IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A genetic perspective of 2019-nCoV in relation to cross species transmission

[1]Rimjhim Dasgupta
[1]Director
[1]4NBIO

Abstract:
Coronaviruses have caused two large scale pandemics severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) in last two decades. It was thought that SARS-related coronaviruses (SARSr-CoV) is mainly found in bats. Previous studies have shown that some bat SARSr-CoVs have the potential to infect humans. The current outbreak of viral pneumonia in the city of Wuhan, China, was caused by a novel coronavirus designated 2019-nCoV by the World Health Organization, as determined by sequencing the viral RNA genome. Many initial patients were exposed to wildlife animals at the Huanan seafood wholesale market, where poultry, snake, bats, and other farm animals were also sold. Here we have taken an attempt to understand the genetic structure of 2019-nCoV and subsequent sequence analysis of multiple regions of its genome to identify unique motifs, receptor binding domain, hypervariable region which may direct some insight to future research for developing effective treatment against this novel coronavirus. We have identified unique motif in spike protein, multiple hypervariable regions, amino acids polymorphism in ORF8. These may affect the conformation of the peptide and shed some light to cross species transmission, and subsequent host adaptation.

Keywords: Genomics, sequence analysis, 2019-nCov, transmission

Introduction:
The Corona Virus Disease 2019 (COVID-19) caused by a novel coronavirus (CoV) named ‘‘2019 novel coronavirus’’ or ‘‘2019-nCoV’’ by the World Health Organization (WHO) is responsible for the recent pneumonia outbreak that started in early December, 2019 in Wuhan City, Hubei Province, China [1, 2]. This outbreak is associated with a large seafood and animal market, and investigations are ongoing to determine the origins of the infection. Many initial patients were exposed to wildlife animals at the Huanan seafood wholesale market, where poultry, snake, bats, and other farm animals were also sold.
Coronaviruses mainly cause respiratory and gastrointestinal tract infections and are genetically classified into four major genera: Alphacoronavirus, Beta coronavirus, Gamma coronavirus, and Delta coronavirus. The former two genera primarily infect mammals, whereas the latter two predominantly infect birds [3]. Human CoVs include HCoV-NL63 and HCoV-229E, which belong to the Alpha coronavirus genus; and HCoV-OC43, HCoVHKU1, severe acute respiratory syndrome coronavirus (SARS-CoV), and Middle East respiratory syndrome coronavirus (MERS-CoV), which belong to the Beta coronavirus Genus. SARS-CoV and MERS-CoV are considered highly pathogenic, and it is very likely that both SARS-CoV and MERS-CoV were transmitted from bats to palm civets or dromedary camels, and finally to humans. There are still controversies about the source of the 2019-nCoV and its intermediate host. Many studies have proved the pathogen of COVID-19 is a novel coronavirus, which belongs to the Coronavirus family, Beta coronavirus genus and Sarbecovirus subgenus, with a linear single-stranded positive-strand RNA genome of about 30 kb [4]. An attempt has been taken in this article to analyse genetic perspective of 2019-nCoV by taking advantage of currently available genome sequence data from NCBI.

Methodology:
Wuhan isolate, SARS-CoV-2 sequence NC_045512.2 (length 29903 nt) was used as a reference sequence and for sequence comparisons. In the present report we have focused on sequence alignments, we have used NCBI BLAST, and CLUSTAL OMEGA

Results and discussion:
Our sequence alignment (between 2019-nCoV and bat coronavirus RaTG13 (GenBank No.: MN996532) result shows that the amino acid homologies of ORF1ab, Nucleocapsid, Spike proteins are 98.55, 99.05 and 97.41% respectively. These suggest the two viruses have a high genetic relationship in agreement with earlier study [5].
A large gene encoding for a polyprotein (ORF1ab) at the 5′ end of the genome is followed by four major structural protein-coding genes: S = Spike protein, E = Envelope protein, M= Membrane protein, and N =Nucleocapsid protein. There are also at least six other accessory open reading frames (ORFs) [6].

