaks by novel mutant strains, intensive and fruitful research is being conducted to produce broad spectrum vaccines and treatment alternatives for corona viruses. Specific aim of the research was to study the genome sequencing of SARS CoV-2 and to study the insights of the role played by variants in disease epidemiology all over the world.

Materials and Methods

1. SARS-CoV-2 sample collection and preparation

SARS-CoV-2 whole genome sequencing was performed on remnant samples from nasopharyngeal and oropharyngeal swabs taken from symptomatic patients. The primary swab sample or extracted RNA were included in these samples. Prior to RNA extraction, the swab samples were heat inactivated in a water bath at 60 oC for 30 minutes in a biosafety level 3 facility. The Viral NA/gDNA Kit on the Chemagic 360 system (Perkin Elmer, Hamburg, Germany), the automated Chemagic 360 equipment (Perkin Elmer, Hamburg, Germany), or the Qiagen Viral RNA Mini Kit were used to extract RNA (QIAGEN, California, USA).

2. Real Time PCR (RT-PCR)

The TaqPath COVID-19 CE-IVD RT-PCR Kit (Life Technologies, Carlsbad, CA) was used to detect the SARS-CoV-2 virus by PCR, according to the manufacturer's instructions. The assays target genomic sections of the SARS-CoV-2 genome (ORF1ab, S protein, and N protein). A Quant Studio 7 Flex Real-Time PCR apparatus was used for RT-PCR (Life Technologies, Carlsbad, CA). Using auto-analysis settings, cycle thresholds (Ct) were evaluated, with the threshold lines occurring inside the exponential phase of the fluorescence curves and above any background signal. A Ct value for RNAse P (an endogenous internal amplification control) and/or the target gene in each reaction is required to accept the results a Ct value for RNAse P (i.e. an endogenous internal amplification control) and/or the target gene in each reaction was confirmed, with undetermined Ct values in the no template control.

3. Illumina Mi-Seq sequencing

This sequencing platform incorporated PCR products for samples that yielded enough material. To generate uniquely indexed paired end libraries of genomic DNA, Illumina1 TruSeq1 Nano DNA Library Prep kits were used according to the manufacturer's methodology. The fragments were examined using the Lab Chip GX Touc, and the libraries were quantified using the Qubit dsDNA high-sensitivity assay on the Qubit 4.0 instrument (Life Technologies)

(Perkin Elmer, Hamburg, Germany). The sequencing libraries were pooled and denatured with 0.2 N sodium acetate after being normalized to 4 nM. 1 percent PhiX was added to the 12 pM sample libraries (PhiX Control v3 adapter-ligated library used as a control). Libraries consisting of 12 samples each were loaded onto a 500-cycle MiSeq Nano Reagent Kit v2 nano v2 Miseq reagent kit and run on the Illumina MiSeq instrument (Illumina, San Diego, CA, USA).

4. Preparing world datasets

On June 30, 2020, the entire SARS-CoV2 genome sequences were downloaded from the GISAID website, and any sequences with lengthy internal gaps or ambiguities (>30bps) were eliminated. There were 46,612 sequences in the final collection. To construct separate collections of samples, the following processing was done on these sequences: Collecting Next strain sequences from subsamples: The Next strain team (subsamples the massive viral dataset into smaller collections of sequences covering different geographical locations to make the phylogenetic tree easier to visualize. There are approximately 5000 sequences in the most recent version (October 2020). We took a sub-sample of this list and got roughly 250 sequences.

5. Phylogenomic analysis

On July 14, 2021, 1427 coronavirus genomes, including 329 SARS, 35 SARS-CoV-2, 61 HCoV-NL63, 521 MERS, 52 HCoV-HKU1,170 HCoV OC43, 97 bovine coronaviruses, and 61 mouse hepatitis viruses, were retrieved from the Virus Pathogen Database and Analysis Resource (ViPR). MAFFT (version 7.407) (Katoh and Standley 2013) software was used to align the sequences, and trimAL (version 1.4.1) (Capella- Gutiérrez *et al*., 2009) was used to trim them. The full genomes were phylogenetically analysed using FastTree (version 2.1.10) software with default parameters, and phylogenetic tree visualisation was done using iTOL (version 5) software. GISAID was used to obtain 21,789 full SARS-CoV2 genomes (Shu, 2020).

