IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Evolutionary Analysis of SARS-CoV-2 genome and protein insights the- origin of the virus, Wuhan

Sudheer Menon
Department of Bioinformatics,
Bharathiar University

ABSTRACT

The family of Coronaviridae, sub-family of *Coronavirinae, Nidovirale's*order and beta-coronavirus genus all are taxonomically categorized as SARS-CoV-2. Its genome consists of positive-sense strand, single-strand of RNA and non-segmented strand.On 31st of December 2019 SARS-CoV-2 was reported in a 41-year-oldpatient.On one week of presentation, the patient was reported with some symptoms like, body pain, cough, temperature, chest rigidity, and weakness. This article is about the origin of SARS-CoV-2 from Wuhan on the base of mutation analysis.To analyze the mutations in this firstly reported SARS-CoV-2 in Wuhan city Wuhan genome from NCBI GeneBank reference NC 045512.2 SARS-CoV-2 and 48,635 SARS-CoV-2 genomic sequences on GISAID server reported in different continents Africa, Asia, Europe, North America, Oceania and South America was used. First of all, MEGA tool was used to tell the classification (Clades) of different strains of SARS-CoV-2 on the phylogenetic tree. After that NUCMER (MUMmer) tool was used and results reported the types of mutations and their prevalence's in every continent. In this article we have reported the most common type of mutational event and the type of nucleotide most mutated.

Keywords: SARS COV2, WUHAN, Coronavirus, genome, protein.

INTRODUCTION

On 31st of December 2019 patient report showed that a 41-year-old man with no background of diabetes, TB and hepatitis. On December 26, 2019, after six days' outbreak of disease, in central hospital of Wuhan's he was admitted and hospitalized to. On one week of presentation, the patient was reported with some symptoms like, body pain, cough, temperature, chest rigidity, and weakness (Y. Huang et al., 2004; Kahn & McIntosh, 2005; Wu et al., 2020). The first evidencewas indicativethe epidemic condition being related with a sea-foodmarketplace in the Wuhan city, which was locked on 1stof January 2020. These agents of aetiologicalwas categorized as a beta-coronavirus similar to SARS, later called SARS-CoV-2 and on NCBI Gen-bank the first whole genome sequence was reported on 5 January 2020 (Hamre & Procknow, 1966; van Dorp et al., 2020; Wu et al., 2020).

From starting so many patients diagnosed with lymphopenia also abnormalities of platelets as well as neutrophils also detected with important body enzymatic infections like aspartate aminotransferase (AST), lactate dehydrogenase (LDH) and also some of the include inflammatory biomarkers. About 10.3% of effected persons had bilateral pneumonia with pleural effusion that was observed by the check up through CT scan and X-rays processing. These patients had a lower level of albumin and platelets with increased concentration of LDH, AST and neutrophils, as compared to general patients these patients have higher level of reactive protein. And these refractory patients had increasing level of bilateral pneumonia and pleural effusion can also be visualized in patients of COVID-19 (C. Huang et al., 2020).Throughout to this pandemic it is observed that it is spreading by the clusters of family members workers of medical facilities dealing with COVID-19 patients, this virus is human to human transmissible via contacts, droplets, fomites and also include the 3 days of incubation periods, while these patients can also spread virus with no symptoms of 0 to 24 days of exposure (Balboni et al., 2012) in diagnosis of bat SARS like CoVs the effectiveness of RT-PCR has been reported. So for the detection of COVID-19 in clinics a reverse transcription based RT-PCR assay is used that is already ongoing of SARS-CoV-2 (Chan et al., 2020).

The family of Coronaviridae, sub-family of *Coronavirinae, Nidovirale's*order and beta-coronavirus genus all are taxonomically categorized as SARS-CoV-2. It's an encapsulated virus consist of positive-sense strand, single-strand of RNA and non-segmented strand. Though SARS-CoV-2 has minimum effects of pathogenicity as compared to coronavirus syndrome SARS-CoV firstly prepared in 2002 to 2003 and other coronavirus syndrome MERS-CoV in Middle-East that appeared in 2012. Studies on MERS-CoVhave reported that the rate at which MERS-CoV transmit between is far high than SARS-CoV-2 (Ahmed et al., 2020; Rehman et al., 2020; Woo et al., 2009). The SARS-CoV-2 genome is composed of the structural and non-structural types of protein, also on non-segmented type of RNA, various accessory protein'sORF's (open reading frame) and 3'and 5' regions of UTR. two to third parts of RNA encoded by three proteins only RNA making constituents and two large sized polyproteins not involved in the structural formation of virus, the most important one is viral polymerase (RdRp) which replicates the viral RNA but they are not directly involved in modulation of host response. (ORF1a-ORF1b) are the remining open reading frames which encodes four proteins which are involved in structure formation of virus spike(S), a membrane (M), an envelope (E) and a nucleocapsid (N) in addition with another set of proteins known as helper proteins(Luk et al., 2019). For the viral particles some important structural proteins involve the nucleocapsid, membrane, spike and envelop proteins. The receptor binding domain, of the spike proteins played vital role for the direct binding of human receptors, for entry of virus, identified the host tropism and their capacity of transmission (Chan et al., 2020; Hogue & Machamer, 2007; Ou et al., 2017; Schoeman & Fielding, 2019). The Spike proteins divided into two main subunits S1 and S2. The function of S1 that directly identifies and binds to the human receptor ACE2. The S2 subunit is very important for the viral life cycle as this subunit interact and submerge itself in the membrane of epithelial cells of lungs in human host after which the virus enter into the cell and start its pathogenicity (Li et al., 2005). In common the RNA viruses such as SARS-CoV2 undergo the diversity of genetics, fast mutation and allowing evolutionaryas a result of some alterations like, affinity of receptor, host tropism, pathogenicity and viral transmissibility.