Our study shows that S protein of the two strains (YP_009724390.1, QHR63300.2) has 33 different amino acids with major differences are located at 439–449 and 482–505 (Figure: 1, marked in box) in receptor binding domain. Apart from that, the 2019-nCoV virus has a unique peptide (PRRA) insertion which is consistent with earlier studies [7, 8]. The PRRA motif is located at the 681 of the 2019-nCoV S protein, but not in the S protein of the bat coronavirus RaTG13. As this motif is closed to furin cleavage site and junction of S1 and S2 subunits of

S protein [8] and contains multiple neighbouring serine residues it may potentially phosphorylates by incorporating a large negative group tethered to the sidechain of Ser [9]. This may induce proteolytic cleavage of the spike protein by cellular proteases, and thus impacts host range and transmissibility.

Figure 1: Sequence alignment Query: 2019-nCoV S protein (YP_009724390.1) Subject: RaTG13 S protein (QHR63300)

Moreover, our sequence alignment results show that the S genes of 2019-nCoV and RaTG13 are longer than other SARSr-CoVs. The major differences in the sequence of the S gene of 2019-nCoV and other SARSr-CoVs are the three short insertions in the N-terminal domain This is consistent with another study [10]. Whether the insertions in the N-terminal domain of the S protein in 2019-nCoV confer sialic-acid-binding activity as it does in MERS-CoV [11] needs investigation. Our previous study showed that if we are focusing on only the spike RBD, pangolin has more identity with 2019-nCoV than RaTG13, so probability to cross host barriers and infect humans more than RaTG13 [8].

We have performed sequence alignment (NCBI) of N (nucleocapsid protein) protein of 2019-nCoV and bat coronavirus RaTG13 (YP_009724397.2, QHR63308.1). The nucleocapsid protein is an important structural protein for the coronaviruses. Its function involves entering the host cell, binding to the RNA, and forming the ribonucleoprotein core. It consists of RNA binding domain (RBD; residues 44-180) in the N-terminal region (N) of the protein, linker peptide (residues 181-246), the dimerization domain (DD; residues 247-364) in the C-terminal region [12]. We found 4 different amino acids, which were 37S/P, 215G/S, 243G/S, and 267A/Q, respectively (Figure:2).

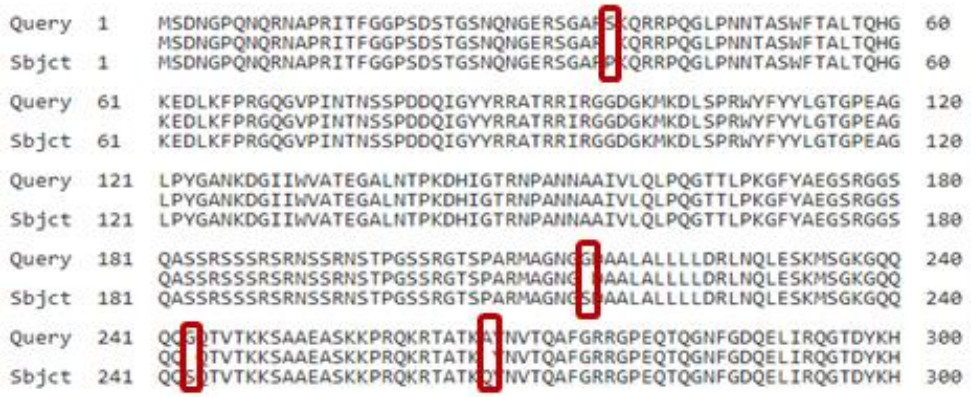

Figure 2: Sequence alignment of N protein of 2019-nCoV (YP_009724397.2) and bat coronavirus RaTG13 (QHR63308.1). A portion of the alignment where the changes observed is presented here

Serine (residue 37 of YP_009724397.2) is polar whereas Proline in QHR63308.1 is non-polar, at position 215 and 243 Glycine is nonpolar and Serine is polar and at 267 alanine of YP_009724397.2 is non-polar whereas in QHR63308.1 Glutamine is polar. These may be responsible for altering the conformation of the protein and hence creates favourable environment for viral passage.

