Genomic analysis and comparative multiple sequences of SARS-CoV2

Abstract:-

Background: On December 31, 2019, China announced the outbreak of a new coronavirus in Wuhan, and the virus has since spread globally. The development of antiviral strategies depends on understanding the molecular mechanisms of genome selection we define the correlation of ten coronavirus (SARS-CoV2) sequences from various countries to analyze the genomic patterns of disease origin

Methods: We apply a Clustalw web service for Multiple sequence alignment (https://www.genome.jp/tools-bin/clustalw)and apply genomic analysis to sequence SARS-CoV2 from GenBank(https://www.ncbi.nlm.nih.gov/genbank/)

Results: Using (NCBI) database and genome alignment for protein sequencing found no differences in amino acid sequences between M and N proteins. The spike (S) protein has two amino acid variations.

INTRODUCTION:-

Infections such as (SARS-CoV)and(MERSCoV), which are spread from animals to humans, have caused pneumonia all over the world. SARS was first discovered in Guangdong, China, in 2002, and has since spread around the world, resulting in 8096 infected cases and 774 deaths. Then a coronavirus appeared in Wuhan, named SARS-CoV2 (International Committee on Taxonomy of Viruses). Coronaviruses are classified into four genera (α, β, ρ and δ). SARSCoV2 (2019) and SARS-CoV1 (2003) are belong to coronaviruses. Coronaviruses contain four proteins: The spike (S) protein helps the virus bind to the host cell's membrane, while the N (nucleocapsid) protein protects the virus's RNA genome. The E (envelope) and M (membrane).

the coronavirus start transmission *when attaches to host cell membrane receptors and enters the host cell. The virus genome's RNA gene 1 then starts to replicate. After that, the virus synthesises subgenomic RNAs with new transcription. and make a lot of copy of virus which attack human body and its Lung.(https://youtu.be/5DGwOJXSxqg)*

SARS-CoV1 is thought to spread from bats and civets to humans, where it causes extreme respiratory illness and a 10% mortality rate. Before being transmitted to humans, Wuhan SARS-coV2 is thought to be transmitted from a bat (Ra TG13).

The s protein is consist two subunit s1 and s2 on the s1 there is receptor binding domain is for ability of binding to human cell and effect him it bind with ACE2 which lung is produce (https://youtu.be/OVDaq_vOQ48)

in this paper, We compared SARS-CoV2 sequences from various countries to examine disease origin and evolution genomic providing genomic knowledge for the creation of new control methods against the worldwide SARS-CoV2 pandemic.