6. Protein structure analysis

The Protein Data Bank provided experimentally determined protein structures (PDB). To create protein structural homology models, researchers utilised Swiss Model (Arnold *et al*. 2006), I Tasser (Yang *et al*. 2014) [56], Raptor X, and an in-house modelling workflow. The transmembrane regions were predicted using Phobius. Raptor X was also utilised for secondary structure prediction and visualisation.

7. Epidemiological analysis and reproductive number estimation

The effective reproduction number (R) was calculated using the observed epidemic growth rate r as well as two hypothesised R relationships previously published in the literature. The current investigation utilised the relationship R = (1+r/b) given in Imperial College London's COVID-19 report 13, where a = m2/s2 and b = m/s2, m the serial interval (SID) mean and the SID standard deviation, and m the serial interval (SID) standard deviation. With m = 4.7 and s = 2.9, the SID distribution utilised was that calculated by (Nishiura *et al*., 2020). The Flaxman *et al*. approach was given to this method (Flaxman *et al*., 2020), (ii) The relationship R = (1+r/sigma) (1+r/delta) was used, with 1/sigma representing the infectious period and 1/delta representing the incubation period, as described by Wallinga and Lipsitch (2007), which is based on an SEIR modelling framework and expects both periods to be exponentially distributed. Exponential distributions were employed, with a mean incubation time of 5.1 days and an infection time of 4 days. The Wallinga *et al*. technique was named after this method. Maximum likelihood estimation in R (function optim) was used to calculate the epidemic growth rate r, which was calculated by fitting the exponential growth model A0ert to the reported time series of cases and deaths (independently), where t is time, A0 is the number of reports at t = 0, and r is the growth rate. Deaths and cases that were reported on a daily basis were utilised. The data for deaths was collected from 27 March to 17 July, and the data for cases was collected from 5 March to 12 July.

Results

The SARS-CoV-2 genome is most closely linked to SARS-related coronaviruses (SARS-CoV-2) from horseshoe bats, with 96 percent similarity to RaTG13, 93 percent similarity to RmYN02, and 88 percent similarity to bat-SL-CoVZC45 and bat-SL-CoVZXC21, as well as SARS-CoV-2 from pangolins. We looked at the temporal occurrences of SARS-CoV-2 genomes throughout the early phases of the pandemic and the global spread of SARS-CoV-2 across human populations, which helped us identify multiple viral clades with shared SNPs. Real-time PCR, phylogenomic analysis, and protein analysis revealed that the newly found SARS CoV-2 variants were deadly and resulted in a large number of deaths throughout the world.

1. Clinical data

Fever 43/61 (70%), cough 14/61 (23%), asthma 9/61 (14.7%), myalgia or tiredness 30/61 (49%), nasal congestion 16/61 (26%), sputum production 4/61 (5.7%), and dyspnea 14/61 (14%) were the most common clinical symptoms in COVID-19 patients (23 percent). Headache or dizziness 3/61 (5%), diarrhea 3/61 (5%), nausea and vomiting 4/61 are all minor symptoms (5.7 percent). All of the patients were able to recover without any problems.

2. Data availability

The nucleotide and amino acid percentage differences across SARS-CoV-2 cases are 0.4 percent and 0.25 percent, respectively, according to an analysis of all SARS-CoV-2 genome sequences in positive patients from around the world. The samples from all across the globe were given the prefix "hCoV-19/CUNCI-HGC" and the sample numbers were appended to it. They were sent to GISAID and GenBank. In addition, GISAID has been used to store all sequences. Supplementary S1 contains a list of all sequences in all databases, as well as their IDs. SARS-CoV2 genome genes and mutations. Based on the reference Wuhan 1 sequence NC 045512.2, all obtained sequences were aligned and trimmed.

3. Genome Sequencing

All sequences were cut to 29698 bp, and the global strains revealed a total of 204 mutations. We discovered that the ORF1ab polyprotein was responsible for more than half of the variants (64 percent). The ORF6 and E protein sequences have the least number of variants (0.5 percent) (Table 1). There were 131 mutations in ORF1ab, followed

by 30 mutations in S, 23 mutations in N, 6 mutations in ORF3a, 6 mutations in ORF7a, 4 mutations in ORF8, 2 mutations in M, 1 mutation in E, and 1 mutation in ORF6. ORF1ab is split into 16 non-structural proteins after being transcribed as a multi-protein (NSPs). NSP3 contains 34 ORF1ab protein variants, including 20 missense, 13 synonymous, and one frameshift. c.2772C > T (p. Phe924Phe) was found in 57 of the 61 samples, followed by c.5019C > T (p. Asn1673Asn) in 6/61 samples, and c.2772delC (p. Tyr925fs) in 3/61 samples of the NSP3.

