doi.org/10.36721/PJPS.2021.34.4.REG.1305-1313.1

Genome wide association study and phylogenetic analysis of novel SARS-COV-2 virus among different countries

Syed Umair Ahmad[1], Muhammad Shehzad Khan[2], Zainab Jan[1], Nayab Khan[3], Asif Ali[4], Naumana Rehman[5], Mohsina Haq[6], Umama Khan[7], Zohaib Bashir[1*], Muhammad Tayyab[8], Ihteshamul Haq[9*], Shumaila Bakht[10] and Fazli Zahir[11]

[1]Department of Bioinformatics, Hazara University, Mansehra, Pakistan
[2]Department of Biophysics, City University of Hong Kong
[3]Department of Zoology, University of Balochistan, Quetta, Pakistan
[4]Institute of Pathology and Diagnostic Medicine, Khyber Medical University, Peshawar, Pakistan
[5]Department of Pathology, Khyber Medical College Peshawar, Pakistan
[6]Department of Pathology, Peshawar Medical College, Peshawar, Pakistan
[7]Department of Microbiology, University of Karachi, Karachi, Pakistan
[8]Institute of Biotechnology and Genetic Engineering, The University of Agriculture, Peshawar, Pakistan
[9]Deaprtmnt of Biotechnology and Genetic Engineering Hazara University, Mansehra, Pakistan
[10]Department of Human Nutrition, University of Agriculture, Peshawar, Pakistan
[11]Department of Allied Health Science Iqra National University, Peshawar, Pakistan

Abstract: Corona Virus (COVID-19) outbreak has threatened the world, since it has become pandemic and spread all over the world. The causative agent SARS-COV2 has proved lethal caused serious public health concern worldwide. Our aims were to describe the SARS-COV-2 genetic connections and check for recombination of all genome. The recombination was investigated by RDP5 and conflicting phylogenetic clustering in individual genomic fragments was established by phylogenetic study by maximum likelihood and Bayesian methods. Our analysis suggests that the available sequences from currently genomes of various strain were retrieved from different countries including Japan, French Republic, Spain, Peru, China, Vietnam, Taiwan, Brazil, U.S.A., South Korea, Sweden, Australia, Nepal, India, Iran, and Italy. The phylogeny of SARS-COV-2 observed the largest number of genome is Vietnam 29891-bp, while France is the smallest member identified with 29679-bp. Using Recombination Detection program5 (RDP5) the china strains was taken as parental strain but there were no recombination in the all strains. In our study we identified the mutation in Pakistani strains in high conserved region of Corona nucleoca super family domain at the nucleotide position (394: C replace with T, Position: 858: C replace with T and Position: 997 G replace A).

Keywords: SARS-COV-2, strains, multiple sequence alignment, phylogeny, recombination.

INTRODUCTION

Coronavirus are a wide family of viruses believed to cause diseases from common cold to more severe diseases. Viral infections tend to grow and pose a significant public health issue, according to the World Health Organization (WHO). In the past 20 years, multiple viral epidemics have been reported, for example in 2002 to 2003, severe acute respiratory syndrome coronavirus (SARS-CoV) and 2009 influenza H1N1. In Saudi Arabia the first identified in 2012 was Middle East respiratory Coronavirus Syndrome (MERS-CoV) (Lee *et al*., 2003; Wan *et al*., 2020; Lai S.T, 2005). A new outbreak in Wuhan, the largest metropolitan city in China's Hubei province, with mysterious low respiratory infections, was first reported to the WHO Country Office in China on 31 December 2019, which is currently in the picture. At a meeting on 30 January 2020, the WHO announced that, according to the International Health Regulation (IHR, 2005), the disease had expanded to 18 countries with 4 countries confirming human to human transmission. The PHEIC was reported to be of international concern (Cascella *et al*., 2020). On 11 February 2020, Dr Tedros Adhanom Ghebreyesus, WHO director general announced that this recent CoV outbreak had been caused by a "SARS-COV-2" acronym, "2019 coronavirus" Two more coronavirus epidemics have been reported over the last 20 years. The new virus is rather infectious and has spread exponentially worldwide. CoVs are the primary vectors for new outbreaks of respiratory diseases. For reasons which are yet to be clarified, these viruses can bypass species barriers and cause disease in men that range from common cold to more severe diseases like the MERS and SARS They are a large family of single-stranded RNA viruses (+ ssRNA) that could be isolated into various animal species (Perlman and Jason, 2009). Because of the appearance on the envelope of spike-glycoproteins coVs are positively beached RNA viruses that look crown-like in electrons microscopic form (coronam is the coronary term in Late).

*Corresponding author: e-mail: ahmadumair927@gmail.com

The Orthocoronavirinae of the Corona viridae (Nidovirus order) classification is classified into four genera: alpha-CoV, Betacoronavirus (beta CoV), DeltaCoV and Gamma-CoV. (Gamma CoV) of the Coronaviridae family are classified into 4 genera. In comparison, the genus BetaCoV breaks into five subgenres or sections. (Chan *et al*., 2013). Currently there are no approval drug are available but many vaccine and drug are under clinical trials (Ahmad *et al.,* 2021).

Genome-wide association studies for a number of specific diseases and quantitative features have also been performed, but do not relate the characteristics of Mendelian diseases. For some circumstances genomic areas that have major connections with the disorder were identified, leading to mutations that relate to susceptibility to diseases such as IBD, autism and type 1 diabetes. In some cases, genomic regions were found. However the linkage study for most specific diseases was limited and the genes can generally explain only a small portion of the overall disease's HERITABILITY (Hirschhorn & Mark, 2005).