Next, we performed sequence alignment of ORF8 between 2019-nCoV ORF8 and RaTG13 ORF8 (YP_009724396.1, QHR63307.1 respectively). We found >95% identity in with 5 changes in positions 3(F/L), 14(A/T), 26(T/A), 65(A/V) and 84 (L/S). At position 3, both amino acids are non-polar, but phenylalanine has a benzoic ring in the side chain which may stiffen the secondary structure by means of aromatic-aromatic, hydrophobic or stacking interactions. Earlier report in Non- Structural Protein 6 (NSP6) by amino acid change stability (ACS) analysis showed that this (leucine to phenylalanine) leads to a lower stability of the protein structure [13]. Threonine (14 and 26) and Serine (84) are polar whereas Alanine (14) and Leucine (84) are nonpolar amino acids, these may affect the conformation of the peptide. Moreover, at position 65, the residue alanine stabilizes the protein due to its hydrophobic nature. On substitution with Valine in the same position, it can potentially alter the affinity of the molecule towards others and creates favourable environment for virus propagation.

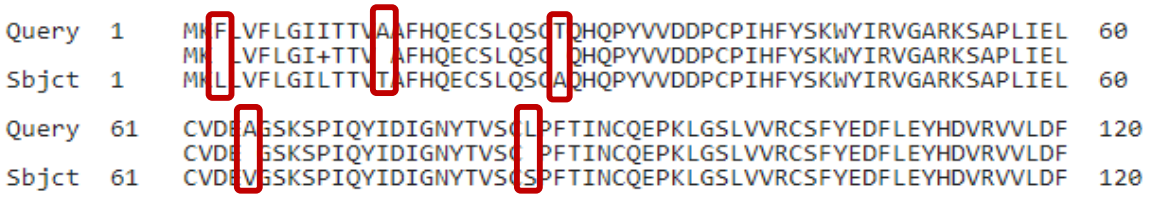

Figure 3: Sequence alignment Query: 2019-nCoV ORF8 YP_009724396.1; Subject: RaTG13 ORF8 QHR63307.1. A portion of the alignment where the changes observed is presented here

Next, we attempted to compare variability in ORF8 between the patient samples. The high sequence similarity (>99%) was observed between the available genome sequences of 2019-nCoVs (from patient samples) with low variability. However, there are at least two hotspots of hypervariability positions at position 24 and 84 ORF8 (Figure 6, provided alignment data for aa84 only) and these cause Ser to Leu and Leu to Ser. Serine is a polar amino acid, and Leucine is nonpolar. This may potentially alter the stability of the protein.

Figure: 4 Sequence alignment between patients 2019-nCoV ORF8, provided alignment data for position 84 only

The levels of genetic similarity between the 2019-nCoV and RaTG13 suggests that the latter does not provide the exact variant that caused the outbreak in humans, but the hypothesis that 2019-nCoV has originated from bats is very likely. It was shown that the novel coronavirus (2019-nCov) is not-mosaic consisting in almost half of its genome of a distinct lineage within the beta coronavirus [14]. However, these genomic features and their potential association with virus characteristics and virulence in humans need further attention. The comprehensive sequence analysis and comparison in conjunction with relative synonymous codon usage (RSCU) bias among different animal species based on the 2019-nCoV sequence suggests that the 2019-nCoV appears to be a recombinant virus between the bat coronavirus and an origin-unknown coronavirus. The recombination occurred within the viral spike glycoprotein, which recognizes cell surface receptor. It was suggested that snake could be the probable wildlife animal reservoir for the 2019-nCoV based on its RSCU bias which is closed to snake compared to other animals [15].

