IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Genomic variation study of COVID-19 sequences from India

Sofia Priyadarsani Das[1]*, Rashmi Ranjeeta Das[2]

[1]Amity Institute of Marine Science and Technology, Amity University Uttar Pradesh, ector-125, Noida

[2]DDCE, Utkal University, Bhubaneswar

Abstract:

The Whole world is facing the pandemic of COVID-19 infection today after it detected in December 2019 in China. From two confirm cases in India at the early 2020, it became 216,919 as of now. Though India is facing a high infection rate, death rate is much lower than other countries. So, as to know the genomic level variation and to analyze the main cause behind the types the present investigation has been prepared. The Present study showed 5 haplotypes from the complete genome sequences publicly available in GenBank which could be further use for the variation studies leading to the types and severity of this virus.

Introduction:

COVID-19 contamination, which is caused by Sars-Cov2, is an RNA virus (Forni et al.2017). The RNA genome of the virus has two interesting character, RNA is unstable due to its chemical nature; therefore, it is susceptible to mutations. The other characteristics are that unlike genomes made of DNA, RNA genomes are incapable of correcting such mutations. Being an RNA virus, Sars-Cov2 usually undergoes much faster genomic changes once they enter their host. Therefore, the aim of the study is to analyse the Sars-Cov2 sequence variation seen in the sequences deposited in the GenBank and search for the variation in the Indian patient sequences. Today, whole world is at a verge of serious crisis due to COVID-19 (coronavirus disease-2019) pandemic caused by a novel coronavirus SARSCoV-2 (severe acute respiratory syndrome-coronavirus-2). Till today the total infection worldwide is about 6,287,771 with 379,941 deaths whereas, in India the confirmed cases are 216,919 and about 6,074 deaths are being noticed. The disease was first reported from Wuhan, China in December 2019 and spread to almost every part of the world in a short span of Time (Zhu et al., 2020, Lu et al. 2020). Till now only two complete genomes have been deposited from India and others are partial ones (Yadav et al, 2020). As Viruses, or any organism, develop minor but permanent changes in their genetic codes, called mutations/ variation, over a period of time due to a variety of factors, including climatic and environmental conditions. The present investigation aims at the SARSCoV-2 sequence variations that are responsible for the diversity.

Materials and Methods:

A total of 29 whole genome sequences have been downloaded from publicly available portal of GenBank shown in Table-1, and put into Bioedit (Hall, 1999) for alignment and pre processing. The length of the sequences of this study is approximately 29kb. We have collected only the complete genome sequences from GenBank and taken for analysis. We have collected sequences from 20 different geographical region that includes, India, South Africa, USA, Spain, Israel, Greece, Peru, Brazil, Italy, Korea, Australia, Japan, Vietnam, Sweden, Hongkong, Turkey, Pakistan, Colombia, Taiwan and China. We have done the variant analysis in Bioedit. The analyzed sequences are then put in MEGAX (Kumar et al., 2018) for phylogenetic analysis.

Results and Discussion:

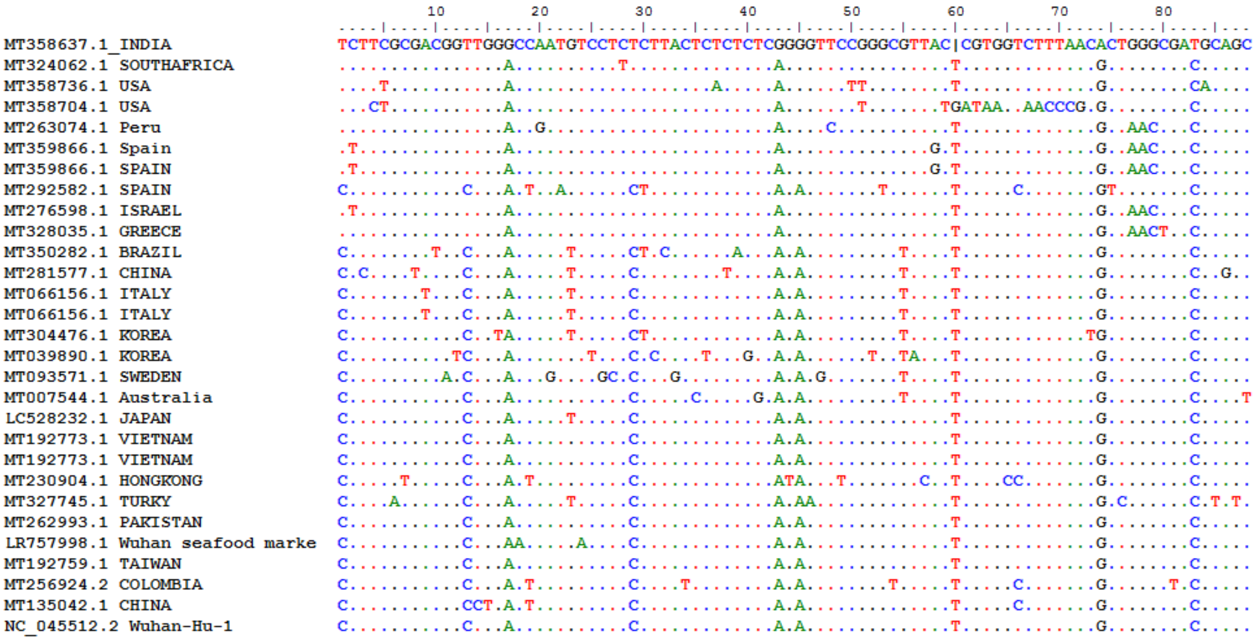
The Variant detection gives us many answers. There are about 88 Variants have been detected from these 29 Complete genome sequences after manual editing. Koyama et al. 2020 also found 80 variants from 48 complete genomes of COVID-19. As we all know the nucleotide sequences can give variation with respect to individuals, as the sequences are from different geographical regions. So if we ignore the individual SNPs found in the sequences, Also there are 5 haplotypes are present. The sequence analysis showed the sequences from all other countries are more or less different from Indian Sequences. Broadly the sequences are divided into mainly two types and then it could be further divided into other sub-groups. The first group comprised of sequences from India, USA, Greece, Spain, Peru, South Africa and Israel and all other falls into the second group (Fig.1 and Fig.2). The phylogenetic tree also depicts the two lineages of the corona viruses. Out-group formed by one of the sequence from USA because that sequence is having a long stretch of nucleotide variation. That could be due to individual variation. More on that could be resolved if the symptom and case history of the particular patient is being known. Scientists track the mutations happening in the viruses keenly, because these are key to understanding their behaviour, and in drug and vaccine development. According to the researchers, including those from the University of Cambridge in the UK, there are three distinct "variants" of COVID-19, consisting of clusters of closely related lineages, which they labelled 'A','B' and 'C'.

Conclusion:

Rapid data sharing is the basis for public health action. In conclusion, it is believe that genetic variation in SARS-CoV-2 may at least somewhat explain variable severity of COVID-19. Also we can further analyze the severity by sequencing the patients from India where death rate is much lower than USA, and comparative analysis of those sequences can give us the clue to win over this pandemic. The present investigation of the gene sequences of the SARSCoV-2 has so far not revealed anything to suggest that the virus present in India is in any way different from that circulating in other parts of the world. There is also no evidence, as of now, of any one particular strain of the virus being more deadly than the other. India, a country, having a huge population but till now we have been sequencing a very few genome of COVID-19 and available at public domain. So, our goal is to encourage all researchers to share their data as quickly and widely as possible.