Phylogenetic is a study of the evolutionary connectivity of different organism groups (e.g. organisms or populations), which shows morphological and molecular sequencing data, for example. Phylogenetic is enriching taxonomy, the study of naming and classifying species, although both disciplines remain distinct in empirical and conceptual terms. Our understanding of evolution and subsequently taxonomy have been changed by recent phylogenetic research of various molecular datasets. The new eukaryotic classification scheme based on current biology, biochemistry and molecular phylogenetic evidence has recently been suggested (Anning *et al*., 2019). In Italy the full length genomes sequence of SARS-Cov-2 (EPI_ISL_412973 and EPI_ISL_412974) strains were isolated from two dfferent peoples one patient were Chinese visitor and the other Italian patient. Both sample were sequenced and analyzed after virus cultivation (Stefanelli, Faggioni *et al*. 2020)

Aim of the current study is to observation of comparative genomics (variation and homology) of SARS-COV-2 virus genome from different countries and also to diversification constructs the evolutionary phylogeny tree and domain analysis in all selected countries.

MATERIALS AND METHODS

Data mining and retrieval

Full genome sequences data of SARS-COV-2 were collected from NCBI Database (http:/www.ncbi.nih.gov/) genome sequences of total 18 different countries: Japan, French Republic, Spain, Peru, China, Vietnam, Taiwan, Brazil, U.S.A., South Korea, Sweden, Australia, Nepal, India, Iran, Italy were retrieved (Iftikhar *et al*., 2021).

Analysis of structural domains

Our data mining methods were conducted and the regions were explored further. At the NCBI "Conserved Domain Database" (CDD), nucleotide sequences have been analyzed using default component recognition criteria (http:/www.ncbi.nih.gov/Struktur/cdd/cdd.shtml) SARS-COV-2 was the target domain for these elements. During the sequence scan, we took screen shots one by one for each of the results received from Conserved Domain Database (CDD). Before that, the sequences of these elements should be accompanied by the word count in order to further process highlighted domains.

Multiple sequence alignment and editing

The FASTA format sequences of SARS-COV-2 virus were imported into MEGA7 in text format. The MEGA7 is used for comparative and phylogenetic research study. Clustal W and Bioedit program were applied for alignment of whole genome of SARS-COV-2.

After the alignment, these sequences were visually checked and manually edited wherever possible. Deletions and tiny parts have been eliminated. Except where codons or framework-shift mutations have occurred, all SARS-COV-2 sequences have been used in the alignment.

Generation of Phylogenetic tree

The phylogenetic analysis was carried out by constructing a tree using the neighbor-joining approach with 500 replicates of the Bootstrap generated in MEGA7. Specific distance equations have been used respectively for determining and estimating the genetic distance between nucleotide and amino acid sequences.

RESULTS

Using the Repbase and NCBI databases, we retrieved the available sequence and unfinished genomes of SARS-COV-2 virus. Total 18 genome strains were identified from different countries including Japan, French Republic, Spain, Peru, China, Vietnam, Taiwan, Brazil, U.S.A, South Korea, Sweden, Australia, Nepal, India, Iran, Pakistan and Italy. This archive has been recovered with several SARS-COV-2 virus sequences which is represented in table 1. That has been showed with their country name strain and accession ID. The total Strains were aligned in ClustalW available in Mega program. The most conserved region of the SARS-COV-2 in all strains is highlighted as shown in fig 1.

The phylogenies of SARS-COV-2 virus were constructed by Neighbor Joining method with 500 bootstrap repetitions in MEGA7. The tree showed a clear separation

Table 1: List of SARS-COV-2 genome in various countries with their accession number and version

Element name	Location/Area	Size of elements (bp)	ORF	Accession number	Version
SARS-COV-2	China	29903	6	MN908947	MN908947.3
SARS-COV-2	Japan	29902	6	LC528232	LC528232.1
SARS-COV-2	USA	29882	6	MN985325	MN985325.1
SARS-COV-2	Australia	29893	6	MT007544	MT007544.1
SARS-COV-2	India	29854	6	MT012098	MT012098.1
SARS-COV-2	Italy	29867	6	MT066156	MT066156.1
SARS-COV-2	South Korea	29903	6	MT039890	MT039890.1
SARS-COV-2	Sweden	29886	6	MT093571	MT093571.1
SARS-COV-2	Nepal	29811	6	MT072688	MT072688.1
SARS-COV-2	Taiwan	29870	6	MT066175	MT066175.1
SARS-COV-2	Brazil	29876	6	MT126808	MT126808.1
SARS-COV-2	Vietnam	29891	6	MT192772	MT192772.1
SARS-COV-2	PAKISTAN	29836	6	MT240479	MT240479.1
SARS-COV-2	Spain	29782	6	MT233519	MT233519.1
SARS-COV-2	Peru	29856	6	MT263074	MT263074.1
SARS-COV-2	Israel	29851	6	MT276597	MT276597.1
SARS-COV-2	France	29679	6	MT320538	MT320538.1
SARS-COV-2	Iran	29828	6	MT320891	MT320891.1

Table 2: List of domain size and name of various countries.