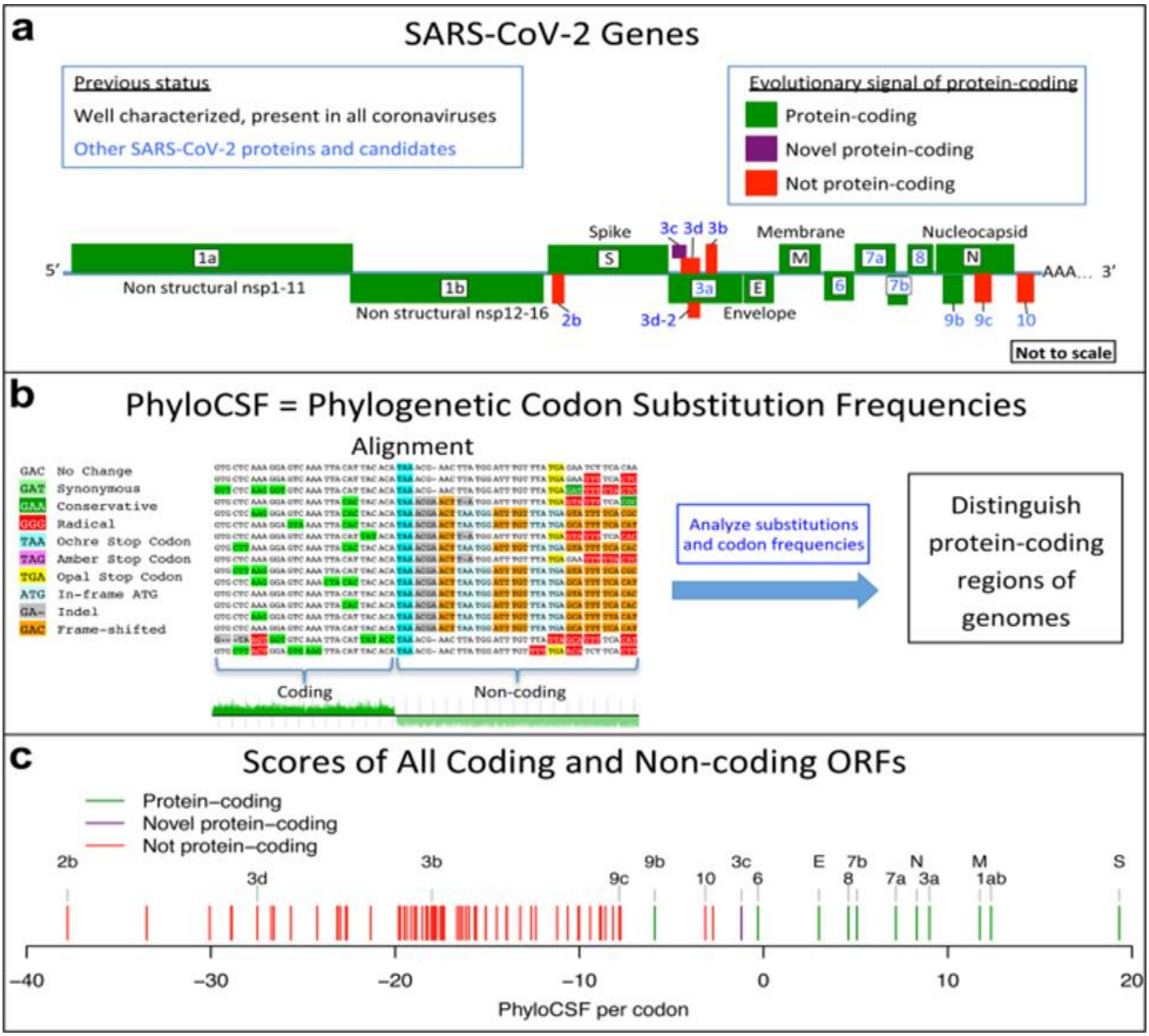

Fig 2: Genomic Sequence of SARS CoV-2

4. Worldwide Frequency of Variants:

Figure (A) depicts the distribution and variance of various variations throughout the genome. Because of its vastness, it's no surprise that the ORF section has the majority of the variegations (148). We compared this pattern of distribution across a variety of datasets from throughout the world. This contrast is shown in Figure (B). The relative frequencies among the genes in the same dataset are also shown in the graph. Figure 3(b) shows that from June to October 2020, the number of variants worldwide grew by a maximum of 24%, which is not proportionate to the doubling of deposited sequences. In addition, the world's relative frequencies in June did not differ from those in October (P-value > 0.05; Chi-square test). This means that particular genes aren't under any more stress than others. In June and October, there was no significant difference in the Relative Frequencies columns between the Egyptian and global populations (P value > 0.05; Chi square).

