Review Article

Genomic annotation of Human Coronavirus strains: A Review

Shivani Singh[1], Dr. Sharique Ahmad[2*], Dr. Saeeda Wasim[3], Dr. Silky Rai[4], Dr. Sudarshana Gogoi[5], Dr. Zarina Farheen[6]

[1-2]Department of Pathology, Era's Lucknow Medical College and Hospital, Era University, Lucknow, Uttar Pradesh-226003 India
[3]Nova IVF Fertility, Hazratganj, Lucknow, Uttar Pradesh-226001, India
[4-6]Junior Resident, Department of Pathology, Era's Lucknow Medical College and Hospital, Lucknow, Uttar Pradesh-226003 India

Abstract

Coronaviruses are the fascinating group of viruses with the unusual molecular mechanism of recombination and transcription and also consist newly emerged pathogens. The emergence of SARS and coronavirus as its etiological agent was found totally surprising in community of coronaviruses. As it was the first strain found to be causing severe illness in humans and also in moderate cases causes death. Till now there are found to be seven strains of coronaviruses affecting humans including newly found coronavirus. The NL63 strains and HKU1 strains both provide model for newly described identified human coronaviruses. Other strains OC43 MERS-CoV, 229E and SARS-CoV associated genome analysis showed identification and characterization of protease and replicase as well as various enzymatic activities of open reading frame ORF1a and 1b of these strains. All this provides us target for which potential drug can be evaluated. Various novel respiratory viruses including SARS-COV and NL63 were found in past 3 years. Then the discovery of another coronavirus HKU1 found in china. The other which had been recently emerged in Wuhan, China named as COVID-19. The genome of all strains of coronaviruses is described in this paper which will be helpful in decoding genome, diagnostics and prognosis of COVID-19.
Keywords: Corona Viruses, Open reading frame (ORF), Respiratory disease, Genome.

INTRODUCTION

Coronaviruses infects animals and humans also it had been identified greater than 50 years before and molecular mechanism associated with replication and several coronaviruses pathogenesis which have actively been studied in 1970. Its name "coronavirus," was termed in 1968 is named corona due to it crown like morphology observed in electron microscope.

The animal's coronaviruses can leads to diarrhea in cows, pigs and in chicken's upper respiratory tract disease. First coronavirus was isolated in 1937 from the bronchitis virus responsible for causing respiratory illness in chickens. Coronaviruses the broadest virus group belongs to order and Nidovirales and also contains Coronaviridae among 3 other Coronavirinae Coronaviridea and Torovirinae. Coronavirinae is further subdivided into alpha, beta, gamma and delta while this virus all 4 categories probable can be present in mammals. The source of gene for alpha and beta belonging coronaviruses are Bat and for gamma and delta associated coronaviruses gene source are avian. Among 100 of coronaviruses 7 are known to infect humans out 7 two are alpha coronaviruses 229E and NL63 and four are beta coronaviruses: OC43, HKU1, MERS, SARS-CoV. This virus classification is based on their phylogeny of which reflects how these viruses strains evolved from their common ancestors.

It reveals that whenever a new virus emerges it is classified accordingly and relates with the other known viruses and if it found to be distinct from all the existing species then it is called a new species. Alternatively it can classify according to serology (by immune system response to virus exposure) in III groups in which I and II refers to mammalian coronaviruses and III includes avian coronaviruses. 229E is classed in group I and includes largely animal pathogens. Group II consist veterinary pathogens and contains OC43, HKU1 and NL63. SARS coronaviruses classed in group II as well. Worldwide people mostly get infected with the 229E, HKU1, NL63 and OC43. Sometimes it can also happen that coronaviruses which infects animal gets evolve and become human coronaviruses includes MERS,SARS-CoV1 and SARS-CoV-2.

It is identified that there are more than one strains of coronavirus, this theory arises from China by analyzing different samples of coronavirus, it turned out that all coronavirus were not same. There were two types which were similar with minute difference in two places L and S types. Among both of them L type is most commonly seen and the mortality rate in L virulent is found to be very high.

229E
The first strain of coronaviruses to be described in mid 6os was extracted from the respiratory tract and this was established by Hamre and Procknow

OC43
It was discovered in 1965 which was investigated by Tyrrell and Bynoe they worked on nasal swab titled B814.

NL63 and HKU1
First found in Netherlands isolated from 7 months old infants consisting respiratory symptoms in year 2004. At that time there was research going on human coronaviruses led to the NL63 and HKU1 discovery, Hong Kong in early 2005.

SARS-CoV
In China 2003 animal was the source not yet clear, bats thought to have given it to other animals, civet cats.

MERS
This was found 2012 in Saudi Arabia transmitted by dromedary camels.

SARS-Cov-2
It had emerged as recently in year 2019 in Wuhan source of it is not yet known but possibly it could be bats. Until the SARS-CoV was not emerged human coronavirus were associated with only two prototypes HCoV- OC43 and HCoV-229E both of them consist same causative agent and primary symptom common cold. The recently identified Covid-19 is the first strains of coronavirus responsible for causing serious illness in human. The identification of SARS-CoV, revealed 2 new human coronaviruses which are were involved in respiratory disease. HKU1 virus was isolated from pneumonia patient there is not much information about this virus biology and the strain HCoV-NL63 of coronavirus was found in 7 month child who was suffering from disease bronchiolitis and conjunctivitis. The strong correlation was seen in strain NL63 with children containing lower respiratory infections and croup which revealed the virus and croup relationship. On the other side pathogenesis HCoV-229E, HCoV-OC43, HCoV-NL63, SARS-CoV is not known as much. The various studies of coronavirus associated with animal had contributed in the understanding the coronavirus associated with human.

Genome analysis of Human Coronaviruses
HCoV-NL63
This strain of coronavirus is a Coronaviridae family member with genus alpha virus and was first identified in 2004 [1]. It is particularly involved in common cold among immunocompromised patients, children and adults [2, 3]. Its genome size is approximately 27 kb with conserved order of genes, orf1ab-spike, orf3-envelope membrane, Nucleocapsid, poly (A) tail. The tropism of this strain is identified through the spike glycoprotein. The enzyme angiotensin converting enzyme 2 (ACE-2) is the receptor for both of the strains of coronavirus NL63 and SARS-CoV for taking entry in their host cell [4, 5]. HCoV-NL63 is identified commonly as an etiology of infection in upper portion of respiratory tract and to be prevalent on worldwide level.

The genomic analysis of HCoV-NL63 were done in the samples taken from nasopharyngeal as a swab from a children hospitalized with in China, 2018 with severe acute respiratory infection. The collected samples were filtered and RNA was isolated and utilized for single primer amplification in sequence independent pattern [6, 7]. Further reverse transcription were performed with primer containing fixed sequence and irregular hexamer at 3' end. After that Klenow fragment polymerase was used for DNA synthesis, further polymerase chain reaction (PCR) was performed using primer FR26. The DNA purified was utilized in sequencing through the method next-generation sequencing (NGS), the sequence read were associated the overlapping sequence of DNA or contigs with new assembly model. Which showed presence of gap region in HCoV-NL63 it was amplified with the specific primer designed for this purpose only. This amplified region is used for assembly of genome with the use of sequencing, 5 variant were found of NL63 by next-generation sequencing (NGS) and Sanger sequencing and named as GDO1GDO2, GDO3, GDO4, GDO5 (27,531 bp, 27,516 bp, 27,516 bp, 27,532 bp, 27,544bp) respectively

Multiple sequence alignment programs were used for alignment of these presented 5 strains of NL63 revealed homology of 98.5 % to 99.1% nucleotide with NL63 virus. Only 2 of the genome sequence of NL63 were found to be involved in acute respiratory illness reported in China. This complete information about the HCoV-NL63 genome provided the evolution and diversity HCoV-NL63 in China.