Here we reported all mutations and satisfy them by highlighting genomically and geographically and alsoemphasizing the uprising of sub-clades and highly genomic variable spots. These studies can be enormouslyvaluable to plan and think about the effectiveness of measuresthat have been classified on continental basis to limit SARS-CoV-2 spreading

OBJECTIVES

1. Evolutionary analysis by comparing the full genome sequences of SARS-CoV-2 present in the data base with genome sequence of Wuhan virus.
2. Enlisting mutation events in different regions of world.

MATERIAL AND METHODS

MEGA

In order to offer statistical methodologies for molecular development via a collective interface on MS-DOS (Microsoft Disc Operating System), it was issued first in 1933. At first, MEGA consist of distance based and highest parsimony methodologies of molecular phylogenetic execution(S Kumar et al., 1994). The data accession and integration of main perspectives for arranging sequences were introduced to dilate MEGA’s scope(Sudhir Kumar et al., 2004). After that, the maximum likelihood (ML) methodologies for analysis of molecular evolution were added(Tamura et al., 2011). Now MEGA consist of methodologies for selection of best fit substitution model(s), assessing distances of evolution and division times, regenerating phylogenies, estimating ancestral distance, examining for selection and identifying disease mutations(Caspermeyer, 2018; Tamura et al., 2021).

GISAID (Global Initiative on Sharing All Influenza Data):

Platform for GISAID was started on event of in May 2008 in Sixty-first World Health Assembly. Made as a substitute to share the model to public domain, sharing the GISAIDs mechanism came into an account of Member States by giving a publicly available databased arranged by scientist for another scientist, to enhance the sharing for data of influenza. For sharing the data between National Influenza Centers and WHO collaborating Centers, GISAD is playing an important role from the time of its launch, for vaccine virus of bi-annual influenza suggestions by WHO GISRS (Global Influenza Surveillance and Response System). This initiative of GISAID increases the fast data sharing from all corona and influenza viruses which are major cause of COVID-19. In order to help the researchers to understand viruses’ evolution and their epidemic and pandemic spreading, it consists of genetic sequences and associated epidemiological and clinical data regarding human viruses, specie specific data and geographical data related to avian and animal viruses(Mercatelli & Giorgi, 2020; Shu & McCauley, 2017).

MUMmer:

For rapid arrangement of protein and DNA sequences, MUMmer is a system. It is system for better arrangement of DNA sequences on larger scales, for data arrangement of millions of nucleotides. MUMmer is an alignment system, has the ability of arranging entire genome of bacteria in less than one minute on a standard computer. The major algorithm for MUMmer requires two input sequences, may be DNA or proteins and determines whole subsequence longer than a specific least length k which are similar among two inputs. There is maximum guarantee of these matches, in which they can't be elongated without finding a mismatch on either end, and they are distinctive in an accurate system, as described below. This entire algorithm is imposed by using the data structure of suffix-tree, which allows very rapid and memory-capable contrast of sequences. The promer usage produces alignments on basis of six frame translations for both input sequences. Promer allows allows the arrangement of genomes for the similar proteins while the sequence of DNA is very large to find similarity (A L Delcher et al., 1999; Arthur L Delcher et al., 2002).

RESULTS AND DISCUSSION:

Phylogenetic analysis of SARS-CoV-2

Understand the strains for the monophyletic distribution of the SARSCoV-2 population. According to prior reports byGISAID the phylogenetic tree exhibited three main clades (S, G, and V clades). By the mutations of L84S (ORF8), D614G (S), and G251V (ORF3a) these clades (S, G, and V clades) were determined (Figure 1). The mention strain (hCoV-19/Wuhan-Hu-1/2019) is specified on a gray screen by bold letters. The amino acid changes here show the substitution of amino acid on left of the stanadrd strain to a variance amino acid of the parallel strain on right. Muatation [D614G (S)] in G clade is significat among the three mutations that determine the clades. D614G (S)is found inthe nearby polybasic cleavage location. However its main function on RDB/ACE2 binding is indistinct. Along with three clades, Phylogenetic study was approved out wrepresent two subclades that belong to G clade named as G.1 and G.2 clades mentioningtheir parent clade name.The G.1 subclade was determined by three mutations[G204R (N), R203K (N), and P214L (ORF1b)] and G.2 subclade was determined by mutation [Q57H (ORF3a)] (Figure 2). G clade and its subclades,phylogenetic tree were determined by D614G (S)Q57H (ORF3a) and G204R (N), R203K (N), as well as P214L (ORF1b), are shared by the G.1 and G.2 clades, respectively.In this regard, the G.1 and G.2 subclades' origin from the G clade demonstrates that a clade can be determined through one or even more mutation.