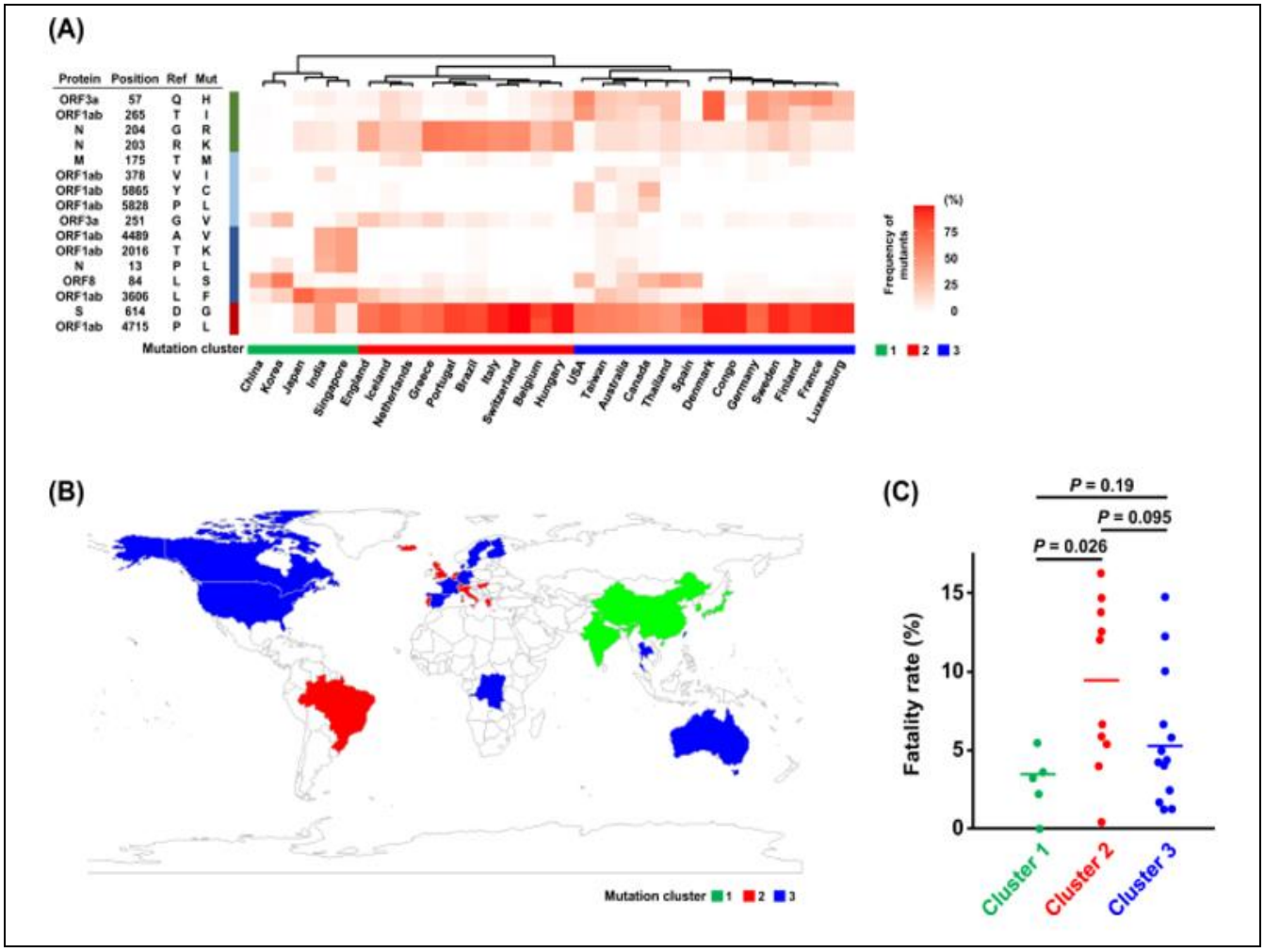

Fig 3: Worldwide frequency and genomic sequencing of SARS CoV-2

5. Epidemiological Data

The epidemiological data of global COVID-19 cases has been considerably disclosed thanks to genome sequencing of novel SARS CoV-2 variants. COVID-19 is impacting 221 nations and territories, according to the WHO. The overall number of cases reported worldwide is 219,332,901, the total population lost to death is 4,545,878 people, and the total number of recovered patients is 196,119,311. (WHO, September 2, 2021).