Element name	Location/ Area	Size of elements (bp)	ORF	Accession number	version	Domain size	Domain name
SARS-COV-2	China	29903	6	MN908947	MN908947.3	1064	Corona_nucleoca super family
SARS-COV-2	Japan	29902	6	LC528232	LC528232.1	1064	Corona_nucleoca super family
SARS-COV-2	USA	29882	6	MN985325	MN985325.1	1064	Corona_nucleoca super family
SARS-COV-2	Australia	29893	6	MT007544	MT007544.1	1064	Corona_nucleoca super family
SARS-COV-2	India	29854	6	MT012098	MT012098.1	1064	Corona_nucleoca super family
SARS-COV-2	Italy	29867	6	MT066156	MT066156.1	1064	Corona_nucleoca super family
SARS-COV-2	South Korea	29903	6	MT039890	MT039890.1	1064	Corona_nucleoca super family
SARS-COV-2	Sweden	29886	6	MT093571	MT093571.1	1064	Corona_nucleoca super family
SARS-COV-2	Nepal	29811	6	MT072688	MT072688.1	1064	Corona_nucleoca super family
SARS-COV-2	Taiwan	29870	6	MT066175	MT066175.1	1064	Corona_nucleoca super family
SARS-COV-2	Brazil	29876	6	MT126808	MT126808.1	1064	Corona_nucleoca super family
SARS-COV-2	VietNam	29891	6	MT192772	MT192772.1	1064	Corona_nucleoca super family
SARS-COV-2	Pakistan	29836	6	MT240479	MT240479.1	1064	Corona_nucleoca super family
SARS-COV-2	Spain	29782	6	MT233519	MT233519.1	1064	Corona_nucleoca super family
SARS-COV-2	Peru	429856	6	MT263074	MT263074.1	1064	Corona_nucleoca super family
SARS-COV-2	Isreal	29851	6	MT276597	MT276597.1	1064	Corona_nucleoca super family
SARS-COV-2	France	29679	6	MT320538	MT320538.1	1064	Corona_nucleoca super family
SARS-COV-2	Iran	29828	6	MT320891	MT320891.1	1064	Corona_nucleoca super family

COVID19 family containing 50 elements in different countries, the complete size of elements with domain name and its size (table 2). The largest number shown is Vietnam (29891), while France is the smallest member identified from 29679.

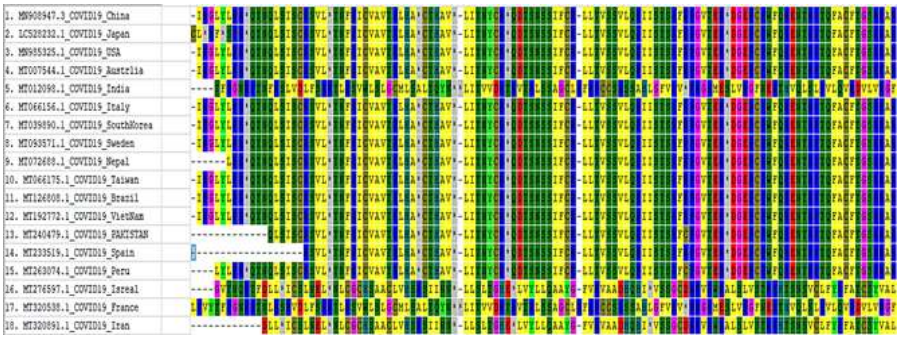

Fig. 1: Multiple alignment of SARS-COV-2. The amino acid sequences were aligned with ClustalW, Small insertions were deleted without altering the frame. The dashes represent gaps or ends of incomplete sequences and steric represent missing data.

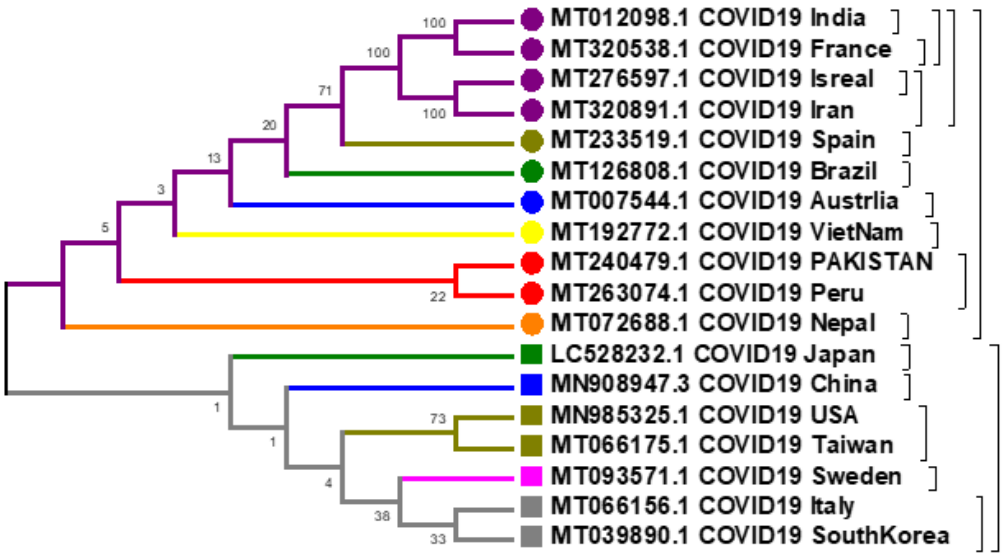

Fig. 2: Phylogenetic tree of *SARS-COV-2* sequences from China, France, India, Israel, Spain, Brazil, Australia, Vietnam, Pakistan, Peru, Nepal, Japan, USA, Taiwan, Sweden, Italy, and South Korea. The tree was constructed by Neighbor-Joining method with 500 bootstrap replicates using the MEGA7 program. The bootstrap support values (%) are shown near the nodes and values> 50% are not shown. The sequences clustered into 2 clades further dividing into sub-clades. The clade 1 represents eleven countries, where India, France, Israel and Spain is represented by purple filled circle; Spain is represented by brown filled circle; Brazil is shown in green, Australia in blue, Vietnam in yellow, Pakistan and Peru in red , and Nepal is represented by orange filled circle. The clade 2 represents seven countries, where Japan is represented by green, China in blue, USA in and Taiwan in brown, Sweden in pink, Italy and South Korea is represented by gray filled circle. The circle shapes are representing clade 1, while square shapes are representing clade 2. The *SARS-COV-2* are followed by the accession numbers given in NCBI database.