The genomic analysis of HCoV-NL63 revealed that RNA was found with typical coronavirus pattern and high quantity of uracil (39%) and lesser quantity of guanine (20%) and cytosine (14%). Infact among coronaviridae NL63 was the most pronounced nucleotide bias. The significant fluctuation was seen in

nucleotide counts among HCoV-NL63 genes, occurrence of ORF-3 and membrane protein extremely rich in uracil. The distinctive composition of nucleotide of few structural genes represents evolutionary origin, demonstrating that few of these functions were newly adapted through other cellular or viral origin by the process of gene transfer. These entire properties mimic the genome pathogenicity of prokaryotes [7]. The

biphasic pattern were observed while investigating genome and nucleotide composition 2/3 of the 5' genome encodes the ORF1a and ORF 1b with stable nucleotide count and with order of uracil>adenine>guanine but the noticeable differences was analyzed in 3' end in 1/3 of genome which encodes proteins involve in structural aspects [Fig. 1].
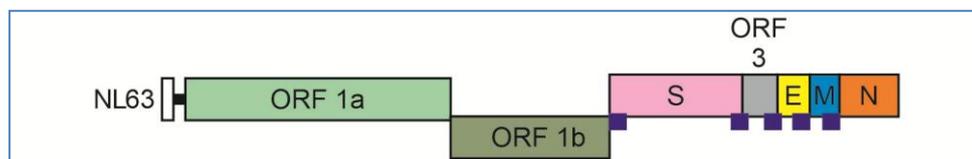

Figure 1

NL63- 5' end consist ORF1a/1b which encodes replicase polyprotein responsible for performing replication and increasing its copy number in host cell, 3' end contains structural protein S- spike, accessory proteins ORF-3 encodes responsible for pathogenicity in host cells, E-envelope, M- membrane, N-nucleoprotein.

Particularly the cytosine content enhances at the guanine and uracil expense. Grigoriev reported the bias of coronaviral genome and switch of nucleotide in 2/3 part of the genome, study was conducted for analysis of guanine and cytosine revealed that there was drop in the of guanine to cytosine ratio due to decrease count of guanine [8]. The investigation of other coronaviral genome insures that cytosine is elevated in one group and guanine decrease in other group. Therefore another mechanism which was known for transcription strategy associated with genome of virus the proportion of 1a/1b of virus positive strand is slightly prone towards the complementary strand annealing during replication due to sub-genomic location of negatively stranded RNA though it is found to be deficient of domain 1a/1b. Genome 1a/1b part is found to be prone towards the risk of cytosine to uracil deamination with correlation of higher uracil count and lower cytosine count. As there are various others cellular conditions and viral properties like secondary structure on 3' part of the genome which could design the genome of coronavirus on evolutionary time scale. This explains the switching of nucleotide in 2/3 part of viral genomes, uracil reaches the elevated level of counts and cytosine at the third position of NL63 counts lowest. The codon usage of the strain NL63 gets affected by genome composition which results either direct or indirect effect on amino acid composition of the coronaviral protein mainly by the non-synonymous codon usage effect [9-11]. Investigation of the viral genome sequence helps for predicting that the ORF1a and ORF1b polyprotein is found to be expressed in the genomic RNA and 3' structural proteins and ORF3 from 5' different subgenomic mRNA.

The viral genome associated investigation leads to predict that ORF1a and ORF1b is expressed from the RNA with 3' structure protein ORF3 from 5' different subgenomic mRNAs. It was observed that subgenomic mRNA are found in abundant amount when correspond to trinucleotide repeat sequences (TRS) is situated at 3' end of genome the exception was found in envelope subgenomic mRNA. A mechanism explained that polymerase of virus has reduced density along genome and the polymerase becomes less prone to perform transmission of trinucleotide repeat sequences (TRS) to long terminal repeats (LTRS) at extended strand synthesis [12].

HCoV-OC43

The HCoV-OC43 was extracted in 1967 and was initially generated in cultures of embryonic and nasal organ in humans [13]. OC43 is associated in all types of common cold at 10-30% mainly in winter and initial time of spring season the incubation period for this strain is 2 to 4 days. As coronavirus can infects everyone with any age its re-infection are frequently seen the infection may be subclinical and found commonly mild but in many cases it become severe mainly in adults and infants lower respiratory tract [14, 15]. A role of HCoV-OC43 as an etiological factor is found in multiple sclerosis is being studied [16-18].

The analysis of HCoV-OC43 genome of 32 kb revealed that the complete genome sequence which was extracted from an adult in 1967 suffering with common cold like symptoms, consists 30,738 nucleotides with the exclusion of polyadenalyted tail at 3' end and consist guanine, cytosine content in 36.9% in their genome. HCoV-OC43 genome contains 11 prominent ORF (open reading frames) which are flanked in UTR (untranslated regions) 5' and 3' consisting 211, 288 nucleotides. The 2 replicase ORFs which were larger 1a and 1b were also seen in two third of genomes. Both ORF 1a and 1b overlap each other as 1a consist of 4,383 codons and overlaps with 2,721 codons consisting ORF1b, replicase polyprotein enzyme of coronavirus contains 7,095 amino acids which synthesized through translation of multiple protein from

single mRNA (1 ribosomal shift) at UUUAAAC which is conserved site with 13335 to 13341 nucleotide. Also consist hemagglutinin-esterase (HE) helping viruses in fusion with membrane and attachment with accessory

protein non-structural protein -2 (nsp2). The 3' proximal parts of OC43 genome consist of various ORFs which codes several structural and accessory proteins [Fig. 2].
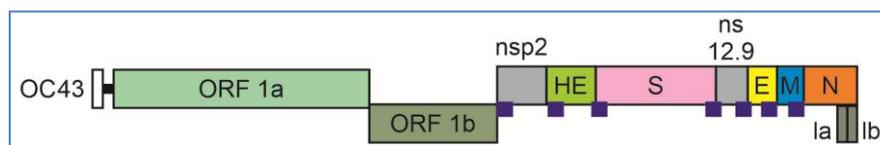

Fig-2

OC43 – 5' end consist ORF1a and ORF 1b encoding proteins for performing replication, non-structural protein 2, hemagglutinin-esterase (HE) involves attachment and fusion with membrane, at 3' end structural protein are present spike, E envelope, M membrane, N nucleoprotein, ns 12.9 in between spike and envelope protein encodes protein of 12.9 kDa.

The HCoV-OC43 prototype which is a laboratory strain that has been extracted in 1997 and been subculture from embryonic tracheal organ 7 times of human culture 15 times from mouse brain cells and unknown times from rectal tumor HRT-18 cells in humans at the time of subculture number of mutation accumulates. The analysis of entire nucleotide sequence of HCoV-OC43 strains which was not associated with mutation resulting in nonfunctional proteins.

For this virus a comparative study was done through amino acid analyses for ORF1a, ORF1b, spike, membrane, nucleoprotein on 58 sequences strains for which complete genomes were available on 76 sequences of spike proteins only. For conducting this analysis the new sequence added were compared with phylogenetic analysis. Each sequence position is needed to be examining separately. By doing this examination there will be low chance of erroneous mutations selection. The aim was to ensure that the informative mutations observed are characteristics of the specific genotype. For both of the cases the main amino acid substitution was identified by the selection of informative sites, for highlighting the conserved sites manually colorization was done [19].