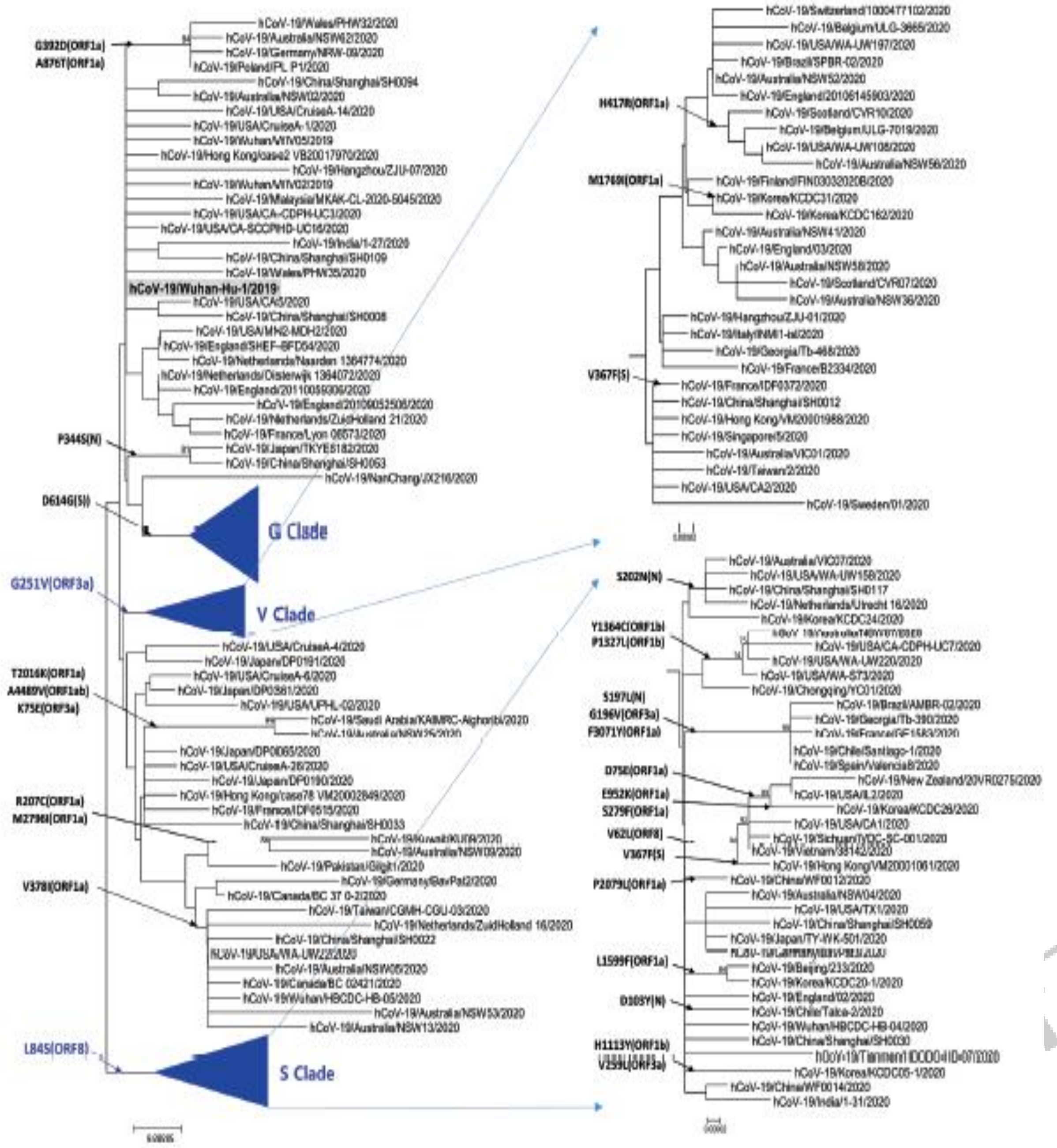
hCoV-19/Wales/PHW32/2020
hCoV-19/Australia/NSW62/2020
hCoV-19/Germany/NRW-09/2020
hCoV-19/Poland/PL P1/2020
hCoV-19/China/Shanghai/SH0094
hCoV-19/Australia/NSW02/2020
hCoV-19/USA/CruiseA-14/2020
hCoV-19/USA/CruiseA-1/2020
hCoV-19/Wuhan/WIV05/2019
hCoV-19/Hong Kong/case2 VB20017970/2020
hCoV-19/Hangzhou/ZJU-07/2020
hCoV-19/Wuhan/WIV02/2019
hCoV-19/Malaysia/MKAK-CL-2020-5045/2020
hCoV-19/USA/CA-CDPH-UC3/2020
hCoV-19/USA/CA-SCCPHD-UC16/2020
hCoV-19/India/1-27/2020
hCoV-19/China/Shanghai/SH0109
hCoV-19/Wales/PHW35/2020
hCoV-19/Wuhan-Hu-1/2019
hCoV-19/USA/CA5/2020
hCoV-19/China/Shanghai/SH0008
hCoV-19/USA/MH2-MDH2/2020
hCoV-19/England/SHEF-BFD54/2020
hCoV-19/Netherlands/Naarden 1364774/2020
hCoV-19/Netherlands/Oisterwijk 1364072/2020
hCoV-19/England/20110059306/2020
hCoV-19/England/20100052506/2020
hCoV-19/Netherlands/ZuidHolland 21/2020
hCoV-19/France/Lyon 08573/2020
hCoV-19/Japan/TKYE6182/2020
hCoV-19/China/Shanghai/SH0053
hCoV-19/NanChang/JX216/2020
G392D(ORF1a)
A876T(ORF1a)
P344S(N)
D614G(S)
G Clade
G251V(ORF3a)
V Clade
hCoV-19/USA/CruiseA-4/2020
hCoV-19/Japan/DP0191/2020
hCoV-19/USA/CruiseA-6/2020
hCoV-19/Japan/DP0381/2020
hCoV-19/USA/UPHL-02/2020
hCoV-19/Saudi Arabia/KAIMRC-Alghoribi/2020
hCoV-19/Australia/NSW25/2020
hCoV-19/Japan/DP0085/2020
hCoV-19/USA/CruiseA-26/2020
hCoV-19/Japan/DP0190/2020
hCoV-19/Hong Kong/case78 VM20002849/2020
hCoV-19/France/IDF0515/2020
hCoV-19/China/Shanghai/SH0033
hCoV-19/Kuwait/KU08/2020
hCoV-19/Australia/NSW09/2020
hCoV-19/Pakistan/Gilgit1/2020
hCoV-19/Germany/BavPat2/2020
hCoV-19/Canada/BC 37 0-2/2020
hCoV-19/Taiwan/CGMH-CGU-03/2020
hCoV-19/Netherlands/ZuidHolland 16/2020
hCoV-19/China/Shanghai/SH0022
hCoV-19/USA/WA-UW22/2020
hCoV-19/Australia/NSW05/2020
hCoV-19/Canada/BC 02421/2020
hCoV-19/Wuhan/HBCDC-HB-05/2020
hCoV-19/Australia/NSW53/2020
hCoV-19/Australia/NSW13/2020
T2016K(ORF1a)
A4489V(ORF1ab)
K75E(ORF3a)
R207C(ORF1a)
M2796I(ORF1a)
V378I(ORF1a)
L84S(ORF8)
S Clade
hCoV-19/Switzerland/1000477102/2020