of two clades further dividing into sub-clades (fig. 2). Clades 1 consist of eleven countries and clade 2 consisting seven countries. In the first clade, India, France, Israel, Spain, Brazil, Australia, Vietnam, Pakistan, Peru and Nepal clustered together, while in the second clade, Japan, China, USA, Taiwan, Sweden, Italy, and South Korea. The India and France share a same group, where Israel makes a sister family with Iran. These respective families clustered together indicating similarities in sequences and thus a common ancestor. Another specific group is observed, where Pakistan makes a sister family with Peru. Of the USA and Taiwan clustered in their genera-specific groups, although Italy and South Korea showed close homology to Sweden suggesting their old history and common ancestry. The largest base pairs number shown is Vietnam which 29891 bp, while France is the smallest member base pair identified which is 29679 bp.

The SARS-COV-2 sequences were aligned in ClustalW available in BioEdit program. The most conserved region is vertically highlighted. The similarity among various sequences is much higher within the members of the same family as compared to other families.

Based on the sequences of SARS-COV-2 genome, the phylogeny were constructed by Neighbor Joining method with 500 bootstrap repetitions in MEGA7. The tree showed a clear separation of two clades further dividing into sub-clades shown in fig. 5. Clade 1 consist of fourteen countries and clade 2 consisting four countries. In the first clade, 6 sub-clades are Japan, France, Spain, Peru, China, and Vietnam clustered together, while in the second clade, 2 sub-clades presenting Taiwan, Brazil, USA, Sweden, Australia and Nepal family in one, India and Iran another sub-clade. The Japan and France share a same group, where Spain makes a sister family with Peru. These respective families clustered together indicating similarities in sequences and thus a common ancestor. Another specific group is observed, where Taiwan makes a sister family with Brazil. Although USA and Sweden showed close homology to Australia and Nepal of the Italy and Israel clustered in their genera-specific groups, while South Korea and Pakistan share another family.

To check out the recombination in whole sequences a Recombination Detection program5 (RDP5) was used in which the China strain was taken as the parental strain and dominant all over other countries as a result there is no any recombination in the SARS-COV-2 genome strains but its shows the similarities among the sequences in some context as shown in fig 6.

Fig 7 represent the Pakistan as a dominant its shows the changes similarity with china strain but no recombination in all available strains of corona_nucleoca. In Pakistani

strain we absorb the mutation in high conserved region accession ID MT240479.1 Pakistan. The nucleotide position (394: C replace with T, Position: 858: C replace with T and Position: 997 G replace A) noticed the mutation in Pakistani strain.

DISCUSSION

Coronaviruses SARS-COV-2 are a large family of viruses that cause diseases ranging from common colds to much more serious respiratory diseases, for instance Extreme Acute Respiratory Synchronization (SARS). COVID-19 is a novel illness that had not existed in people until December 2019. As such, the population is not immune, meaning that we are all vulnerable to infection and COVID-19 is not actually possible to spread effectively. Thus, there is a lack of immunity. SARS-COV-2 cause a number of animal diseases, while human infections are almost exclusively associated with RTI infections. The Coronavirinae substrates into 4 genera (alpha, beta,