Discussion

The goal of this study was to look at the genomic sequences of several strains of SARS CoV-2 from all around the world. This sequence data showed the disease's epidemiology and gave researchers a better understanding of the link between sequencing patterns and epidemiology. This study identified five significant variations (G614, S84, V251, I378, and D392) of completely sequenced SARS-CoV-2 genomes in the GISAID database, which will be available until July 28, 2021. The distribution of variations across different geographical regions was shown by clustering these genomes. This trend differs from those of other epidemic coronaviruses, such as MERS-CoV, which have different geographic areas (Kim *et al*. 2016).

For example, clade G614, which was discovered on January 28, is currently prevalent in sampled genomes all around the world, and its proportion is rising from Europe to North America, Oceania, and finally Asia. This pattern indicates effective viral transmission by frequent transcontinental travel during a time when international travel restrictions were infrequent, allowing the virus to spread to numerous distant sites in a short amount of time. This finding emphasises the significance of limiting foreign travel and establishing limitations early in pandemics, as well as enforcing social distance, to keep viruses from spreading globally.

In different regions of the world, we've seen a varied distribution of the primary variations (Figure 3.2). The majority of viral genomes that haven't been attributed to a major variation have been discovered in Asia, with discovery dates in January and February, around the onset of the pandemic in China. The study's findings indicated that the global spread of viral strains has been decreasing over time.

Conclusions

This work establishes a baseline for SARS CoV-2 genome sequencing and elucidates the importance of this data in revealing the epidemiological characteristics of various strains throughout the world. It will be beneficial for tracking changes in SARS-CoV-2 circulating variants in different geographical locations over time. The epidemiological data will be utilized to develop efficient disease control methods across the world.

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J *et al*. A novel coronavirus from patients with pneumonia in