hCoV-19/Belgium/ULG-3665/2020
hCoV-19/USA/WA-UW197/2020
hCoV-19/Brazil/SPBR-02/2020
hCoV-19/Australia/NSW52/2020
hCoV-19/England/20106145903/2020
hCoV-19/Scotland/CVR10/2020
hCoV-19/Belgium/ULG-7019/2020
hCoV-19/USA/WA-UW108/2020
hCoV-19/Australia/NSW56/2020
H417R(ORF1a)
M1769I(ORF1a)
hCoV-19/Finland/FIN03032020B/2020
hCoV-19/Korea/KCDC31/2020
hCoV-19/Korea/KCDC162/2020
hCoV-19/Australia/NSW41/2020
hCoV-19/England/03/2020
hCoV-19/Australia/NSW58/2020
hCoV-19/Scotland/CVR07/2020
hCoV-19/Australia/NSW36/2020
hCoV-19/Hangzhou/ZJU-01/2020
hCoV-19/Italy/INMI1-isl/2020
hCoV-19/Georgia/Tb-468/2020
hCoV-19/France/B2334/2020
V367F(S)
hCoV-19/France/IDF0372/2020
hCoV-19/China/Shanghai/SH0012
hCoV-19/Hong Kong/VM20001988/2020
hCoV-19/Singapore/5/2020
hCoV-19/Australia/VIC01/2020
hCoV-19/Taiwan/2/2020
hCoV-19/USA/CA2/2020
hCoV-19/Sweden/01/2020
S202N(N)
hCoV-19/Australia/VIC07/2020
hCoV-19/USA/WA-UW158/2020
hCoV-19/China/Shanghai/SH0117
hCoV-19/Netherlands/Utrecht 16/2020
hCoV-19/Korea/KCDC24/2020
Y1364C(ORF1b)
P1327L(ORF1b)
hCoV-19/USA/CA-CDPH-UC7/2020
hCoV-19/USA/WA-UW220/2020
hCoV-19/USA/WA-S73/2020
hCoV-19/Chongqing/YC01/2020
S197L(N)
G196V(ORF3a)
F3071Y(ORF1a)
hCoV-19/Brazil/AMBR-02/2020
hCoV-19/Georgia/Tb-390/2020
hCoV-19/Chile/Santiago-1/2020
hCoV-19/Spain/Valencia8/2020
D75E(ORF1a)
E952K(ORF1a)
S279F(ORF1a)
hCoV-19/New Zealand/20VR0275/2020
hCoV-19/USA/IL2/2020
hCoV-19/Korea/KCDC26/2020
hCoV-19/USA/CA1/2020
V62L(ORF8)
V367F(S)
hCoV-19/Sichuan/IVDC-SC-001/2020
hCoV-19/Vietnam/38142/2020
hCoV-19/Hong Kong/VM20001061/2020
P2079L(ORF1a)
hCoV-19/China/WF0012/2020
hCoV-19/Australia/NSW04/2020
hCoV-19/USA/TX1/2020
hCoV-19/China/Shanghai/SH0059
hCoV-19/Japan/TY-WK-501/2020
L1599F(ORF1a)
hCoV-19/Beijing/233/2020
hCoV-19/Korea/KCDC20-1/2020
hCoV-19/England/02/2020
hCoV-19/Chile/Talca-2/2020
D103Y(N)
hCoV-19/Wuhan/HBCDC-HB-04/2020
hCoV-19/China/Shanghai/SH0030
H1113Y(ORF1b)
V259L(ORF3a)
hCoV-19/Tianmen/HBCDC-HB-07/2020
hCoV-19/Korea/KCDC05-1/2020
hCoV-19/China/WF0014/2020
hCoV-19/India/1-31/2020