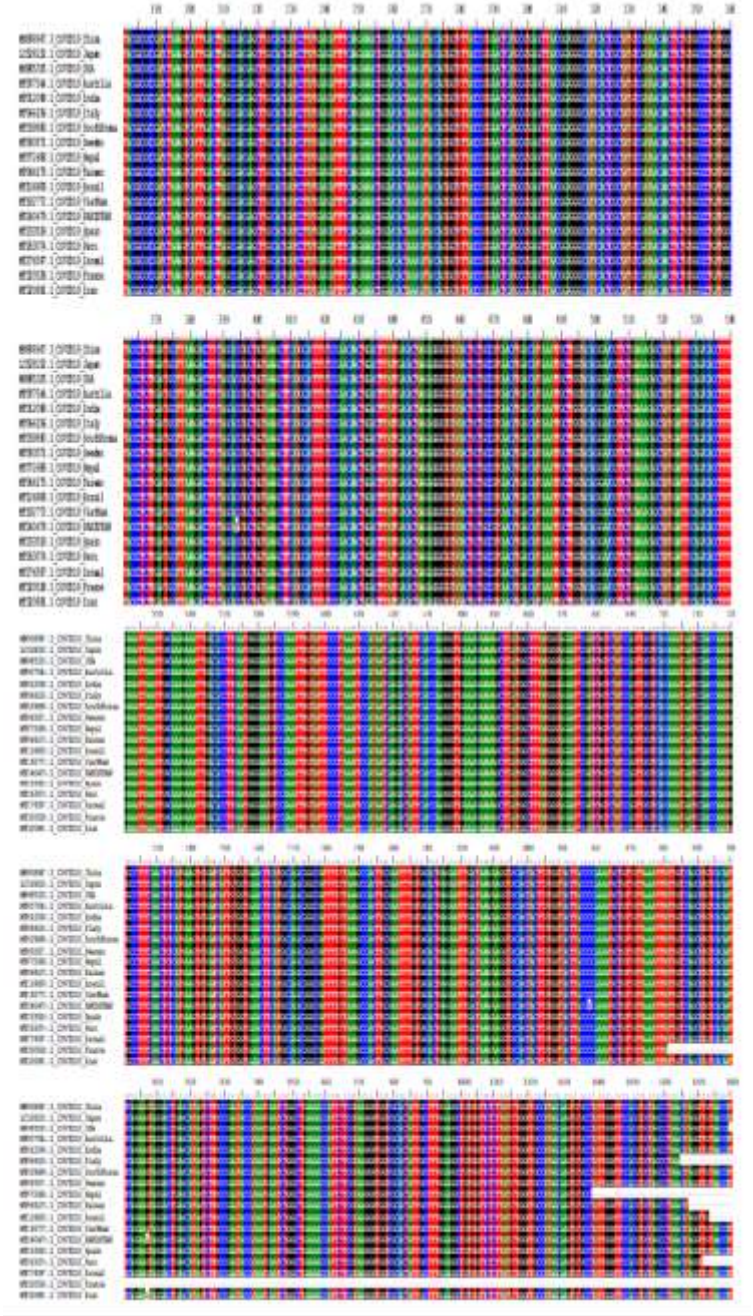

Fig. 3: Nucleotide alignments of conserved region from Corona_nucleoca superfamily. This region was retrieved from databases and aligned using Clustal W. Dashes indicate deletions; vertical colored lines indicating homology show conserved regions, while N donates the missing data. The names at left identify individual elements.

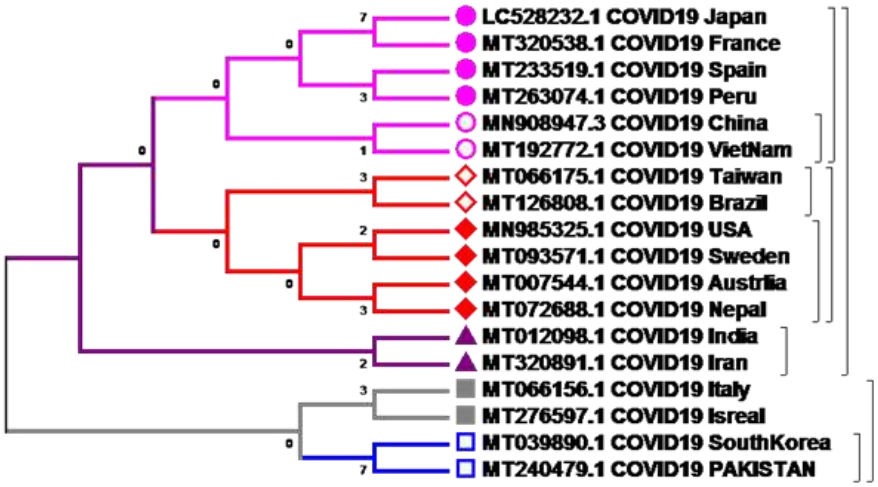

Fig. 4: Phylogenetic tree of SARS-COV-2 nucleoca superfamily sequences from Japan, France, Spain, Peru, China, Vietnam, Taiwan, Brazil, USA, Sweden, Australia, Nepal, India, Iran, Italy, Israel, South Korea, and Pakistan. The tree was constructed by Neighbor-Joining method with 500 bootstrap replicates using the MEGA7 program. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The analysis involved 18 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There was a total of 838 positions in the final dataset. The sequences clustered into 2 clades further dividing into sub-clades. The clade 1 represents fourteen COVID19 Corona_nucleoca superfamily, where Japan, France, SPAIN, Peru, China and Vietnam is represented by pink open/filled circle; Taiwan, Brazil, USA, Sweden, Australia and Nepal is represented by red open/filled diamond; India and Iran is shown in purple filled triangle. The clade 2 represents four COVID19 Corona_nucleoca superfamily, where Italy and Israel are represented by gray filled square; South Korea and Pakistan is represented by blue open. The square, diamond and triangle shapes are representing clade 1, while square shapes are representing clade 2. The names of the COVID19 are followed by the accession numbers given in NCBI database.