It was suggested that this virus shows genetic and antigenic similarity with Bovine coronavirus remarkable. The complete analysis of genome and ORFs were performed in comparison of BCoV. It shows deletions of 290 nucleotides in between the region of spike and membrane protein in comparison of BCoV ORFs, nucleotides and amino acid similarity were identified HCoV-OC43 ORFs and other groups of coronaviruses. The highest similarity was identified between the ORF of OC43 and BCoV with the envelope gene exception. The spike gene sequence associated molecular clock analysis shows the zoonotic transmission of both BCoV and OC43. In order of gaining awareness for the process associated with non-

human to human adaptation of coronavirus this is first pair of animal to human coronavirus transmission as observed. This is crucial for knowing the origin and transmission events which is leading to severe acute respiratory syndrome and its outbreak. All of this reveals that HCoV-OC43 provides the viral promiscuity and also the insight of non-human coronavirus to human host adaption. This plays important role for understanding the interspecies transmission events causing SARS outbreak. So this HCoV-OC43 can be a powerful weapon for SARS coronavirus transmission associated understanding.

Middle East Respiratory syndrome Coronavirus (MERS-CoV)

The strain was initially found in Saudi Arabia in 2012 and through these record further cases recorded in United Kingdom [20-23]. Saudi Arabia is the place where people were most heavily get impacted with the MERS-CoV by 38% of deaths in 64 cases and almost 50% of fatality rate [24]. The genome of MERS-CoV is compared with Coronavirus sequences for performing this Bioinformatics is a useful tool for analyzing the protein, new gene or whole genome. Also annotations and analysis of sequence are the helpful in finding new function, drug targets and infectious disease recently emerged associated observation [26-30]. For comparing MERS-CoV with various other coronavirus requires only low to medium identity are evident. The 4 middle east isolates were taken from Al-Hasa as a reference, conclusion were made that it consist 46% homology with SARS coronavirus genome sequence, it shares low identity with the animal origins isolates of CoV. The symptoms of MERS-CoV are respiratory distress fever and renal failure in few of the cases. These symptoms are similar to SARS-CoV but molecular mechanism reveals that both of these strains of coronaviruses were not same. MERS-CoV does not use the receptor of SARS-CoV for entering in the host cells. Further the binding mechanism of MERS-CoV with the receptor is different from SARS CoV [31]. The phylogenetic maps revealed that MERS-CoV and SARS-CoV both shares 40% of their sequence at least. Its contains single stranded genome of 30.1 kb which encodes proteins, replicase, open reading frame 1a and 1b and structural proteins nucleoprotein, membrane protein and spike surface protein also 5 non –structural ORFs 3, 4a, 4b, and 5 proteins [32-33] [Fig. 3].
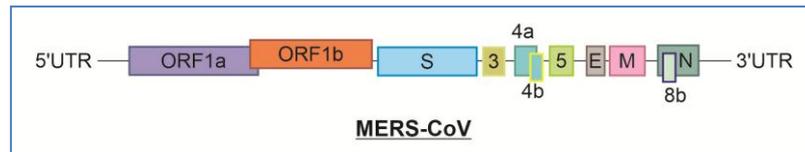
Fig-3

The strain MERS-CoV genome consist single stranded RNA genome encoding two large ORFs 1a and 1b, spike, envelope, membrane and Nucleocapsid structural protein encoded at 3'end of the strand which is commonly found in all the strains of coronaviruses, and also they encodes an accessory protein 4a/4b 5 at 3' end 4a is responsible for suppressing the antiviral response.

Sample (sputum) from the patient was collected on May 2015 and the MERS strain was inoculated 3 times to Vero cells and passed to viral culture solution 3 times and then, reverse transcription was done with reverse primers, further cDNA was amplified through primer of PCR which were overlapping. All these procedure for genome analysis results in total 2,814,805 sequences reads of MERS-CoV and 2,617,936 reads were found to be consensus sequence [34]. The whole genome sequence was obtained finally through the mapping result with 3,605.95 average coverage. The size of genome based on assembly was found approximately 30,108 bp with 41.15% guanine to cytosine content. Tamura et al performed a sequence analysis in 53 complete human MERS- CoV genomes of South Korean [35]. The entire

MERS-CoV/KOR/KNIH/002_05_2015 genome revealed 99.5% to 99.8% overall nucleotide identities with 53 human MERS-CoV. This observation shows Korean MERS-CoV consist 29 nucleotides and 12 variant of amino acid in comparison of MERS-CoV associated 53 full sequence. Variation in N-terminal domain was observed Arg 137 Ser and other at Leu530Val at receptor-binding domain. Its spike protein mediates entry in host and observed only in MERS-CoV/KOR/KNIH/002_05_2015 cell culture in comparison of other studies of spike protein mediated receptor binding [36].

Human coronavirus 229E

Generally this strain causes common cold but can also be involved with severe respiratory infections in persons who are already been dealing with any kind of disease [37-38]. Its genome is positively stranded RNA of $6X\ 10^6$ nucleotide of 7 Kb approximately extending from 3'end of genome has been known. It codes structural protein nucleoprotein, membrane and spike protein and OR4a, 4b and 5 in between spike and membrane protein also ORF1a, 1b are encoded in this strain [Fig. 4].
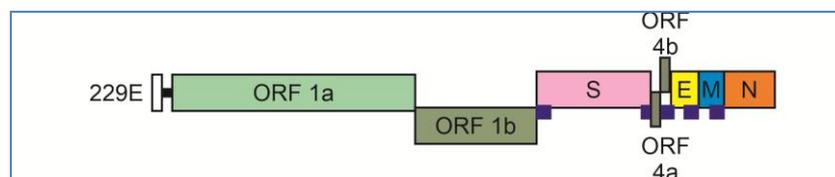

Fig-4

5' end codes for replication associated ORFs ORF1a and ORF1b and ORF 4b, 4a in between spike and envelope protein act as viroporin, and at 3' end structural protein are coded.

Till now full genome of HCoV-229E had been determined and this reference had been obtained from the infectious clone of cDNA which was based on laboratory-adapted prototype of 229E deposited in 1973 [39]. This prototype was extracted by a medical student with upper respiratory infection in Chicago University [40, 41] Chibu *et al.* have investigated the HCoV-229E evolution by spike and nucleoprotein associated sequencing obtained by the clinical samples[42].

The entire genome of HCoV-229E prototype 0349 and J0304 contains approximately 27.240 nucleotide. The proportion of guanine and cytosine are 38.07 % and 38.13 % in 0349 and J0304 respectively. The HCoV-229E strains have 7 putative protein coding

genes, 8 in reference sequence. This difference is due to the fact that clinical strains have ORF-4 and reference has ORF4A and 4B. The full genome alignment of the strains of HCoV-229E with reference sequence revealed substitutions of 168 in 0349 strain and 175 in J0304, in replicase 1a gene 66 non-synonymous are present in both of the strains of HCoV-229E. In the 1b replicase gene substitution of 20 and 19 amino acid are present and also 1 deletion and an insertion were seen. In strain 0349 and J0304 74 nucleotide substitution were seen in which 34 non-synonymous, 21 nucleotide insertions at 286 position and 6 nucleotide deletion in the 316-321 position were found. The gene NSP4 gene (Endoplasmic reticulum associated gene) is present in between two autoproteolytical proteins NSP3 and 5, 21 nucleotide substitutions in strain 0349 results in changes of 5 amino acid. Non-structural protein (NSP5) function in proteolytic activity of cysteine which is important for replication and named as main protease (Mpro)[43].