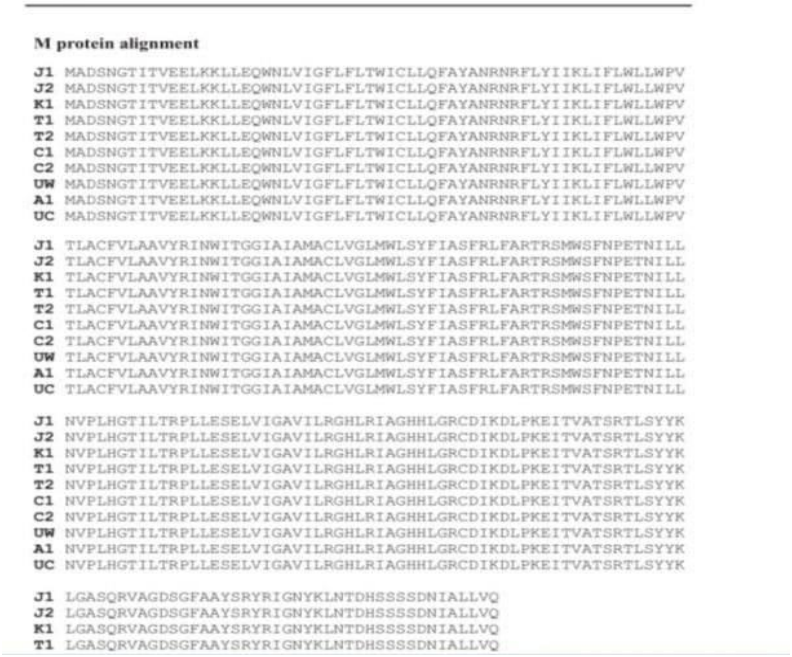
Related work:-

- *COVID-19 Genome Analysis using Alignment-Free Methods:-*
  The genome sequences for COVID-19 strains, taken from NCBI's GenBank, are compared and analyzed with AF methods for: (1) extracting frequent patterns of nucleotides in COVID-19 genome sequences, (2) finding the similarity/dissimilarity between COVID-19 genome sequences by using difference distance measure and (3) Phylogeny construction with various AF methods for COVID-19 genome sequence
- *Mutation hot spots in Spike protein of COVID-19:-*
  Multiple sequence alignments were done using alignment tool of NCBI virus server as well as CLUSTAL Omega. Sequence alignments from CLUSTAL Omega was viewed using MView tool. Multiple sequence Alignment of COVID-19 spike protein sequences from United States of America showed multiple mutations at few frequent locations
- *A lot of web service that we can use it to make alignment sequence as :-*
  MAFFT ---------→https://www.genome.jp/tools-bin/mafft
  PRRN----------→https://www.genome.jp/tools-bin/prrn

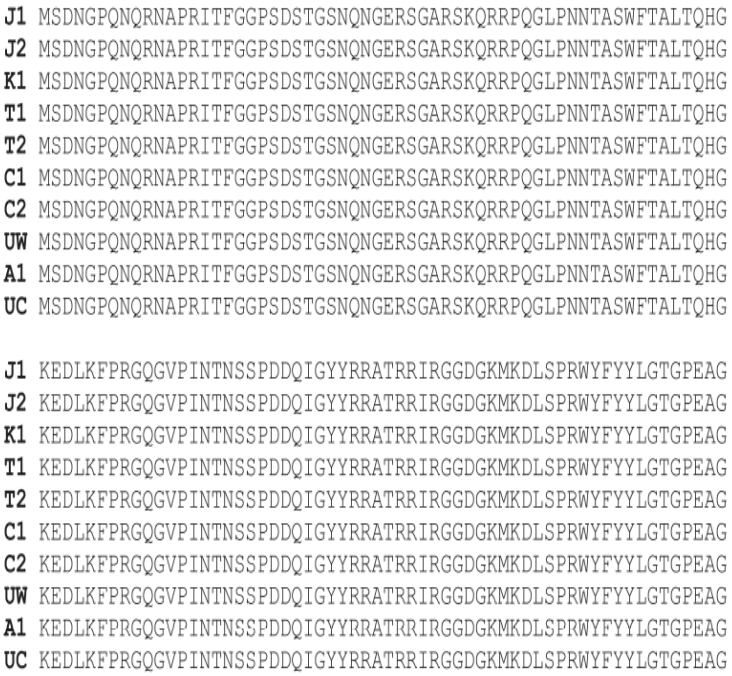
Methodology:-

. we make genomic analysis to discover covid 19 and focus on evolutionary and phylogenetic analyses have applied in disease. Using Clustalw web service to Multiple Sequence Alignment

We found one amino acid mutation at "H" from South Korea comparing the "L" from other nine sequences. Yellow line indicates the difference in 10 sequence alignments



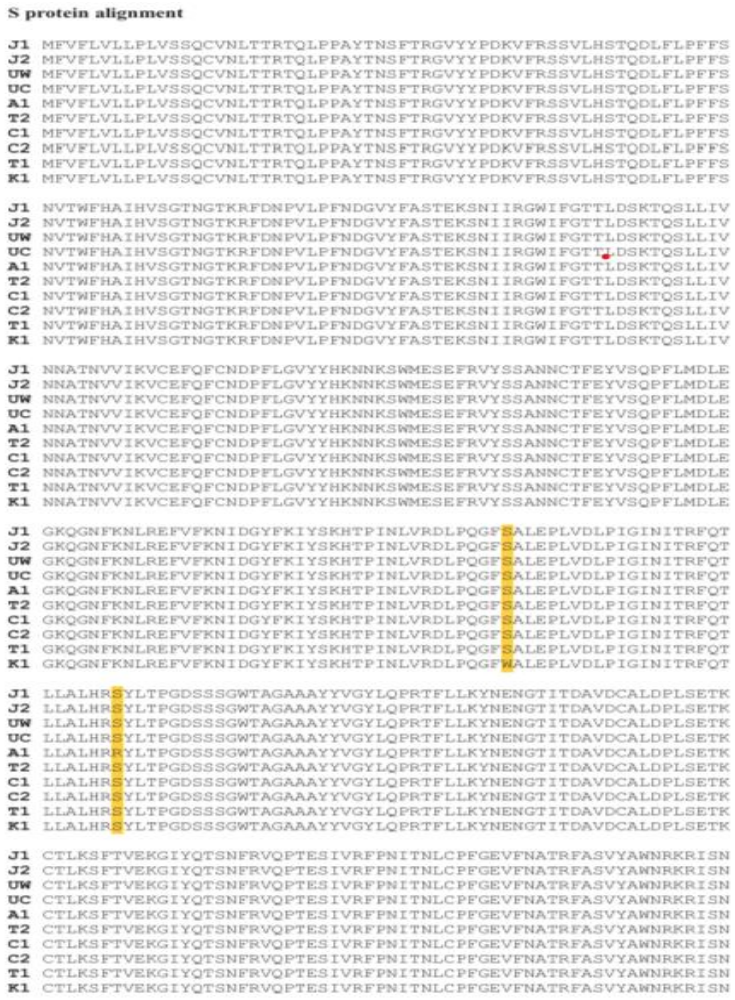
not found any mutation in 10 sequences of M protein region