China, 2019, N. Engl. J. Med,2020:382:727–733,
2. Rao P, Shukla A, Parmar P, Rawal RM, Patel BV, Saraf M *et al*. Proposing a fungal metabolite-flaviolin as a potential inhibitor of 3CLpro of novel coronavirus SARS-CoV-2 identified using docking and molecular dynamics, J. Biomol. Struct. Dyn, 2020, 1-13,
3. Sariol CA, White LJ. Utility, limitations, and future of non-human primates for dengue research and vaccine development, Front. Immunol,2014: 5:1–15.
4. Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2, Cell Discov 6, 2020.
5. Sudheer Menon. "Preparation and computational analysis of Bisulphite sequencing in Germfree Mice" International Journal for Science and Advance Research In Technology,2020:6(9):557-565.
6. Sudheer Menon, Shanmughavel Piramanayakam, Gopal Agarwal. "Computational identification of promoter regions in prokaryotes and Eukaryotes" EPRA International Journal of Agriculture and Rural Economic Research (ARER),2021:9(7):21-28.
7. Sudheer Menon. "Bioinformatics approaches to understand gene looping in human genome" EPRA International Journal of Research & Development (IJRD),2021:6(7):170-173).
8. Sudheer Menon. "Insilico analysis of terpenoids in Saccharomyces Cerevisiae" international Journal of Engineering Applied Sciences and Technology, ISSN No. 2455-2143, 2021:6(1):43-52.
9. Rao P, Shukla A, Parmar P, Rawal RM, Patel B, Saraf D *et al*. Reckoning a fungal metabolite, Pyranonigrin A as a potential Main protease (Mpro) inhibitor of novel SARS-CoV-2 virus identified using docking and molecular dynamics simulation, Biophys. Chem, 2020:264:106425.
10. Rao P, Patel R, Shukla A, Parmar P, Rawal RM, Saraf M *et al*. Identifying structural–functional analogue of GRL0617, the only well-established inhibitor for papain-like protease (PLpro) of SARS-CoV2 from the pool of fungal metabolites using docking and molecular dynamics simulation, Mol. Divers, 2021.
11. Khan JS, McIntosh K. History and recent advances in coronavirus discovery. Pediatr Infect Dis J,2005:24(11):S223–227.
12. Sudheer Menon. "Computational analysis of Histone modification and TFBs that mediates gene looping" Bioinformatics, Pharmaceutical, and Chemical Sciences (RJLBPCS), June,2021:7(3):53-70.
13. Sudheer Menon Shanmughavel piramanayakam, Gopal Prasad Agarwal (2021) "FPMD-Fungal promoter motif database: A database for the Promoter motifs regions in fungal genomes" EPRA International Journal of Multidisciplinary research,2021:7(7):620-623.
14. Sudheer Menon, Shanmughavel Piramanayakam, Gopal Agarwal. Computational Identification of promoter regions in fungal genomes, International Journal of Advance Research, Ideas and Innovations in Technology,2021:7(4):908-914.
15. Sudheer Menon, Vincent Chi Hang Lui and Paul Kwong Hang Tam. Bioinformatics methods for identifying hirschsprung disease genes, International Journal for Research in Applied Science & Engineering Technology (IJRASET), 2021:9(7):2974-2978.
16. Paules CI, Marston HD, Fauci AS. Coronavirus infections—more than just the common cold. JAMA,2020:323(8):707–8.
17. Grubaugh ND, Ladner JT, Lemey P, Pybus OG, Rambaut A, Holmes EC, *et al*. Tracking virus outbreaks in the twenty-first century. Nat Microbiol,2019:4(1):10–9.
18. Yang S, Cao P, Du P, Wu Z, Zhuang Z, Yang L, *et al*. Early estimation of the case fatality rate of COVID-19 in mainland China: a data-driven analysis. Ann Transl Med,2020:8(4):128-128. Available from: https://pmc/articles/PMC7049028/.
19. Sudheer Menon. Bioinformatics approaches to understand the role of African genetic diversity in disease, International Journal Of Multidisciplinary Research In Science, Engineering and Technology (IJMRSET),2021:4(8):1707-1713.
20. Sudheer Menon. Comparison of High-Throughput Next generation sequencing data processing pipelines, International Research Journal of Modernization in Engineering Technology and Science (IRJMETS),2021:3(8):125-136.
21. Sudheer Menon. Evolutionary analysis of SARS-CoV-2 genome and protein insights the origin of the virus, Wuhan, International Journal of Creative Research Thoughts (IJCRT),2021:9(8):b696-b704.
22. Sudheer Menon, Vincent Chi Hang Lui and Paul Kwong Hang Tam. A step-by-step work flow of Single Cell RNA sequencing data analysis, International Journal for Scientific Research and Development (IJSRD),2021:9(6):1-13.
23. Chowdhury R, Heng K, Shawon MSR, Goh G, Okonofua D, Ochoa-Rosales C, *et al*. Dynamic interventions to control COVID-19 pandemic: a multivariate prediction modelling study comparing 16 worldwide countries. Eur J Epidemiol,2020:35(5):389–99.
24. Karatayev VA, Anand M, Bauch CT. Local lockdowns outperform global lockdown on the far side of the COVID-19 epidemic curve. Proc Natl Acad Sci U S A,2020:117(39):24575–80.
25. Weiss SR, Leibowitz JL. Coronavirus pathogenesis. Adv. Virus Res,2011:81:85–164.
26. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H *et al*. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, 2020.
27. Sudheer Menon. Computational characterization of Transcription End sites in Human Genome, International Journal of All Research Education and Scientific Methods (IJRESM),2021:9(8):1043-1048.
28. Sudheer Sivasankaran Menon, Shanmughavel Piramanayakam. Insilico prediction of gyr A and gyr B in *Escherichia coli* insights the DNA-Protein interaction in prokaryotes, International Journal of Multidisciplinary Research and Growth Evaluation, (IJMRD),2021:2(4):709-714.
29. Sudheer Menon, Vincent Chi Hang Lui, Paul Kwong Hang Tam. Bioinformatics tools and methods to analyze single cell RNA sequencing data, International Journal of Innovative Science and Research Technology, (IJISRT),2021:6(8):282-288.
30. Sudheer Menon. Computational genome analysis for identifying Biliary Atresia genes, International Journal of Biotechnology and Microbiology,