Mpro pathway is conserved in all the strains of coronaviruses and it cleaves numerous sites for producing 13 mature proteins [44]. While at the nucleotide level there were only 7 changes and 0349 prototype only 1 amino acid difference were present, other prototype J0304 consist 10 substitutions which results only 2 alteration in amino acid level. Gene NSP6 encodes transmembrane protein, 8 substitution of nucleotide were seen in 0349 and 7 in J0304, among which 1 was non- synonymous [45]. In NSP7 gene only 1 substitution were seen and which do not have any type of consequences on encoded protein, 6 nucleotides were found to replace in NSP8 in which except one all are synonymous . NSP9 consist 5 nucleic acid substitution and NSP10 6 nucleic acid substitution in which two leads to change the amino acid sequence. NSP 12 codes RNA dependent polymerase contains amino acid 927 in which 35 nucleotide substitutions is observed which results to 12 amino acid changes in comparison to reference sequence. However, the cleavage site of both NSP10 and NSP12 are different. NSP13 gene encodes a viral helicase protein in which 19 nucleotide substitutions were observed among them 3 leads to the changes at amino acid level same pattern is seen in NSP14 which encodes exoribonuclease leads to 17 nucleotide substitutions. In NSP14, 13 mutations were seen causing changes in amino acid and NSP15 and 16 consist same number of nucleotide substitutions 7 nucleotides.

The spike protein associated S gene contains 2 domains S1 and S2 no cleavage site for Furin enzyme were present in between these domain as seen in the reference prototype. Spike protein gene alignment revealed 2 deletions and 117 nucleotides substitution in both of the clinical strains 0349 and J0304. The change in amino acid and deletions are seen in S1 domain, the 0349 associated deletion leads to S1 two codon deletions at 228 and 354 positions of amino acid. On other side J0304 results to 3 codon deletion in S1 at the position 228, 354, and 355 of amino acid. The changes in S1 domain associated amino acid results in escaping virus from neutralizing antibodies. The HCoV-229E is basically a aminopeptidase N of human which is identified by S1 domain. The amino acid between 417 and 517 is an important position for receptor binding [46]. Aminopeptidase N binding is also found prominently for the area between 278 and 329 amino acids this was revealed in an study.[47] The envelope protein consist of 77 amino acid it has only 2 nucleotides difference in comparison of reference and in which 1 results to change in amino acid. Membrane gene consist 678 nucleotides and 225amino acid with 10 nucleotide difference observed in clinical isolates in

comparison of reference sequence. The protein of nucleocapsid plays a important function in RNA synthesis and assembly of virus [48]. 1170 nucleotide encodes the nucleoprotein protein with 26 nucleotides and the prototype 0349 and J0304 consist 7 changes in amino acid which were not in hotspot area between amino acid 224 and 228 amino acid. The nucleoprotein gene sequence present in Genbank shown changes in 6-7 amino acid found in circulating HCoV-229E strains and it consist 224-228 hot spot. All of this reported the alignment of HCoV-229E genome sequence, nucleotide and protein sequences which also shows various difference and genetic drift in the spike gene.

Severe acute respiratory syndrome coronavirus (SARS-CoV)

This strain of coronavirus is a single stranded positive RNA virus its virus profile is similar physically and genotypically with the other identified strains [49-51]. Its genome consist of 5 proteins replicase , spike, envelope, membrane and nucleoplasid, were aligned homologically with other coronaviruses. The other unknown structure and function of proteins were previously named as putative uncharacterized proteins (PUPs) but few of the PUPs were matched with NCBI database. Initially annotations of SARS-CoV genome revealed that by combining various prediction gene and the use of program predicting multiple genes in genome (FGENSV) 134 open reading frame were predicted F1 – F14 and among all of them 2 were newly formed F14 were found towards minus end of the strand. The genes which were already known beside them 9 ORF were predicted called G-G9 by the bioinformatics tool Gene Locator and Interpolated Markov ModelER (GLIMMER) it is basically involved in finding gens in prokaryotes, viruses. The other 9 genes were recognized through BGFV tool of bioinformatics named B1-B9, apart from this prediction manual prediction provides 5 more ORFs were seen as candidate's genes, whose chromosomal location is associated with specific disease. Each candidate genes at its upstream region consists trinucleotide repeat sequence (TRS) which have translated sequence of longer the 40 amino acids. All of these ORFs identified were present uniformly according to its initial location in the genome and also its physiochemical features were predicted very well

The SARS-CoV genome consist approximately about 11 open reading frame and 10 intergenic regions (IGRs) with positive single stranded RNA which is of 30 Kb approximately. The primary 2 ORFs were found to be overlapped at the 5'end and encompases 2/3 genome part while remaining ORFs at 3' end 7a, 8a, 9b [Fig. 5].
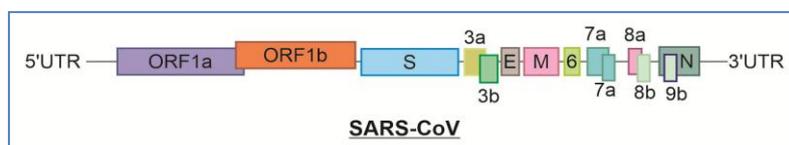

Fig-5

SARS-CoV- This strain consist 11 ORFs two polyprotein ORF1a and 1b overlapping at 5' end, the genome of SARS-CoV is smaller than MERS-CoV. It consist polA tail at 3' end, Spike, envelope, membrane, Nucleocapsid structural protein with 8 accessory proteins ORF 3a, 3b, 7a, 7b, 8a, 8b, 9b at 3' end, 3a function in apoptosis induction and 3b, ORF6, ORF7a signaling inhibition, 7b,8a,9b function are not known till now.

The genome analysis revealed variation in comparatively 38 isolates which results in the conclusion that all the isolates of SARS-CoV were homologous. Most of the isolates consists similar structure of nucleotide with similar 5' and 3' ends but polyA tail were found to be of different length at 3' end. Among 312 single nucleotide polymorphism (SNPs) 2 was only present in intergenic regions (IGRs). In 38 isolates three consist long insertions within the sequence and 2 isolates consist huge number of deletions and insertions individually with different SNPs positions. Therefore entire 38 isolates contains higher degree of homology may be grouped into 3 groups with 2 subgroup in accordance of their polymorphism sequence with reflection of their similarities and dissimilarities. Therefore, sequence of isolates have higher homology, different features are presented in more clear way through classification tree generated through these kinds of qualitative analysis,

then the phylogenetic tree obtained through multialigned sequence by the CLUSTALW program [52] Due to multiple single nucleotide polymorphism change in amino acids were observed they were 15 in group A, which contains similar nucleotide structure to reference group ("regular isolates") and group C isolates consisting ("individual with various insertions"), 34 in group B (isolates with "long insertions). Among all of these 10 silent mutations in A, C and 7 for group B, spike protein is important in context of antigenicity and affinity of receptors. One feature which is noticeable was all amino acid change is found mainly in outer membrane of all groups. So this can be useful for further identifying the genome regions responsible for causing infections SARS-CoV-2.