Conclusion:

Taken together this analysis provides some insight about the genomic structure, sequence similarity with other viruses, unique motif in spike protein, receptor binding domain, core positions of high variability and amino acid polymorphism in ORF8. The mutation in ORF8 resulting in one of its two variants, ORF8-L and ORF8-S, is predicted to be affecting the structural disorder of the protein. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission. All together these findings may shed some cautiously light on the possibility of finding effective treatment for this novel coronavirus, starting from already existing anti-betacoronaviridae compounds, which will be dealing with a relatively homogenous viral population. Finally, considering the wide spread of 2019-nCoV in their natural reservoirs, future research should be focused on active surveillance of these viruses for broader geographical regions. Probably in the long term, broad-spectrum antiviral drugs and vaccines may be useful for emerging infectious diseases that are caused by this cluster of viruses in the future.

Declaration of Competing Interest

Author reports no conflict of interest related to the submitted work.

References:

1. Chan JFW, Yuan S, Kok KH, et al. 2020, "A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to person transmission: a study of a family cluster". Lancet. 395, 514-523.
2. Li Q, Guan X, Wu P, et al. 2020, "Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia". N Engl J Med. NEJMoa2001316. https://doi.org/10.1056/NEJMoa2001316
3. Tang, Q., Song, Y., Shi, M., Cheng, Y., Zhang, W., and Xia, X.Q. 2015, "Inferring the hosts of coronavirus using dual statistical models based on nucleotide composition". Sci. Rep. 5, 17155.
4. Ceraolo C, Giorgi FM. Genomic variance of the 2019-nCoV coronavirus. J Med Virol. 2020;92: 522–528. https://doi.org/10.1002/jmv.25700
5. Wu et al. Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. Cell Host & Microbe 27, March 11, 2020; https://doi.org/10.1016/j.chom.2020.02.001
6. Song et al. 2019, "From SARS to MERS, Thrusting Coronaviruses into the Spotlight" Viruses 2019, 11(1), 59; https://doi.org/10.3390/v11010059
7. Li X, Zai J, Zhao Q, et al. Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV-2. J Med Virol. 2020; 1–10. https://doi.org/10.1002/jmv.25731
8. Rimjhim Dasgupta,2020 "Comparative Genomics of Receptor Binding Domains of Spike Protein and Receptor Interaction in COVID-19 Patient". AIJR Preprints,118, version 1
9. Rimjhim Dasgupta, 2020, "Mutations in structural proteins of SARS-CoV-2 and potential implications for the ongoing outbreak of infection in India". AIJR Preprints, 202, version 1.
10. Liu Z, Xiao X, Wei X, et al. Composition and divergence of coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2. J Med Virol. 2020; 1–7. https://doi.org/10.1002/jmv.25726
11. Li W, Hulswit RJG, Widjaja I, et al. 2017, "Identification of sialic acid-binding function for the Middle East respiratory syndrome coronavirus spike glycoprotein". *Proc Natl Acad Sci U S A*. 114(40), E8508-E8517. doi:10.1073/pnas.1712592114
12. Weihong Zeng, Guangfeng Liu, Huan Ma, Dan Zhao, Yunru Yang, Muziying Liu, Ahmed Mohammed, Changcheng Zhao, Yun Yang, Jiajia Xie, Chengchao Ding, Xiaoling Ma, Jianping Weng, Yong Gao, Hongliang He, Tengchuan Jin, 2020, "Biochemical characterization of SARS-CoV-2 nucleocapsid protein". *Biochemical and Biophysical Research Communications* ,527, pp. 618-623, 2020
13. Domenico Benvenuto, Ayse Banu Demir, Marta Giovanetti, Martina Bianchi, Silvia Angeletti, Stefano Pascarella, Roberto Cauda, Massimo Ciccozzi, Antonio Cassone. 2020 "Evolutionary analysis of SARS-CoV-2: how mutation of Non-Structural Protein 6 (NSP6) could affect viral autophagy" *Journal of Infection* 81, e24–e27, https://doi.org/10.1016/j.jinf.2020.03.058
14. Paraskevis et al. 2020, "Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event". Infection, Genetics and Evolution, 79, 104212. https://doi.org/10.1016/j.meegid.2020.104212
15. Ji et al. Homologous recombination within the spike glycoprotein of the newly identified coronavirus 2019-nCoV may boost cross-species transmission from snake to human. doi: 10.1002/fut.22099