gamma & delta) contemporary taxonomy (King *et al*., 2020). In the 1960s, the affected alpha and beta species are the only species detected by humans (Hamre & John, 1966; Almedia & Tyrrell, 1967) and by the four species. Detected in the last 10 years: extreme coronavirus syndrome (SARS-CoV) (Drosten *et al*., 2003, Peiris *et al*., 2003). HCoV-NL63 (Hoek *et al*., 2004; Pyrc *et al*., 2007), HCoV-HKU1 (Woo *et al*., 2005) and Middle East Respiratory Syndrome (MERS-CoV) (Zaki *et al*., 2012). Their recorded species are just the alpha and beta species found in the 1960s. Coronavirus are the coronaviridae family enveloped viruses, which are found in the class of Nidovirus (Zuniga *et al*., 2010, Gorbalenya *et al*., 2014). They contain positive, single-stranded RNAs (ssRNAs), the largest known viral RNA genome, of around 30 kb. Their genomes are positive. 5'2/3 of genome RNA encodes the proteins in replicates. The 3 'section codes the structural and non-structural proteins in the genome. The computational function of the coronavirus nucleocapsid (N protein) is RNA Synthesis. The length and primary sequence of N proteins from various coronaviruses varies. Moreover, some biologically important reasons are also maintained and N proteins are sequence-like, shared by a three-domain association (Micintosh *et al*., 1967). A hierarchical arrangement with two standardized areas separated by a region of long disruptions was recently proposed for coronavirus N protein based on disorder predictions. RNA chapter one function has been identified as an average behavior of all coronavirus-nucleocapside proteins in extreme and acute respiratory syndrome (SARS- CoV) N proteins (Tyrrell & Bynoe, 1965). Many COVID-19 infected people suffer from mild to moderate respiratory diseases and recover without special treatment. Most likely to contract a debilitating illness is the elderly population and those with existing medical issues such as cardiovascular diseases, asthma, chronic respiratory disorder and cancer. If an affected person is flipping or sneezing, COVID-19 primarily spreads by the drips of saliva or nose discharge. Hence it is also critical that you practice respiratory etiquette (for instance by coughing into a flexed elbow). COVID-19 is characterized by fiver, dry cough, nausea, sputum, shortness of breath, myalgia and arthralgia, sore throat and headache. A limited number of patients (5 per cent) have been registered for nausea or vomiting.

Bioinformatics applied Insilco approaches to find out the location of gene, predict the transcripts of a particular gene and the structure and function of a specific protein transcribed from that gene inside cell and the disease(s)

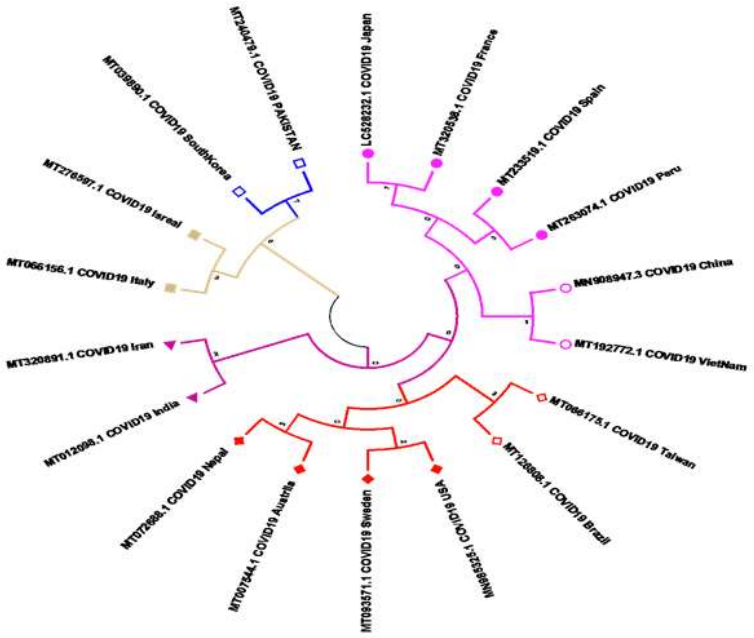

Fig. 5: Phylogenetic tree analysis of COVID-19 on basis of Corona_nucleoca superfamily.

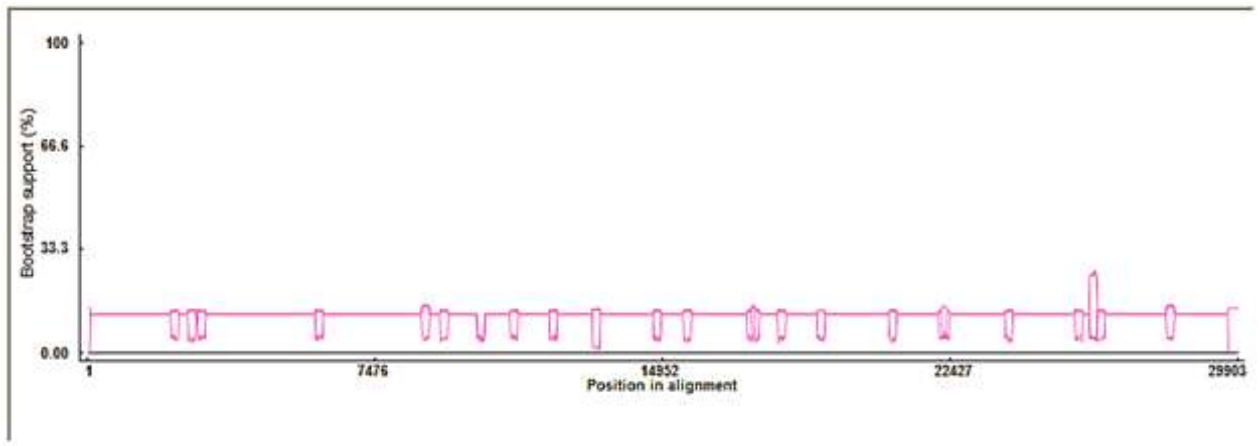

Fig. 6: RDP5 shows the result of SARS-COV-2 genome. When we placed the china as parental strain the no recombination were noticed. Red color indicates the similarity region among the remaining available strain of different countries.

associated with the abnormal structure or function of that particular protein (Ahmad *et al*. 2020; Ahmad *et al*., 2021; Jan *et al*., 2021)