HKU1 (CoV-HKU1)

The complete sequencing of newly identified virus had been proposed which was designated as HKU1 coronavirus (CoV-HKU1). It is basically of 29,926 nucleotide, polyadenylated RNA. The presence of nitrogenous base guanine and cytosine are found to 32% which lowest among all of the known coronavirus. This strain genome organization is similar to the other cornaviruses strains with gene order replicase, spike, envelope, membrane, nucleocapsid. Its 5' and 3' end consist short UTR regions with 5' end contains putative 5 leader sequence of genome [53, 54] [Fig. 6].
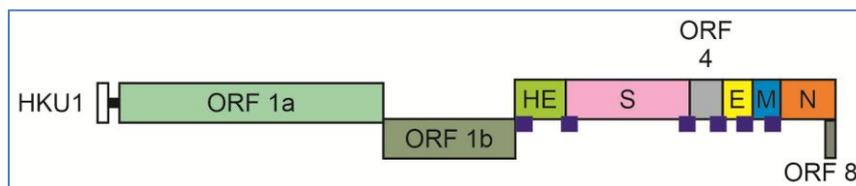

Fig-6

HKU1 it consist 5' UTR and 3' UTR in between them 5' end consist 1a and 1b performs replication, hemagglutinin-esterase (HE) it ensures efficient circularization of virus through one cell to other and 3' end consist ORF encoding structural protein spike (S), envelope (E), membrane (M), Nucleocapsid protein (N), ORF8 in between M and N encodes N2 a hypothetical protein of Nucleocapsid coronavirus family member.

Putative transcription regulatory sequence motif is found at 3' end of leader sequence and leads every ORFs which are translated through ribosome with the exception of ORF5. In sialodacryoadenitis virus (SDAV) infects the upper part of respiratory tract and ORF of mouse hepatitis virus (MHV) encodes the putative envelope protein, it may have same transcription regulatory sequence with ORF4 revealing that envelope protein translated through internal ribosome entry site (IRES) mechanism in cap independent fashion [55]. The Mouse hepatitis virus (MHV) consist 13 nucleotide stretch which was found

to be similar to internal ribosome entry site (IRES and initiation codon upstream of envelope protein was found in HKU1 [56]. Additionally several investigations should be performed for identifying the sequence if it act as IRES for this ORF or whether the sequence known was real transcription regulatory sequence (TRS) for CoV-HKU1. Further analysis of genomes CoV-HKU1 reveled the consistent feature and possible role of CoV-HKU1 genome in recombination. The 3UTR consist a region with predicted stem-loop structure downstream of nucleoprotein gene of 2 to 66 nucleotides. Downstream of the stem loop structure consist 63 to 115 nucleotides with the presence of pseudoknot structure and nucleoprotein gene. Pseudoknot structure is conserved in all of the strain of coronaviruses and plays important role in replication associated with RNA [57]. ORF-2 encodes the hemagglutinin-esterase (HE) glycoprotein present in enveloped virus with 386 amino acid. CoV-HKU1 associated HE consist 50-57% similar amino acid with other groups of coronaviruses HE. Protein family database (PFAM) and Inter Pro Scan observation of

ORF shown that predicted protein of 1 to 349 amino acid residues were family member of hemagglutinin esterase. Prosite Scan analysis of CoV-HKU1 associated the HE protein revealed 8 potential glycosylation N-linked; ORF3 encodes the spike glycoprotein with 1,356 amino acids. The spike protein of CoV-HKU1 contains similarity of 60% in amino acid with other strain of coronaviruses. Prosite Scan of CoV-HKU1 spike protein showed 28 potential N-linked glycosylation. Mostly the spike protein was found to be exposed at extracellular region of virus with transmembrane domain at carboxy terminus with cysteine rich cytoplasmic tail. Multiple alignment analysis revealed the presence of 2 heptad repeats in comparison of other strains of coronavirus. ORF4 encodes the 109 amino acids predicted protein and found to be overlapped with ORF which encodes the envelope protein. Analysis with PFAM revealed that predicted protein was non-structural protein NS2 of coronavirus family member. ORF6 encodes the membrane protein of 223 amino acid. The membrane protein of CoV-HKU1 consist identities of 76 to 84% amino acid with other coronaviruses. Analysis of ORF by PFAM showed predicted membrane protein was of matrix glycoprotein coronavirus family member. ORF7 encodes the predicted nucleoprotein of 441 amino acids. It has similarity of N-protein with other N-proteins of coronavirus of 57-68% amino acid. ORF8 encodes a hypothetical protein (N2) with 205 amino acid with ORF it encodes nucleoprotein.

The N2 hypothetical protein of CoV-HKU1 consist identities with amino acid of 32-39% with N2 protein of other coronaviruses. The prevention for this strain in humans causing respiratory tract infections is yet to known. The clinical serological and phylogenetic studies further would be helpful in determining the importance of CoV-HKU1 in comparison of various respiratory tracts associated viruses, its seroprevalence and genesis of virus.

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2)

SARS-CoV-2 is structurally spherical enveloped particles with genome size of 32 kb approximately single stranded RNA of positive sense and nucleoprotein inside a capsid containing the matrix protein. It consists 6 open reading frames in its genome in which ORF1a and 1b present at 5' end and other at 3' end, encodes structural proteins. All proteins of this strain are translated through sub genomic RNA (sgRNA) of Coronaviruses. It possesses the largest genome among the various known virus with varying percentage of guanine and cytosine. This viral genome consist a different features, specific N-terminal fragments in spike proteins. The major genes for structural proteins of coronavirus occurs in 5'-3' order [58]. There are seen 6 ORFs in all types of strains of coronaviruses genome beside gama coronaviruse consist nsp1, ORF1a and 1b consist frameshift which produces 2 polypeptide 1a encodes NSP3, NSP5 they synthesis Papain like protease and Chymotrypsin like protease (3CLpro) activity and 1ab encodes NSP 12 and 13. Both of these polypeptides encodes accessory and structural proteins are transcribed through single genomic RNA (sgRNA). SARS-CoV2 is composed of nucleocapsid, membrane, spike and envelope protein [Fig. 7].
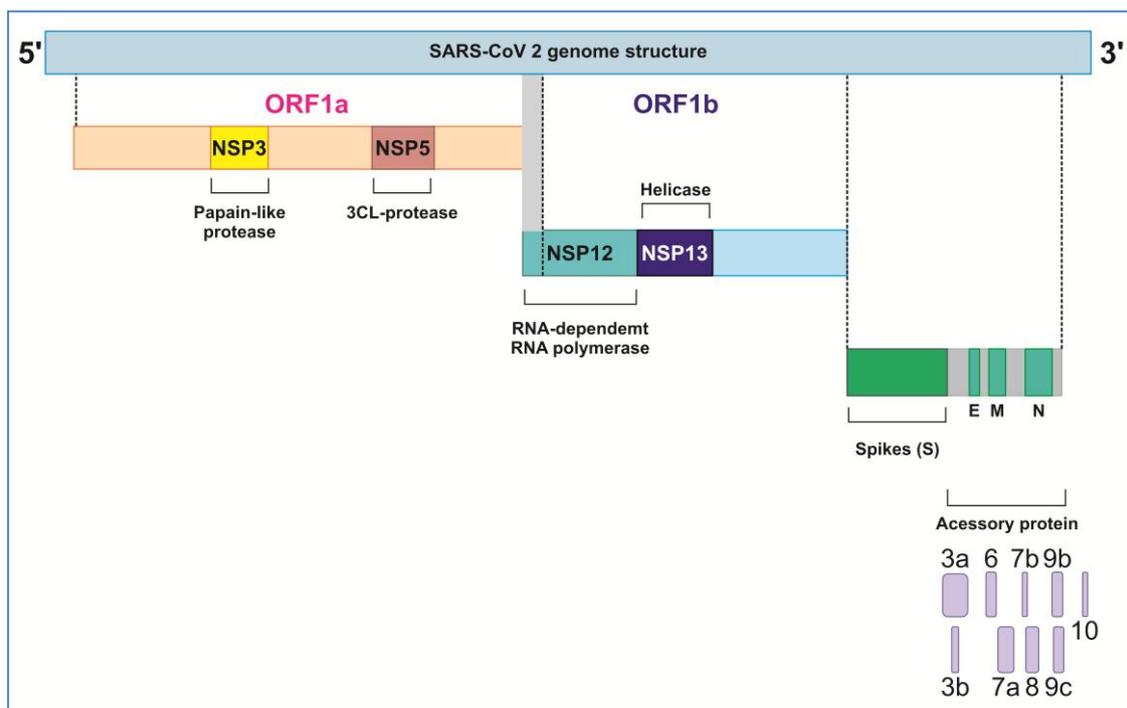