References

1) Forni D, Cagliani R, Clerici M, Sironi M. (2017) Molecular Evolution of Human Coronavirus Genomes. *Trends in Microbiology* **25**: 35-48.
2) Hall TA. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. p. 95–98.
3) Kumar S, Stecher G, Li M, Knyaz C, and Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution* **35**:1547-1549.
4) Lu H, Stratton CW, Tang YW (2020). Outbreak of Pneumonia of Unknown Etiology in Wuhan China: the Mystery and the Miracle. Journal of Medical Virology,
5) Koyama T, Platt D, Parida L (2020) Variant analysis of COVID-19 genomes. Bull World Health Organ. E-pub: doi: http://dx.doi.org/10.2471/BLT.20.253591
6) Yadav PD, Potdar VA, Choudhary ML, Nyayanit DA, Agrawal M, Jadhav SM, Majumdar TD, Shete-Aich A, Basu A, Abraham P, Cherian SS (2020). Full-genome sequences of the first two SARS-CoV-2 viruses from India. *Indian Journal of Medical Research*, **151**: 200-209



```
                                        10        20        30        40        50        60        70        80
                                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|...
MT358637.1_INDIA                TCTTCGCGACGGTTGGGCCAATGTCCTCTCTTACTCTCTCTCGGGGTTCCGGGCGTTAC|CGTGGTCTTTAACACTGGGCGATGCAGC
MT324062.1 SOUTHAFRICA          ................A..........T..............A................T.............G........C.....
MT358736.1 USA                  ....T...........A...................A.....A......TT........T.............G........CA....
MT358704.1 USA                  ...CT...........A.........................A.......T.......TGATAA..AACCCG.G........C.....
MT263074.1 Peru                 ................A..G......................A....C...........T.............G..AAC...C.....
MT359866.1 Spain                .T..............A.........................A..............G.T.............G..AAC...C.....
MT359866.1 SPAIN                .T..............A.........................A..............G.T.............G..AAC...C.....
MT292582.1 SPAIN                C...........C...A.T..A......CT............A.A.......T......T.....C.......GT.......C.....
MT276598.1 ISRAEL               .T..............A.........................A................T.............G..AAC...C.....
MT328035.1 GREECE               ................A.........................A................T.............G..AACT..C.....
MT350282.1 BRAZIL               C........T..C...A.....T.....CT.C......A...A.A.........T....T.............G........C.....
MT281577.1 CHINA                C.C....T....C...A.....T.....C........T....A.A.........T....T.............G........C..G..
MT066156.1 ITALY                C.......T...C...A.....T.....C.............A.A.........T....T.............G........C.....
MT066156.1 ITALY                C.......T...C...A.....T.....C.............A.A.........T....T.............G........C.....
MT304476.1 KOREA                C...........C..TA.....T.....CT............A.A.........T....T............TG........C.....
MT039890.1 KOREA                C..........TC...A.......T...C.C....T...G..A.A......T..TA...T.............G........C.....
MT093571.1 SWEDEN               C.........A.C...A...G....GC.C...G.........A.A.G.......T....T.............G........C.....
MT007544.1 Australia            C...........C...A...........C.....C.....G.A.A.........T....T.............G........C....T
LC528232.1 JAPAN                C...........C...A.....T.....C.............A.A..............T.............G........C.....
MT192773.1 VIETNAM              C...........C...A...........C.............A.A..............T.............G........C.....
MT192773.1 VIETNAM              C...........C...A...........C.............A.A..............T.............G........C.....
MT230904.1 HONGKONG             C.....T.....C...A.T.........C.............ATA...T.......C..T....CC.......G........C.....
MT327745.1 TURKY                C....A......C...A.....T.....C.............A.AA.............T.............G.C......C.T.T.
MT262993.1 PAKISTAN             C...........C...A...........C.............A.A..............T.............G........C.....
LR757998.1 Wuhan seafood marke  C...........C...AA.....A....C.............A.A..............T.............G........C.....
MT192759.1 TAIWAN               C...........C...A...........C.............A.A..............T.............G........C.....
MT256924.2 COLOMBIA             C...........C...A.T.........C....T........A.A........T.....T.....C.......G......T.C.....
MT135042.1 CHINA                C...........CCT.A.T.........C.............A.A..............T.....C.......G........C.....
NC_045512.2 Wuhan-Hu-1          C...........C...A...........C.............A.A..............T.............G........C.....
```