In this research we used the computational method to find out the phylogenetic analysis of different strains of SARS-COV-2 from different countries. In previous study a phylogenetic network analysis were carried out a total 160 complete human respiratory syndrome SARS-COV-2 where they found three variants by amino acid changes (Forster, Forster *et al*. 2020). In our study the strain of Pakistan with accession ID MT240479.1 we identified the mutation in nucleotide position like C394T, C858T and G997A mutation in Pakistani strain. Our analysis included sequences SARS-COV-2 from China, Germany, Germany and India. Based on the findings of this study we collected sequences from Brazil, Australia, Vietnam, Pakistan, Peru, Nepal, Japan, the United Arab States, Taiwan, Sweden, Italy, and South Korea. This quest index allows one to identify and analyze the phase of evolution quickly. Aligning each other and creating a tree of each grouping is the superfamily of China, Germany, France, India, Israel, Brazil, Australia, Vietnam, Peru, Nepal, Japan, USA, Taiwan, Sweden, Italy and South Korea. The justification for doing so was to look at the gap between them. The findings show the points at the beginning of the measurement of superfamily, which may be triggered either by mutation or replication or removal, as seen in the effects of alignment. By using RDP5 program for detection of recombination in total retrieval Strains when

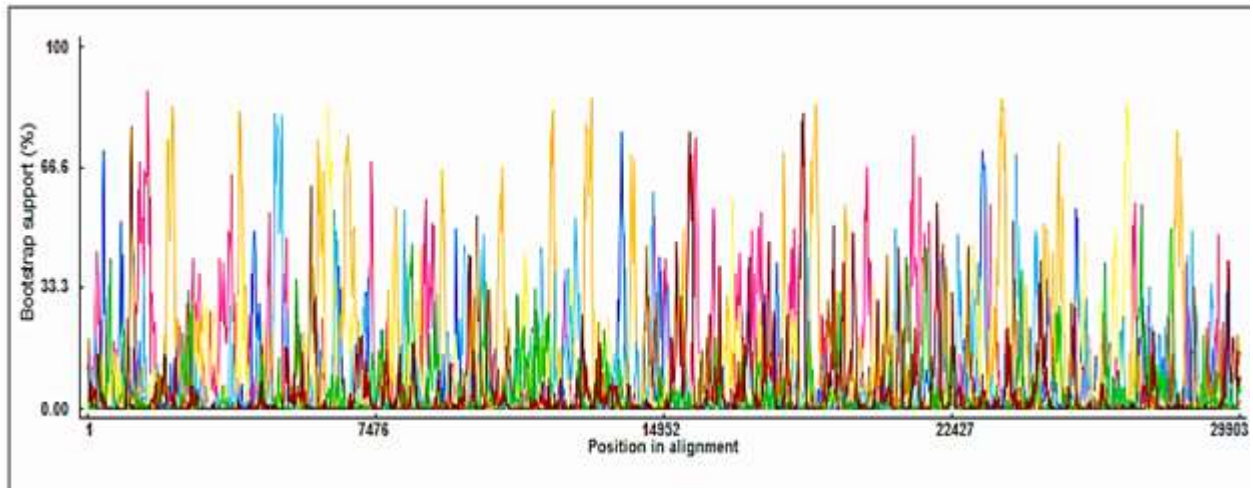

Fig. 7: RDP5 analysis Pakistan strain it also shows similarity with all other country strain and having no recombination is noticed in it. Different colors indicate the different countries similarity and replication

china was kept as parental strain as result there were no recombination were showed in all strains. The RDP5 represent the aligned and mutation region among the sequences. But when we take the Pakistan strain as a dominant or parental stain, shows the resemblance with china.

The current research is more helpful for scientist in the understanding of novel Corona virus genomic studies. This outcome can help in the discovery of drug and vaccine designing against the COVID-19. This research is done by computational method therefore the wet lab validation is needed to verify all these outcome results.

CONCLUSION

This research enhanced the awareness of COVID-19 arrangement, abundance, dissemination, mobilization and evolution. The features of the several countries are observed. The result enables the form and type of completed genome, their derivatives and their flowing genomic sequences and mutations understood in the various additions. Our analysis included sequences SARS-COV-2 virus from China, Germany, Germany and India. Based on the findings of this study we collected sequences from Brazil, Australia, Vietnam, Pakistan, Peru, Nepal, Japan, the United States of America, Taiwan, Sweden, Italy, and South Korea. This quest index allows one to identify and analyze the phase of evolution quickly. Aligning each other and creating a tree of each grouping is the superfamily of China, Germany, France, India, Israel, Brazil, Australia, Vietnam, Peru, Nepal, Japan, USA, Taiwan, Sweden, Italy and South Korea. It helps to annotate and classify many similar sequences and is used in the study of variation between closely related cultivars and varieties as molecular markers. The findings show the points at the beginning of the measurement of superfamily, which may be triggered either by mutation or replication or removal, as seen in the effects of alignment. All the available of COVID-19 strain no recombination is found only changed of nucleotide is noticed.

REFERENCES

Ahmad SU, Qadus A, Ahmad B, Khan JA, Shah ZW, Saeed A and Mahmood T (2020) 35. In-silco investigation of EGFR network in kidney cancer: A drug discovery approach. *Pure Appl. Biol. (PAB)* **9**(2): 1583-1595.

Almeida JD and Tyrrell DA (1967). The morphology of three previously uncharacterized human respiratory viruses that grow in organ culture. *J. Gen. Virol*, **1**(2): 175-178.