Fig-7: SARS-CoV 2

Its genomes contains 2 large ORFs 1a and 1b encoding NSPs 1-16 these are processed for forming transcription complex of replication, NSP3, 5 function in inhibiting innate response of immune system and NSP12 RNA-dependent polymerase and NSP15 helicase activity of unwinding for DNA SARS-CoV-2 beside structural proteins consist high number of accessory protein at its 3' end ORF 3a and b present between Spike and Envelope protein and ORF 6, 7a and b, 8a and b are present in between membrane and envelope protein and 9b in Nucleocapsid. All these accessory proteins are associated in preventing interferon response, immunosuppression and pathogenesis.

Apart from these four major structural proteins various strains of coronavirus encodes special accessory and structural proteins, like HE, 3a/b, and 4a/b. These proteins play important function in maintenance of genome and its replication.[59] SARS-CoV-2 encodes accessory gene at 3' end ORFs 3a and 3b, 6a,7a,7b, 8a,8b,9b these all are associated in antagonist function of interferon which play important role in innate immunity though they suppress the immune response. Coronaviruses membrane consist 3 to 4 viral proteins and all of the membrane glycoprotein spans the lipid bilayer 3 times is a transmembrane protein inserting long carboxy terminal inside the virion and short terminal of amino acid at outer side of virus [60]. The spike protein consist peplomers which is important for specificity of host and infectivity of virus is a type-1 membrane glycoprotein also this protein is mainly involve in inducing the antibodies associated neutralization. In between the envelope protein molecular interaction exist a molecular interaction which determines the coronaviral membrane composition and formation. Membrane protein has predominant function in virus particles genesis intracellularly independently without the involvement of spike proteins. Among all of them most crucial is spike protein as it binds to the host cell receptors and enters into it. In humans the previously known SARS-CoV and currently known SARS-CoV2 are member of β- coronavirus class and associated with binding to Angiotensinogen converting enzyme receptor-2 (AER-2). Spike protein consist S1 and S2 two subunit function in binding the receptor of host cell and S2 perform function of membrane fusion of host cell and virus. This process is assisted by serine protease TMPRRS2 of cells which primed the S-protein. After the viral RNA genome enters in cytoplasm it translates into polyproteins and structural proteins then viral genome replication starts. Further newly formed glycoprotein insert in ER or Golgi and leading to nucleoprotein formation on account of RNA genome and nucleoprotein formation association. The ER-Golgi intermediate is the place where the particles of virus germinate and finally the vesicle of virus particles fused with plasma membrane to release the virus. This virus associated infection primarily occurs through the respiratory droplets. The person got infected with this strain exhibit the symptoms like cough, fever, fatigue, production of sputum, headache, diarrhea, dyspnoea, lymphopenia. The period of incubation for this disease is 1-14 days before a patient observe any symptoms. Replication occur in mucosal epithelial of upper respiratory tract with multiplication in gastrointestinal mucosa and lower respiratory tract. Coronaviruses in presence of tunicamycin antibiotic produces proteins which were non-infectious and spikeless, consist membrane protein but deprived of Spike protein. Few strains SARS-CoV-2, SARS-CoV and MERS consist intermolecular and intramolecular interaction associated 50 and 30 UTRs which are important for interaction between RNA-RNA and also for binding of both cellular and viral protein mediated. [61] At the 5'end the first ORF for entire genome length is Pb1ab encodes non- structural proteins with 29844 bp 29751bp, 30119bp of SARS-CoV-2, SARS-CoV and MERS, respectively. Genetic observation revealed that SARS-CoV-2 was homologous to SARS-CoV 79% and MERS-CoV 50%. It enters in host cell disrupt the coating of the genome, which were synthesized transcribed and translated at the cytoplasmic membrane involve the continuous and discontinuous synthesis of RNA in coordinated manner through viral replication. Replication associated component, replicase complex contains various cellular proteins and 16 viral subunit apart from the component regularly used by RNA virus helicase, protease, RNA polymerase. Replicase involve in SARS-CoV-2 was assumed to be involved in various processing of RNA enzyme which are not found in various other viruses.

The SARS-CoV-2 associated alignment of Nucleotide sequence shown the 1,516 mutations total across the whole genome of the virus strain in comparison of other NCBI reference strain Wuhan-Hu-1. Among the identified nucleotide 661 were involved in polyprotein region of the virus. Although the regions of structural proteins had 382 nucleotide mutation consisting 183, 17, 148, 34 mutation in Spike, membrane and nucleoprotein mutation respectively. However, in the ORFs in total 1247 nucleotide mutation were seen and among them missense mutation were present in 503 nucleotides. In spike protein 11 amino acid substitutions were there which have been present at the receptor binding domain (RBD) and Spike associated subunit S1 in various countries France, USA and many more. A different kind of mutation was identified in spike protein of SARS-CoV-2 in India. This was not found in other strains in distinct geographical regions. Also 12 and 5 amino acid replacements were recognized in the repeat motif of amino acid Heptad repeat -1 and 2 (HR1,2) containing 892–1,013 and 1,145–1,195 place spike protein S2 subunit and in topological domain 1–18 and 71–78 of membrane protein 5 amino acid are found to be replaced. Envelope and nucleoprotein had replacement of 10 and 75 amino acid respectively. Many

investigators observed the molecular understanding of neutralizing cross-reacting antibody CR3022 with the cryptic epitope23 which is highly conserved on the domain of SARS-coV-2 receptor binding spike protein, distal to receptor binding domain and linked host cell to (ACE2) at the cell surface at the time of infection. Only 1 amino acid is found to be substitute (K378R) among 28 epitope residues of SARS-CoV-2 strain this mutation was not present in coronavirus other strains revealed by various study. It conclude that this may be due to sequence error or evolutionary selection in specific strains. Though, ACE2 binding to spike protein associated residue do not revealed any mutational variation. The genome analysis showed 1,516 variations of nucleotide at distinct positions in entire genome of this strain. The nucleotide excision was seen in 12 sites of entire genome beside the ORF8 spike and ORF7a specifically in polyprotein coding sequence recognized previously. The mutation and protein profile analysis revealed huge number of substitution in amino acid which exhibits the heterogeneous proteins of the SARS-CoV-2 associated viral protein [62, 63]

CONCLUSION

Currently there is not any specific treatment regarding SARS-CoV-2 available this virus have high rate of transmission between humans. As this disease is worldwide pandemic reported so, for revealing its treatment there is need of understanding the other strains of coronaviruses. The genome analysis of all its strains provides the information for knowing the newly emerged SARS-CoV-2 nature. Which somewhat found to be identical to SARS-CoV previously found in China. Further investigation focus on structural and phenotypic consequences of substitution of amino acid overlapping of open reading frame. The identification of mutation present in protein structure and cis acting untranslated region is important for studying the virulence. This mutational divergence is important for further studies with inclusion of metabolic pathway intraviral and virus host interaction for the prognosis of this newly emerged strain of coronavirus. The research undergoing for understanding the SARS-COV-2 includes mechanism of viral replication, molecular determinants of virulence and immune response.