Figure 1: Phylogenetic Tree Clade S and Clade V, Wuhan virus (SARS-CoV-2)

Genome Sequence Alignment for mutations detection Geographically

On June 26, 2020, GISAID (Shu and McCauley, 2017) was used to obtain 48,635 SARS-CoV-2 genomic sequences.Only viruses that infect humans were chosen, with low-quality sequences (>5% NNNs) removed and only full-length sequences (>29,000 nt) used. A total of 48,624 sequences were assigned to a geographic location, with 514 from Africa, 3,340 from Asia, 31,818 from Europe, 10,250 from North America, 2,127 from Oceania, and 575 from South America. There were eleven sequences that were not linked to any continent. NCBI GenBank provided the reference NC 045512.2 SARS-CoV-2 Wuhan genome (Gorbalenya et al., 2020), which is 29,903 nucleotides long. A GFF3 annotation associated with the reference, displaying genomic locations for all SARS-CoV-2 protein sequences. Non-structural proteins were separated from the vast ORF1 polyprotein (NSPs).

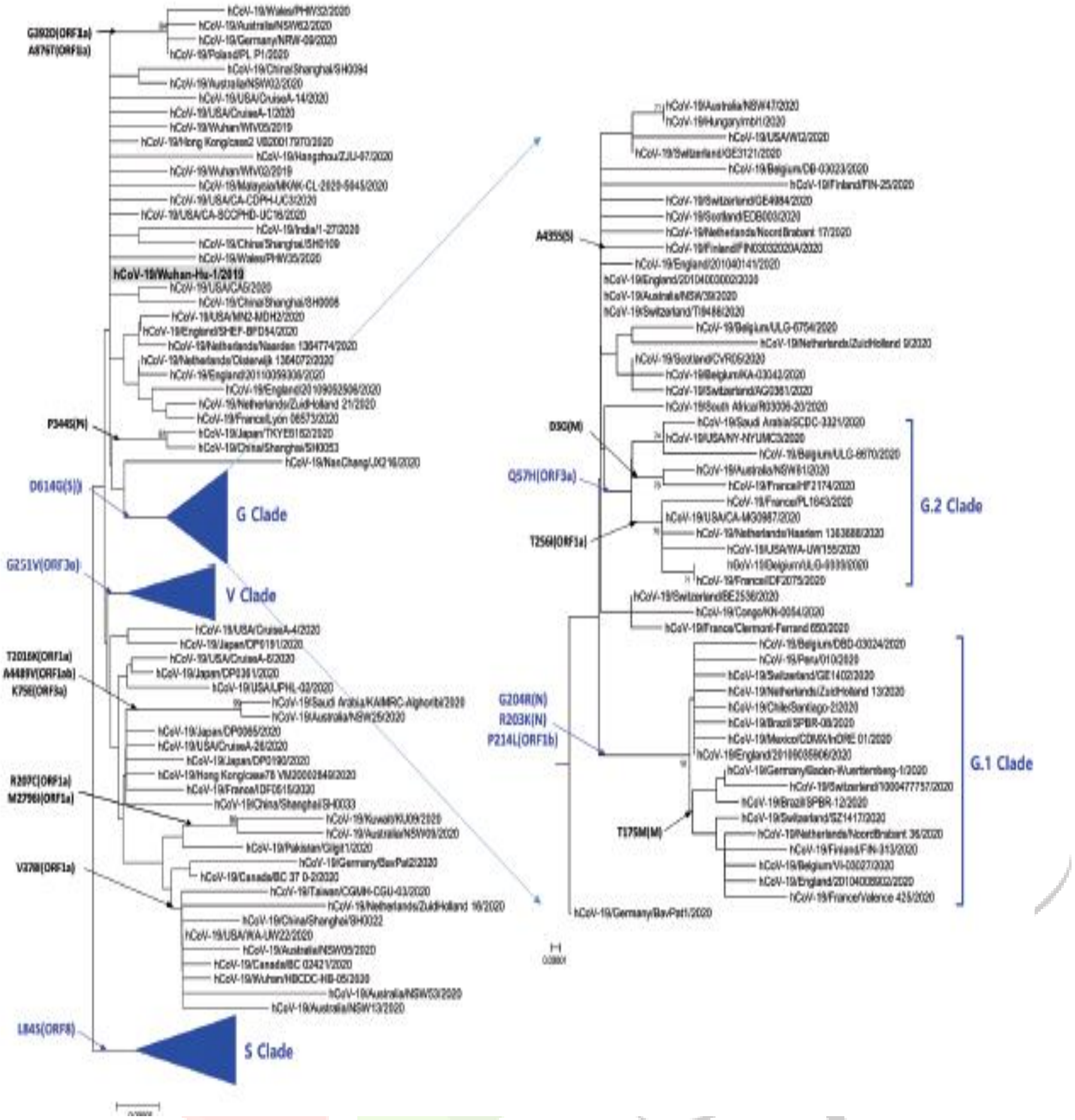

Figure 2: Phylogenetic Tree Clade G and Subclade G1 and G2, Wuhan virus (SARS-CoV-2)

The NSP12 gene, which encodes the viral Gene RNA polymerase, was annotated as two sections, NSP12a as well as NSP12b, according to the domains like during a ribosomal frameshift, which occurs when nucleotide 13,468 is translated but both the last first and nucleotide of a codon. Version 3.1 of NUCMER(Arthur L Delcher et al., 2002). Over the NC 045512.2 reference, all 48,635 genomic sequences were aligned. Just use an internally designed R SARS-CoV-2 annotation technique, the alignment result was converted to an attributed list of all point mutation. 7.23 on average, with only a few samples having and over 15 events. Overall, no region differs considerably from the mean mutation rate, however there is a significant variation in the average value of mutations each sample between states (one-way ANOVA p = 9.55 10205). These countries have such a slightly but significantly greater number of detected mutations per s of the top 40 countries with the largest number of sequenced complete viral genomes, when compared to the world's average shown in Table 1.

Table 1: Showing type and number of mutation events in different regions.

Sr #	Region	Sample	Nr of events (Highest)
1	Africa	514	2K (SNP)
2	Asia	3340	12K (SNP)
3	Europe	31818	120K (SNP)
4	North America	10250	40K (SNP)
5	Oceania	2127	8K (SNP)
6	South America	575	2K (SNP)

We looked examined the type of each mutation and discovered that single-nucleotide polymorphisms (SNPs) outnumber short implantation events (indels) in every continent. We found 205,482 amino acid(aa)-changing SNP occurrences worldwide (58.2% of all SNPs), with fewer than 50% of silent SNPs occurring in coding areas (27.6 percent , with 97,573 events). There are 44,345 occurrences in coding region (12.6%) of the SARS-CoV-2 RNA sequence, mostly in the 5′UTR and 3′UTR. Short frameshift losses account for 0.8 percent of all observed mutational events in the SARSCoV-2 population, following by in-frame deletions (3x deletions diminishing the viral protein length without generating stop codons), which account for 0.6 percent. SNPs that result in a stop codon are relatively uncommon (496 observed events, 0.1 percent of the total). Insertions are exceedingly rare, accounting for less than 0.1 percent of all SARS-CoV-2 mutations discovered thus far. All continents had similar profiles and comparative percentages, implying a conserved molecular basis for SARS-CoV-2 evolution. The SARS-CoV-2 mutations were then categorised according to its nature, with SNP transitions (purine->purine and pyrimidine->pyrimidine) outnumbering SNP transversions (purine->pyrimidine and vice versa), a finding that parallels that of SARS-CoV. The C>T transition is perhaps the most common occurrence, both globally and across continents, accounting for 55.1 percent of all identified worldwide viral mutations Table 2.