Genomic analysis of N protein amino acid sequence.

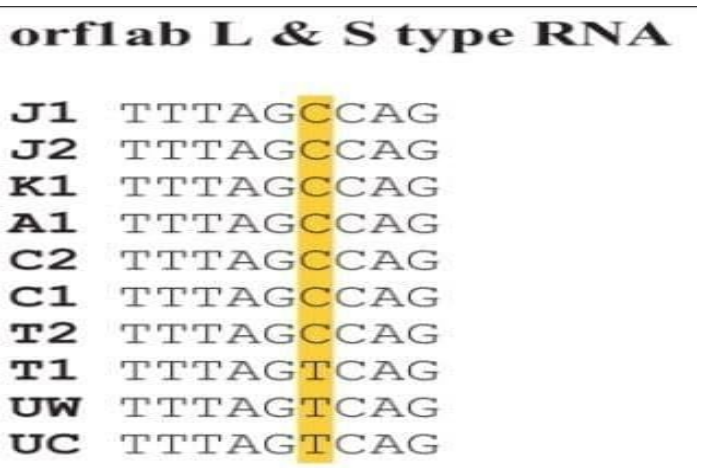
We do not observe any mutation in 10 sequences of N protein region

s



In s protein one amino acid mutation at "W" from South Korea comparing "S" in other nine sequences. One amino acid mutation at "R" from Australia was observed comparing "S" from another nine sequences. Two yellow lines indicate the difference in 10 sequence alignments.

Genomic analysis of ORF8 protein amino acid sequence. possible found in OFR8 with L and S subtypes



Genomic analysis of orf1ab protein amino acid sequence. subtypes were were found orf1b C and T.

Result:-

. E protein :-

E protein has a short and hydrophilic N-amino terminus consisting of 7–12 amino acids, followed by a large hydrophobic transmembrane domain of 25 amino acids, and ends with a long, hydrophilic C-carboxyl terminus and found one mutation in korea

S protein:-

S protein mediates attachment of SARS-CoV1 to the host cell surface receptors and facilitate viral entry into the host cell and by alignment found two mutation one in Australia and another in south korea

M and N proteins:-

The M protein is abundant which defines the shape of the viral envelope. N functions primarily to bind to RNA genome of SARS-Co2V, making up the nucleocapsid.15 Although N protein is most involved in processes viral genome signaling, it is also involved RNA replication cycle with host cellular response to viral infection ,by alignment not found any mutation in N and M protein

DISCUSSION:-

. point mutation:-

. we found point mutation in E protein and no mutation observed in M and N protein .in s protein found 2 mutation at “W” , “R” Report mentioned a single amino acid reversion (L294-to-Q) in the S protein is sufficient to abrogate the phenotype and grows well at and below 32ºC

. SNP or subtype:-

we found that SNPs at locations 8782 (orf1ab: T8517C, synonymous) and 28144 (ORF8: C251T, S84L) showed possible linkage in 10 sequences from different countries. As report “On the origin and continuing evolution of SARS-CoV-2” emphasized two subtypes of “L” and “S” from their data (defined as “L” type because T28144 is in the codon of leucine) and other “TC” haplotype (defined as “S” type because C28144 is in the codon of serine) at these two sites.

. conclusion:-

we analyzed 10 sequences from the NCBI data base by genome alignment and found no difference in amino acid sequences within M and N proteins. There are two amino acid variances in S protein region. One mutation found from South Korea sequence is verified. Two possible “L” and “S” SNPs found in ORF1ab and ORF8 regions are detected

.