Fig.1: Complete Genome Sequence variation of COVID-19 from 29 different individuals

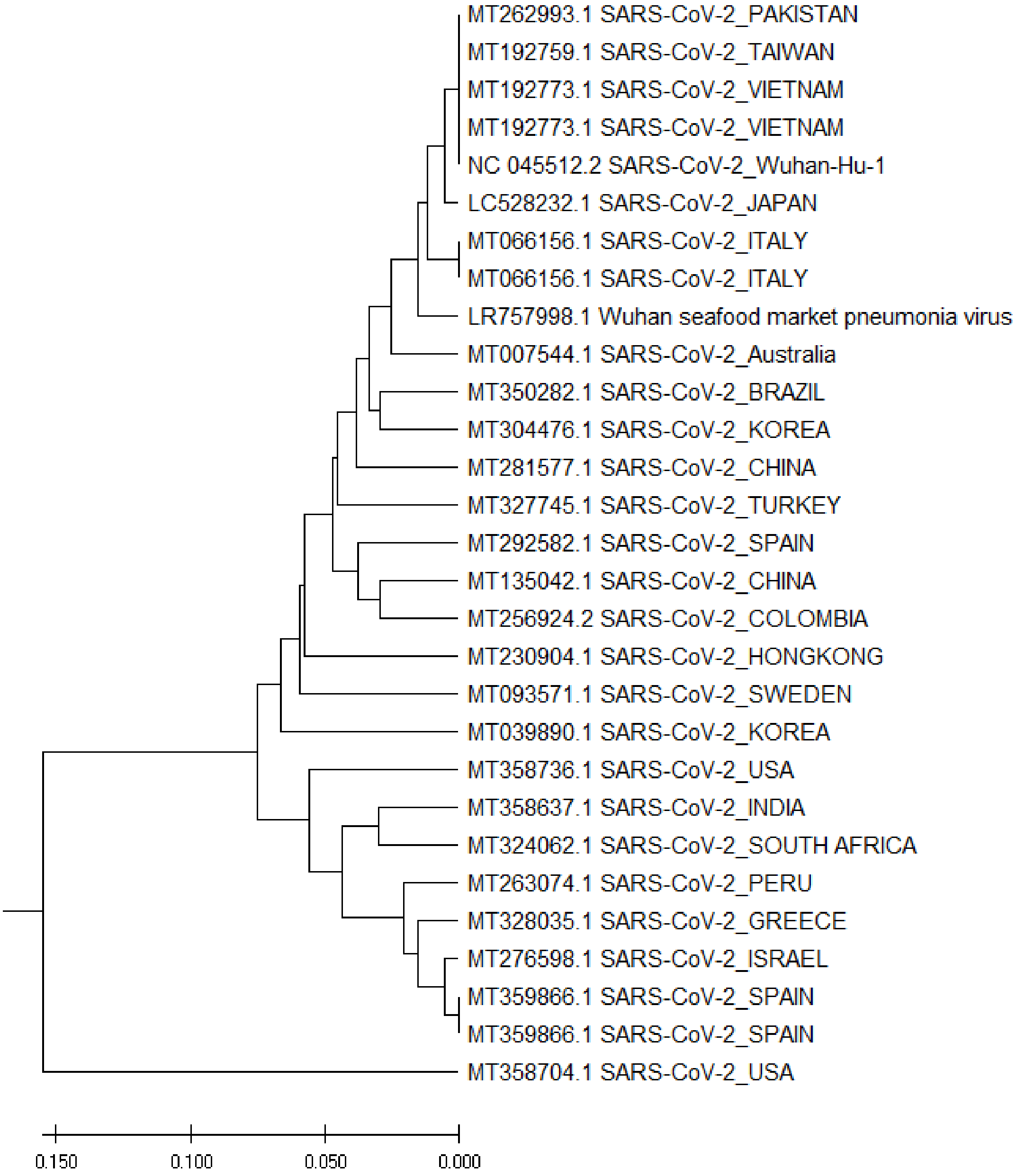

MT262993.1 SARS-CoV-2_PAKISTAN
MT192759.1 SARS-CoV-2_TAIWAN
MT192773.1 SARS-CoV-2_VIETNAM
MT192773.1 SARS-CoV-2_VIETNAM
NC 045512.2 SARS-CoV-2_Wuhan-Hu-1
LC528232.1 SARS-CoV-2_JAPAN
MT066156.1 SARS-CoV-2_ITALY
MT066156.1 SARS-CoV-2_ITALY
LR757998.1 Wuhan seafood market pneumonia virus
MT007544.1 SARS-CoV-2_Australia
MT350282.1 SARS-CoV-2_BRAZIL
MT304476.1 SARS-CoV-2_KOREA
MT281577.1 SARS-CoV-2_CHINA
MT327745.1 SARS-CoV-2_TURKEY
MT292582.1 SARS-CoV-2_SPAIN
MT135042.1 SARS-CoV-2_CHINA
MT256924.2 SARS-CoV-2_COLOMBIA
MT230904.1 SARS-CoV-2_HONGKONG
MT093571.1 SARS-CoV-2_SWEDEN
MT039890.1 SARS-CoV-2_KOREA
MT358736.1 SARS-CoV-2_USA
MT358637.1 SARS-CoV-2_INDIA
MT324062.1 SARS-CoV-2_SOUTH AFRICA
MT263074.1 SARS-CoV-2_PERU
MT328035.1 SARS-CoV-2_GREECE
MT276598.1 SARS-CoV-2_ISRAEL
MT359866.1 SARS-CoV-2_SPAIN
MT359866.1 SARS-CoV-2_SPAIN
MT358704.1 SARS-CoV-2_USA
0.150
0.100
0.050
0.000

Fig.2: Phylogenetic analysis of COVID-19 from 29 different individuals

Table 1: Sequences along with the accession number and region of the country