Table 2: Showing type and number of nucleotide mutation events in different regions.

Sr #	Region	Sample	Nr of events (Highest)
1	Africa	514	2K (C-T)
2	Asia	3340	12K (C-T)
3	Europe	31818	120K (C-T)
4	North America	10250	40K (C-T)
5	Oceania	2127	8K (C-T)
6	South America	575	2K (C-T)

CONCLUSION

Phylogenetic analysis done by MEGA tool has shown that on the basis of mutations, Phylogenetic tree of SARS-CoV-2 contain three clades (S, G and V) based on the mutation of different genes of genome L84S (ORF8), D614G (S), and G251V (ORF3a). For mutational analysis Wuhan strain genome (reference NC 045512.2 SARS-CoV-2) is selected from NCBI GenBank and 48,635SARS-CoV-2 genomic sequences from different continents published on GISAID were used. NUCMER v3.1 (MUMmer) tool was used to align 48,635 SARS-CoV-2 genomic sequences with Wuhan strain genome (reference NC 045512.2 SARS-CoV-2) and the results have shown that each continent has highest number of SNP mutation events and the mostcommon mutation in nucleotides is of type C>T in every continent. Depending upon results it was concluded that Wuhan strain genome (reference NC 045512.2 SARS-CoV-2) is the first SARS-CoV-2 reported and all other strains were originated from Wuhan strain caused by the mutational events in the genome.

REFERENCES

Ahmed, S. F., Quadeer, A. A., & McKay, M. R. (2020). Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. *Viruses*, *12*(3), 254–269. https://doi.org/10.3390/v12030254

Balboni, A., Gallina, L., Palladini, A., Prosperi, S., & Battilani, M. (2012). A real-time PCR assay for bat SARS-like coronavirus detection and its application to Italian greater horseshoe bat faecal sample surveys. *Scientific World Journal*, *2012*, 989514. https://doi.org/10.1100/2012/989514

Caspermeyer, J. (2018). MEGA Software Celebrates Silver Anniversary. In *Molecular biology and evolution* (Vol. 35, Issue 6, pp. 1558–1560). https://doi.org/10.1093/molbev/msy098

Chan, J. F.-W., Kok, K.-H., Zhu, Z., Chu, H., To, K. K.-W., Yuan, S., & Yuen, K.-Y. (2020). Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging Microbes & Infections*, *9*(1), 221–236. https://doi.org/10.1080/22221751.2020.1719902

Delcher, A L, Kasif, S., Fleischmann, R. D., Peterson, J., White, O., & Salzberg, S. L. (1999). Alignment of whole genomes. *Nucleic Acids Research*, *27*(11), 2369–2376. https://doi.org/10.1093/nar/27.11.2369

Delcher, Arthur L, Phillippy, A., Carlton, J., & Salzberg, S. L. (2002). Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Research*, *30*(11), 2478–2483. https://doi.org/10.1093/nar/30.11.2478

Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, R. J., Drosten, C., Gulyaeva, A. A., Haagmans, B. L., Lauber, C., Leontovich, A. M., Neuman, B. W., Penzar, D., Perlman, S., Poon, L. L. M., Samborskiy, D. V, Sidorov, I. A., Sola, I., Ziebuhr, J., & Viruses, C. S. G. of the I. C. on T. of. (2020). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology*, *5*(4), 536–544. https://doi.org/10.1038/s41564-020-0695-z

Hamre, D., & Procknow, J. J. (1966). A new virus isolated from the human respiratory tract. *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine (New York, N.Y.)*, *121*(1), 190–193. https://doi.org/10.3181/00379727-121-30734

Hogue, B. G., & Machamer, C. E. (2007). Coronavirus Structural Proteins and Virus Assembly. In *Nidoviruses* (pp. 179–200). https://doi.org/doi:10.1128/9781555815790.ch12

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., … Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet (London, England)*, *395*(10223), 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5

Huang, Y., Yang, Z., Kong, W., & Nabel, G. J. (2004). Generation of Synthetic Severe Acute Respiratory Syndrome Coronavirus Pseudoparticles: Implications for Assembly and Vaccine Production. *Journal of Virology*, *78*(22), 12557–12565. https://doi.org/10.1128/JVI.78.22.12557-12565.2004

Kahn, J. S., & McIntosh, K. (2005). History and recent advances in coronavirus discovery. *The Pediatric Infectious Disease Journal*, *24*(11 Suppl), S223—7, discussion S226. https://doi.org/10.1097/01.inf.0000188166.17324.60

Kumar, S, Tamura, K., & Nei, M. (1994). MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Computer Applications in the Biosciences : CABIOS*, *10*(2), 189–191. https://doi.org/10.1093/bioinformatics/10.2.189