REFERENCES

1. Hoek, L.V.D., Pyrc, K., Jebbink, M.F., -Oost, W.V., Berkhout, R.J., Wolthers, K.C. (2004). Identification of a new human coronavirus. *Nat. Med*; 10: 368 –373.
2. Hoek, L.V.D., Pyrc, K., Berkhout, B. (2006). Human coronavirus NL63, a new respiratory virus. *FEMS Microbiol. Rev*, 30:760 –773.
3. Kaiser, L., Regamey, N., Roiha, H., Deffernez, C., Frey, U. (2005). Human coronavirus NL63 associated with lower respiratory tract symptoms in early life. *Pediatr. Infect. Dis. J,* 24:1015–1017.
4. Hofmann, H., Pyrc, K., Hoek, L.V.D., Geier, M., Berkhout, B., Pöhlmann, S. (2005). Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. *Proc. Nat.l Acad. Sci*, 102:7988 – 7993.
5. Kuba, K., Imai, Y., Rao, S., Gao, H., Guo, F., Guan, B. (2005). A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. *Nat. Med*, 11:875– 879.
6. Reyes, G.R., Kim, J.P. (1991). Sequence-independent, single-primer amplification (SISPA) of complex DNA populations. *Mol. Cell. Probes*, 5:473– 481.
7. Wang, Y., Zhu, N., Li, Y., Lu, R., Wang, H., Liu, G. (2016). Metagenomic analysis of viral genetic diversity in respiratory samples from children with severe acute respiratory infection in China. *Clin. Microbiol. Infect*, 22:458:451– 458.
8. Hacker, J., Kaper, J.B (2000). Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol*, 54:641-679.
9. Grigoriev, A. (2004). Mutational patterns correlate with genome organization in SARS and other coronaviruses. *Trends. Genet*, 20:131-135
10. Sueoka, N. (1961). Correlation between Base Composition of Deoxyribonucleic Acid and Amino Acid Composition of Protein. *Proc Natl. Acad. Sci*, 47:1141-1149.
11. Berkhout, B., van, Hemert, F.J. (1994). The unusual nucleotide content of the HIV RNA genome results in a biased amino acid composition of HIV proteins. *Nucleic Acids Res*, 22:1705-1711.
12. Hemert, F.V.J., Berkhout, B. (1995). The tendency of lentiviral open reading frames to become a-rich: constraints imposed by viral genome organization and cellular tRNA availability. *J. Mol. Evol*, 41:132-140.
13. McIntosh, K., Becker, W.B., Chanock, R.M. (1967). Growth in suckling mouse brain of "IBV-like" viruses from patients with upper respiratory tract disease. *Proc. Natl. Acad. Sci*, 58:2268–2273.
14. Gagneur, A., Sizun, J., Vallet, S., Legr, M.C., Picard, B., Talbot, P.J. (2002). Coronavirus-related nosocomial viral respiratory infections in a neonatal and paediatric intensive care unit: a prospective study. *J. Hosp. Infect*, 51:59–64.
15. Vabret, A., Mourez, T., Gouarin, S., Petitjean, J., Freymuth, F. (2003). An outbreak of coronavirus OC43 respiratory infection in Normandy, France. *Clin. Infect. Dis*, 36:985–989.
16. Arbour, N., Day, R., Newcombe. J., Talbot, P.J. (2000). Neuroinvasion by human respiratory coronaviruses. *J. Virol*, 74:8913–8921.
17. Dessau, R.B., Lisby, G., Frederiksen, J.L. (2001). Coronaviruses in brain tissue from patients with multiple sclerosis. *Acta Neuropathol*, 101: 601–604.
18. Edwards, J.A., Denis, F., Talbot, P.J. (2000). Activation of glial cells by human coronavirus OC43 infection. *J. Neuroimmunol*, 108:73–81.

19. Hogue, B.G., Brian, D.A. (1986). Structural proteins of human respiratory coronavirus OC43. Virus Res. 5:131–144

20. Zaki, A.M., Boheemen, V.S., Bestebroer, T.M., Osterhaus, A.D.M.E., Fouchier, R. A. M. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *The New England Journal of Medicine,* 367: 1814-20.

21. Bermingham, A., Chand, M.A., Brown, C.S., Aarons, E., Tong. C., Langrish, C. (2012). Severe respiratory illness caused by a novel coronavirus, in a patient transferred to the United Kingdom from the Middle East, *September 2012. Euro. Surveill,* *17*(40): 20290.

22. Buchholz, U., Müller, M. A., Nitsche, A., Sanewski, A., Wevering, N., Bauer-Balci, T., ... & Muth, D. (2013). Contact investigation of a case of human novel coronavirus infection treated in a German hospital, October-November 2012. *Eurosurveillance*, *18*(8), 20406.

23. Mailles, A., Blanckaert, K., Chaud, P., Van der Werf, S., Lina, B., Caro, V., ... & Paty, M. C. (2013). First cases of Middle East Respiratory Syndrome Coronavirus (MERS-CoV) infections in France, investigations and implications for the prevention of human-to-human transmission, France, May 2013. *Eurosurveillance*, *18*(24), 20502.

24. Balkhair, A., Al Maamari, K., & Alawi, F. B. (2013). The struggle against MERS-CoV (the novel coronavirus). *Oman medical journal*, *28*(4), 226.

25. Kandeel, M., Nakanishi, M., Ando, T., El-Shazly, K., Yosef, T., Ueno, Y., & Kitade, Y. (2008). Molecular cloning, expression, characterization and mutation of Plasmodium falciparum guanylate kinase. *Molecular and biochemical parasitology*, *159*(2), 130-133.

26. Kandeel, M., Kitamura, Y., & Kitade, Y. (2009, September). The exceptional properties of Plasmodium deoxyguanylate pathways as a potential area for metabolic and drug discovery studies. In *Nucleic acids symposium series* (Vol. 53, No. 1, pp. 39-40). Oxford University Press.

27. Kandeel, M., Miyamoto, T., & Kitade, Y. (2009). Bioinformatics, enzymologic properties, and comprehensive tracking of Plasmodium falciparum nucleoside diphosphate kinase. *Biological and Pharmaceutical Bulletin*, *32*(8), 1321-1327.

28. Kandeel, M., Ando, T., Kitamura, Y., Abdel-Aziz, M., & Kitade, Y. (2009). Mutational, inhibitory and microcalorimetric analyses of Plasmodium falciparum TMP kinase. Implications for drug discovery. *Parasitology*, *136*(1), 11.

29. Kandeel, M., & Kitade, Y. (2013). Computational analysis of siRNA recognition by the Ago2 PAZ domain and identification of the determinants of RNA-induced gene silencing. *PloS one*, *8*(2), e57140.