(IJBM),2021:3(2):29-33.
31. Te Velthuis AJW, Van Den Worm SHE, Snijder EJ. The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension. Nucleic Acids Res,2012:40(4):1737–1747.
32. Isaacs A, Gledhill AW, Andrewes CH. Influenza A viruses; laboratory studies,with special reference to European outbreak of 1950-1. Bull World Health Organ,1952:6:287–315.
33. Vassilara F, Spyridaki A, Pothitos G, Deliveliotou A, Papadopoulos A. A rare caseof human coronavirus 229E associated with acute respiratory distress syndrome in a healthy adult. Case Rep Infect Dis,2018,1-4,
34. World Health Organization. Coronavirus press conference 11 February, 2020Speaker. J Chem Inf Model,2019:53:1689-99.
35. Sudheer Menon. Recent Insilco advancements in genome analysis and characteristics of SARS-Cov2. International Journal of Biology Research, (IJBR),2021:6(3):50-54.
36. Sudheer Menon. Bioinformatics methods for identifying Human disease genes, International Journal of Biology Sciences, (IJBR),2021:3(2):1-5.
37. Sudheer Menon. SARS-CoV-2 Genome structure and protein interaction map, insights to drug discovery, International Journal of Recent Scientific Research, (IJRSR),2021:12(8):42659-42665.
38. Sudheer Menon. Insilico Insights to Mutational and Evolutionary aspects of SARS-Cov2, International Journal of Multidisciplinary Research and Development, (IJMRD),2021:8(8):167-172.
39. Arguin PM, Navin AW, Steele SF, Weld LH, Kozarsky PE. Health communication during SARS, Emerging infectious diseases,2004:10(2):377.
40. YL Lau. SARS: future research and vaccine, Pediatric respiratory reviews,2004:5(4):300-303.
41. Bauch CT, Lloyd-Smith JO, Coffee MP, Galvani AP. Dynamically modeling. SARS and other newly emerging respiratory illnesses: past, present, and future, Epidemiology, 2005, 791–801.
42. Salje H, Tran Kiem C, Lefrancq N, Courtejoie N, Bosetti P, Paireau J, *et al*. Estimating the burden of SARS-CoV-2 in France. Science (New York, NY),2020:369:208–11.
43. Cauchemez S, Kiem CT, Paireau J, Rolland P, Fontanet A. Lockdown impact on COVID-19 epidemics in regions across metropolitan France. Lancet,2020:396: 1068–9.
44. Chen L, Zhong L. Genomics functional analysis and drug screening of SARS-CoV-2. Genes Dis,2020:7(4):542–50,
45. Hatmal MM, Alshaer W, Al-Hatamleh MAI, Hatmal M, Smadi O, Taha MO, *et al*. Comprehensive structural and molecular comparison of spike proteinsof SARS-CoV-2, SARS-CoV and MERS-CoV, and their interactions with ACE2.Cells,2020:9:2638.
46. Premkumar L, Segovia-Chumbez B, Jadi R, Martinez DR, Raut R, Markmann AJ, *et al*. The receptor-binding domain of the viral spike protein is an immunodominant and highly specific target of antibodies in SARS-CoV-2 patients. Sci. Immunol,2020:5(48):e8413,
47. Tomaszewski T, DeVries RS, Dong M, Bhatia G, Norsworthy MD, Zheng X. New pathways of mutational change in SARS-CoV-2 proteomes involve regions of intrinsic disorder important for virus replication and release. Evol. Bioinform. Online,2020:16:1176.
48. Ganesh B, Rajakumar T, Malathi M, Manikandan N, Nagaraj J, Santhakumar A. Epidemiology and pathobiology of SARS-CoV-2 (COVID-19) in comparison with SARS, MERS: an updated overview of current knowledge and future perspectives. Clin Epidemiol Glob Health,2021:10:100694,
49. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol, 2013.
50. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics,2009:25(15):1972–3.
51. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data – from vision to reality. Eurosurveillance, 2017.
52. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics,2014:30(9):1312–3.
53. Treangen TJ, Ondov BD, Koren S, Phillippy AM. The harvest suite for rapid coregenome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol, 2014, 15(11).
54. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang LL, *et al*. 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), 2012.
55. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. Bioinformatics, 2006.
56. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER suite: Protein structure and function prediction. Nature Methods, 2014.
57. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. Int J Infect Dis, 2020.
58. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. Proc R Soc B Biol Sci,2007:274:599-604.