Kumar, Sudhir, Tamura, K., & Nei, M. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Briefings in Bioinformatics*, *5*(2), 150–163. https://doi.org/10.1093/bib/5.2.150

Li, F., Li, W., Farzan, M., & Harrison, S. C. (2005). Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science (New York, N.Y.)*, *309*(5742), 1864–1868. https://doi.org/10.1126/science.1116480

Luk, H. K. H., Li, X., Fung, J., Lau, S. K. P., & Woo, P. C. Y. (2019). Molecular epidemiology, evolution and phylogeny of SARS coronavirus. *Infection, Genetics and Evolution*, *71*, 21–30. https://doi.org/https://doi.org/10.1016/j.meegid.2019.03.001

Mercatelli, D., & Giorgi, F. M. (2020). Geographic and Genomic Distribution of SARS-CoV-2 Mutations. *Frontiers in Microbiology*, *11*(July), 1–13. https://doi.org/10.3389/fmicb.2020.01800

Ou, X., Guan, H., Qin, B., Mu, Z., Wojdyla, J. A., Wang, M., Dominguez, S. R., Qian, Z., & Cui, S. (2017). Crystal structure of the receptor binding domain of the spike glycoprotein of human betacoronavirus HKU1. *Nature Communications*, *8*(1), 15216. https://doi.org/10.1038/ncomms15216

Rehman, S. U., Shafique, L., Ihsan, A., & Liu, Q. (2020). Evolutionary trajectory for the emergence of novel coronavirus SARS-CoV-2. *Pathogens*, *9*(3), 1–12. https://doi.org/10.3390/pathogens9030240

Schoeman, D., & Fielding, B. C. (2019). Coronavirus envelope protein: current knowledge. *Virology Journal*, *16*(1), 1–69. https://doi.org/10.1186/s12985-019-1182-0

Shu, Y., & McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data - from vision to reality. In *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* (Vol. 22, Issue 13). https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494

Sudheer Menon (2020) "Preparation and computational analysis of Bisulphite sequencing in Germfree Mice" *International Journal for Science and Advance Research In Technology*, 6(9) PP (557-565).

Sudheer Menon, Shanmughavel Piramanayakam and Gopal Agarwal (2021) "Computational identification of promoter regions in prokaryotes and Eukaryotes" EPRA International Journal of Agriculture and Rural Economic Research (ARER), Vol (9) Issue (7) July 2021, PP (21-28).

Sudheer Menon (2021) "Bioinformatics approaches to understand gene looping in human genome" EPRA International Journal of Research & Development (IJRD), Vol (6) Issue (7) July 2021, PP (170-173).

Sudheer Menon (2021) "Insilico analysis of terpenoids in Saccharomyces Cerevisiae"international Journal of Engineering Applied Sciences and Technology, 2021 Vol. 6, Issue1, ISSN No. 2455-2143, PP(43-52).

Sudheer Menon (2021) "Computational analysis of Histone modification and TFBs that mediates gene looping" Bioinformatics, Pharmaceutical, and Chemical Sciences (RJLBPCS), June 2021, 7(3) PP (53-70).

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,7(7) PP (620-623).

Sudheer Menon, Shanmughavel Piramanayakam and Gopal Agarwal (2021) Computational Identification of promoter regions in fungal genomes, International Journal of Advance Research, Ideas and Innovations in Technology, 7(4) PP (908-914).

Sudheer Menon, Vincent Chi Hang Lui and Paul Kwong Hang Tam (2021) Bioinformatics methods for identifying hirschsprung disease genes, International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 9 Issue VII July, PP (2974-2978).

Sudheer Menon, (2021), Bioinformatics approaches to understand the role of African genetic diversity in disease, International Journal Of Multidisciplinary Research In Science, Engineering and Technology (IJMRSET), 4(8), PP 1707-1713.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, *28*(10), 2731–2739. https://doi.org/10.1093/molbev/msr121

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, *38*(7), 3022–3027. https://doi.org/10.1093/molbev/msab120

van Dorp, L., Acman, M., Richard, D., Shaw, L. P., Ford, C. E., Ormond, L., Owen, C. J., Pang, J., Tan, C. C. S., Boshier, F. A. T., Ortiz, A. T., & Balloux, F. (2020). Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infection, Genetics and Evolution*, *83*(April), 104351. https://doi.org/10.1016/j.meegid.2020.104351

Woo, P. C. Y., Lau, S. K. P., Huang, Y., & Yuen, K.-Y. (2009). Coronavirus diversity, phylogeny and interspecies jumping. *Experimental Biology and Medicine (Maywood, N.J.)*, *234*(10), 1117–1127. https://doi.org/10.3181/0903-MR-94

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C., & Zhang, Y.-Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, *579*(7798), 265–269. https://doi.org/10.1038/s41586-020-2008-3