30. Kandeel, M., & Kitade, Y. (2013). In silico molecular docking analysis of the human Argonaute 2 PAZ domain reveals insights into RNA interference. *Journal of computer-aided molecular design*, *27*(7), 605-614.

31. Chen, Y., Rajashankar, K. R., Yang, Y., Agnihothram, S. S., Liu, C., Lin, Y. L. ... & Li, F. (2013). Crystal structure of the receptor-binding domain from newly emerged Middle East respiratory syndrome coronavirus. *Journal of virology*, *87*(19), 10777-10783.

32. Coleman, C. M., & Frieman, M. B. (2014). Coronaviruses: important emerging human pathogens. *Journal of virology*, *88*(10), 5209-5212.

33. Mackay, I. M., & Arden, K. E. (2015). Middle East respiratory syndrome: an emerging coronavirus infection tracked by the crowd. *Virus research*, *202*, 60-88.

34. Cotten, M., Lam, T. T., Watson, S. J., Palser, A. L., Petrova, V., Grant, P., ... & Kellam, P. (2013). Full-genome deep sequencing and phylogenetic analysis of novel human betacoronavirus. *Emerging infectious diseases*, *19*(5), 736.

35. Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution*, *30*(12), 2725-2729.

36. Wang, N., Shi, X., Jiang, L., Zhang, S., Wang, D., Tong, P., & Arledge, K. C. (2013). Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell research*, *23*(8), 986-993.

37. Elden, V.L.J., Loon, V.A.M., Alphen, V.F., Hendriksen, K.A., Hoepelman, A.I., Kraaij M.G.J. (2004). *Infect. Dis*, 189, 652–657.

38. Hoek, L.V.D. (2007). Antivir. Ther, 12, 651–658.

39. Thiel, V., Herold, J., Schelle, B., & Siddell, S. G. (2001). Infectious RNA transcribed in vitro from a cDNA copy of the human coronavirus genome cloned in vaccinia virusThe GenBank accession number of the sequence reported in this paper is AF304460. *Journal of General Virology*, *82*(6), 1273-1281.

40. Hamre, D., & Procknow, J. J. (1966). A new virus isolated from the human respiratory tract. *Proceedings of the Society for Experimental Biology and Medicine*, *121*(1), 190-193.

41. Hamre, D., Kindig, D. A., & Mann, J. (1967). Growth and intracellular development of a new respiratory virus. *Journal of virology*, *1*(4), 810-816.

42. Chibo, D., & Birch, C. (2006). Analysis of human coronavirus 229E spike and nucleoprotein genes demonstrates genetic drift between chronologically distinct strains. *Journal of general virology*, *87*(5), 1203-1208.

43. Ziebuhr, J., Snijder, E. J., & Gorbalenya, A. E. (2000). Virus-encoded proteinases and proteolytic

processing in the Nidovirales. *Journal of General Virology*, *81*(4), 853-879.

44. Ziebuhr, J. (2005). The Coronavirus Replicase. *Curr. Top. Microbiol. Immunol*, 287:57–94.

45. Ziebuhr, J., Schelle, B., Karl, N., Minskaia, E., Bayer, S., Siddell, S. G., & Thiel, V. (2007). Human coronavirus 229E papain-like proteases have overlapping specificities but distinct functions in viral replication. *Journal of virology*, *81*(8), 3922-3932.

46. Bonavia, A., Zelus, B. D., Wentworth, D. E., Talbot, P. J., & Holmes, K. V. (2003). Identification of a receptor-binding domain of the spike glycoprotein of human coronavirus HCoV-229E. *Journal of Virology*, *77*(4), 2530-2538.

47. Bonavia, A., Zelus, B. D., Wentworth, D. E., Talbot, P. J., & Holmes, K. V. (2003). Identification of a receptor-binding domain of the spike glycoprotein of human coronavirus HCoV-229E. *Journal of Virology*, *77*(4), 2530-2538.

48. Enjuanes, L., Almazán, F., Sola, I., & Zuñiga, S. (2006). Biochemical aspects of coronavirus replication and virus-host interaction. *Annu. Rev. Microbiol.*, *60*, 211-230.

49. Lai, M. M., Baric, R. S., Brayton, P. R., & Stohlman, S. A. (1984). Studies on the mechanism of RNA synthesis of a murine coronavirus. In *Molecular Biology and Pathogenesis of Coronaviruses* (pp. 187-200). Springer, Boston, MA.

50. Baric, R. S., Shieh, C. K., Stohlman, S. A., & Lai, M. M. (1987). Studies into the mechanism of MHV transcription. In *Coronaviruses* (pp. 137-149). Springer, Boston, MA.

51. Boursnell, M. E. G., Brown, T. D. K., Foulds, I. J., Green, P. F., Tomley, F. M., & Binns, M. M. (1987). Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. *Journal of General Virology*, *68*(1), 57-77.

52. CLUSTALW
[ftp://ftp.ebi.ac.uk/software/dos/clustalw].

53. Lai, M., Baric, R. S., Brayton, P. R., & Stohlman, S. A. (1984). Characterization of leader RNA sequences on the virion and mRNAs of mouse hepatitis virus, a cytoplasmic RNA virus. *Proceedings of the National Academy of Sciences*, *81*(12), 3626-3630.

54. Makino, S., Stohlman, S. A., & Lai, M. (1986). Leader sequences of murine coronavirus mRNAs can be freely reassorted: evidence for the role of free leader RNA in transcription. *Proceedings of the National Academy of Sciences*, *83*(12), 4204-4208.

55. Thiel, V., & Siddell, S. G. (1994). Internal ribosome entry in the coding region of murine hepatitis virus mRNA 5. *Journal of General Virology*, *75*(11), 3041-3046.

56. Jendrach, M., Thiel, V., & Siddell, S. (1999). Characterization of an internal ribosome entry site within mRNA 5 of murine hepatitis virus. *Archives of virology*, *144*(5), 921-933.

57. Williams, G. D., Chang, R. Y., & Brian, D. A. (1999). A phylogenetically conserved hairpin-type 3′ untranslated region pseudoknot functions in coronavirus RNA replication. *Journal of virology*, *73*(10), 8349-8355.

58. Woo, P. C., Huang, Y., Lau, S. K., & Yuen, K. Y. (2010). Coronavirus genomics and bioinformatics analysis. *viruses*, *2*(8), 1804-1820.

59. Boheemen, V.S., Graaf, M.D., Lauber, C., Bestebroer, T.M., Raj, V.S, Zaki, A.M.(2012). Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. Bio, 3:6:00473–12

60. De Haan, C. A., Kuo, L., Masters, P. S., Vennema, H., & Rottier, P. J. (1998). Coronavirus particle assembly: primary structure requirements of the membrane protein. *Journal of virology*, *72*(8), 6838-6850.

61. Yang, D., & Leibowitz, J. L. (2015). The structure and functions of coronavirus genomic 3′ and 5′ ends. *Virus research*, *206*, 120-133.

62. Casadevall, A., Pirofski, L.A Te. (2020). Convalescent sera option for containing COVID-19. *J. Clin. Invest*, 130:4: 1545–1548.

63. Sardar, R., Satish, D., Birla, S., & Gupta, D. (2020). Comparative analyses of SAR-CoV2 genomes from different geographical locations and other coronavirus family genomes reveals unique features potentially consequential to host-virus interaction and pathogenesis. *bioRxiv*.