Sl. No	Accesion Number	Length	Region	Date of Release
1.	MT358637	29903 bp	India	20-APR-2020
	MT324062	29903 bp	South Africa	13-APR-2020
	MT358736	29901 bp	USA	27-MAY-2020
	MT358704	29901 bp	USA	27-MAY-2020
	MT263074	29856 bp	Peru	06-APR-2020
	MT359866	29893 bp	Spain	20-APR-2020
	MT359866	29893 bp	Spain	20-APR-2020
	MT276598	29870 bp	Israel	06-APR-2020
	MT328035	29903 bp	Greece	13-APR-2020
	MT350282	29903 bp	Brazil	17-APR-2020
	MT281577	29903 bp	China	06-APR-2020
	MT066156	29867 bp	Italy	13-APR-2020
	MT066156	29867 bp	Italy	13-APR-2020
	MT304476	29882 bp	South Korea	07-APR-2020
	MT093571	29886 bp	Sweden	06-APR-2020
	MT007544	29893 bp	Australia	11-FEB-2020
	MT039890	29903 bp	South Korea	11-FEB-2020
	MT135042	29903 bp	China	06-APR-2020
	NC_045512	29903 bp	China	30-MAR-2020
	LC528232	29902 bp	Japan	29-FEB-2020
	MT192773	29890 bp	Viet Nam	06-APR-2020
	MT192773	29890 bp	Viet Nam	06-APR-2020
	MT230904	29891 bp	Hong Kong	20-APR-2020
	MT327745	29832 bp	Turkey	13-APR-2020
	MT262993	29836 bp	Pakistan	06-APR-2020
	LR757998	29866 bp	China	06-MAR-2020
	MT192759	29862 bp	Taiwan	06-APR-2020
	MT292582	29782 bp	Spain	06-APR-2020
	MT256924	29782 bp	Colombia	13-APR-2020