Anning AS, Kwakye-Nuako G, Ameyaw EO, Mosore MT and Asare KK (2019). *In vitro* activity of Erythrophleum ivorense extract against the promastigote stage of cutaneous Leishmania parasite, a member of Leishmania (Mundinia) enriettii complex isolates from Ghana. *Access Microbiology*, **1**(7): 1-8.

Antonio GE, Wong KT, Hui DS, Lee N, Yuen EH, Wu A and Ahuja AT (2003). Imaging of severe acute respiratory syndrome in Hong Kong. *Am. J. Roentgenol*, **181**(1): 11-17.

Bashir Z, Ahmad SU, Kiani BH, Jan Z, Khan N, Khan U and Mahmood T (2021). Immunoinformatics approaches to explore B and T cell epitope-based vaccine designing for SARS-CoV-2 virus. *Pak. J. Pharm. Sci.*, **34**(1): 345-352.

Bashir Z, Ullah B, Jan Z, Ahmad SU, Khan N, Bashir Z, Zafar I and Sajjad W (2021). *In-silico* studies of braf signaling network expression in colorectal cancer; A systematic approach for multi-targeted therapy. *Biosci. Res.*, **18**(1): 1015-1023.

Chan JFW, To KKW, Chen H and Yuen KY (2015). Cross-species transmission and emergence of novel viruses from birds. *Curr. Opin. Virol*, **10**(1): 63-69.

Drosten C, Günther S, Preiser W, Van Der Werf S, Brodt HR, Becker S and Doerr HW (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.*, **348**(20): 1967-1976.

Drummond AJ, Suchard MA, Xie D and Rambaut A (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol*. **29**(8): 19691973.

Forster P, Forster L, Renfrew C and Forster M. (2020). Phylogenetic network analysis of SARS-CoV-2 genomes. *Proc. Natl. Acad. Sci. USA*, **117**(17): 9241-9243.

Gorbalenya AE, Enjuanes L, Ziebuhr J and Snijder EJ (2006). Nidovirales: Evolving the largest RNA virus genome. *Virus Research*, **117**(1): 17-37.

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS and Manolio TA (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci*. USA, **106**(23): 9362-9367.

Hirschhorn JN and Daly MJ (2005). Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics*, **6**(2): 95-108.

Jan Z, Ahmad SU, Amara Qadus YA, Sajjad W, Rais F, Tanveer S and Haq I (2021). *In silico* structural and functional assessment of hypothetical protein L345_13461 from Ophiophagus hannah. *Pure Appl. Biol*. **10**(4): 1109-1118.

King AM, Adams MJ, Carstens EB and Lefkowitz EJ (2012). Virus taxonomy. *Virus Taxonomy*. doi: 10.1016/b978-012465330-6/50022-1.

Lai ST (2005). Treatment of severe acute respiratory syndrome. *Eur. J. Clin. Microbiol. Infect. Dis.*, **24**(9): 583-591.

MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E and Parkinson H (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.*, **45**(1): 896-901.

Marra MA, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YS and Roper RL (2003). The genome sequence of the SARS-associated coronavirus. *Science*, **300**(3): 1399-1405.

McIntosh K, Becker WB and Chanock RM (1967). Growth in suckling-mouse brain of "IBV-like" viruses from patients with upper respiratory tract disease. *Proc. Natl. Acad. Sci.* USA**, 58**(6): 2268-2273.

Medicine P (1962). (30734) Dorothy Hamre. (30734), pp. 190-193.

Perlman S and Netland J (2009). Coronaviruses post-SARS: Update on replication and pathogenesis. *Nature Reviews Microbiology*, **7**(6): 439-450.

Pyrc K, Berkhout B and van der Hoek L (2007). 'Identification of new human coronaviruses. *Expert Rev. Anti-infect. Ther*, **5**(2): 245-253.

Stefanelli P, Faggioni G, Presti AL, Fiore S, Marchi A, Benedetti E and ISS COVID-19 study group. (2020). Whole genome and phylogenetic analysis of two SARS-CoV-2 strains isolated in Italy in January and February 2020: additional clues on multiple introductions and further circulation in Europe. *Eurosurveillance*, **25**(13): 2000305.

Tyrrell DAJ and Bynoe ML (1965). Cultivation of a novel type of common-cold virus in organ cultures. *Br. Med. J*. **1**(5448): 1467-1470.

Wan Y, Shang J, Graham R, Baric RS and Li F (2020). Receptor recognition by the novel coronavirus from wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *J. Virol.*, **94**(7): 1-9.

Woo PC, Lau SK, Chu CM, Chan KH, Tsoi HW, Huang Y and Yuen KY (2005). Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J. Virol.*, **79**(2): 884-895.

Zafar I, Iftikhar R, Ahmad SU and Rather MA (2021). "Genome wide identification, phylogeny, and synteny analysis of sox gene family in common carp (*Cyprinus carpio*). *Biotechnol. Rep.*, **30**: e00607.

Zaki AM, Van Boheemen S, Bestebroer TM, Osterhaus AD and Fouchier RA (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.*, **367**(19): 1814-1820.

Zúñiga S, Cruz JL, Sola I, Mateos-Gómez PA, Palacio L, and Enjuanes L (2010). Coronavirus nucleocapsid protein facilitates template switching and is required for efficient transcription. *J. Virol*,, **84**(4): 2169-2175.

Van Der Hoek L, Pyrc K, Jebbink MF, Vermeulen-Oost W, Berkhout RJ, Wolthers KC and Berkhout B (2004). Identification of a new human coronavirus. *Nat. Med.*, **10**(4): 368-373.