AJ ACADEMIC JOURNALS
expand your knowledge

Journal of General and Molecular Virology

Full Length Research Paper

SARS-CoV-2 Molecular Clock and Zoonosis

ALINDA Alfred Khamala[1*], AMBUTSI Michael John[1], KAKAI Shem Khakame[2] and OKOTH Patrick Kirsteen[1]

[1]Department of Biological Sciences, School of Natural Sciences (SONAS), Masinde Muliro University of Science and Technology (MMUST) Kenya.
[2]Department of Agricultural Resource Management, School of Agriculture, Embu University, Embu, Kenya.

SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2), associated with Corona Virus Disease 2019 (COVID-19) coined by World Health Organization, belongs to single stranded RNA viruses (ssRNA Viruses) under Betacoronaviruses. The virus' molecular dynamics are necessary in the wake of human-human transmissions globally with mortality cases on the rise, thusly the race for a vaccine. As the viral genome expresses more human-biased mutations, the coronavirus disease 2019 continues to infect people in their millions, with the available detection kits limiting the numbers detected out of the population. Understanding the molecular basis of the virus through bioinformatics would speed up the viral diagnostics, management and vaccine generation. Currently, the scientific community seeks to give varied perspectives of what is known of the virus at a cellular level. The knowledge is scattered and requires a consolidated flow on thematic understanding in order to ensue further build up towards curbing the disease. The structure and function of the virus, genome and revealed mutations are critical in directing the SARS-CoV-2 virus and disease understanding. Here, we analyze and review published knowledge on the virus in relation to the molecular specs and evolutionary relationships of the virus.

Key words: SARS-CoV-2 infection, COVID-19, RNA viruses.

INTRODUCTION

Coronaviruses fall under the Coronaviridae family that has four genera: Alpha (α), Beta (β), Delta and Gamma (Woo et al., 2007; Lefkowitz et al., 2018). They have a positive-sense single-stranded RNA genome of length 26.4 - 31.7 kb (largest of RNA viruses). The virus causes severe respiratory sickness, dubbed COVID-19 disease, having spread from human-human worldwide at an alarming rate with high fatalities (Huang et al., 2020; WHO, 2020).

SARS-CoV-2 virus is a β-coronavirus of diameter 60 to 140 nm and polymorphically rounded or oval in shape (Zhu et al., 2020). Beta-coronaviruses are enveloped, with a positive-sense, single-strand RNA (+ssRNA Virus), of mammal hosts (Cui et al., 2019). The Coronaviridae family includes SARS-CoV-2, SARS-CoV, common cold viruses (229E, NL63, OC43, and HKU1) and Middle East respiratory syndrome coronavirus (MERS-CoV) (Su et al., 2020). Before December 2019, Cui et al. (2019) reported four Beta-coronavirus, (MERS-CoV, HKU1, OC43 and SARS-CoV), as causes of severe human respiratory diseases. The fifth is SARS-CoV-2 and is responsible for human pneumonia also termed as COVID-19

(Gorbalenya et al., 2020; Wu et al., 2020; Zhou et al., 2020; Zhu et al., 2020). Similar studies by Su et al. (2016) and Chan et al. (2015) document six coronaviruses in humans (Alpha (α), 229E and NL63, Beta (β), OC43 and HKU1, Middle East respiratory syndrome coronavirus (MERSr-CoV) and SARS-associated coronavirus (SARSr-CoV)).

Due to its classification under Sarbecovirus subgenus, SARS-CoV-2 suggestively is of bat origin (Li et al., 2020b; Zhou et al., 2020), transmitted to humans through an intermediate host such as a pangolin (Cyranoski, 2020; Lam et al., 2020; Xiao et al., 2020; Zhang et al., 2020)). Coronaviruses are viruses that transmit around classes Mammalia and Aves of the Kingdom Animalia. Studies document that SARS-CoV and MERS-CoV are zoonotic viruses of bat origin (Lau et al., 2005; Li et al., 2005) where SARS-CoV observes the route bats- palm civets- humans, (Song et al., 2005; Wang et al., 2005) and MERS-CoV observes the route bats-camels-humans (Müller et al., 2014; Chu et al., 2014).

Other than a high rate of transmission, SARS-CoV-2 demonstrates longer incubation period prior to being symptomatic or asymptomatic as is the case in some patients, hence organ failures in selected patients and up to a 3% mortality rate (Huang et al., 2019; Chen et al., 2020; Guan et al., 2019).

Whole SARS-CoV-2 genome structure

Initial genome sequence data placed SARS-CoV-2 in the genus Betacoronavirus, subgenus Sarbecovirus, together with SARS-CoV. However, MERS-CoV fell under subgenus Merbecovirus of the same genus (Lu et al., 2020; Wu et al., 2020; Zhou et al., 2020; Zhu et al., 2020). Prior similarity comparisons reveal a 79% similarity between SARSCoV-2 and SARS-CoV though variations are great at the gene level with 72% base sequence homology on spike (S) protein that binds to host receptors. The novel SARS-CoV-2 as such, has a similar structure with Betacoronavirus viruses whose gene order is (50-replicase ORF1ab-S-envelope (E)-membrane (M)-N-30. The longest replicases gene is ORF1ab at 21 kb length with 16 non-structural proteins and open reading frames (ORFs) downstream. This comparative genomic analysis was enhanced by related virus from *Rhinolophus affinis* bat in 2013 from Yunnan, China (Zhou et al., 2020).

Subsequently, virus RaTG13 has 96% similarity to SARS-CoV-2 with key genomic differences including; SARS-CoV-2 having a furin (polybasic) cleavage site insertion between S1 and S2 subunits of S protein site, that is similar to insertions in HCoV-HKU1 and avian influenza virus but is absent in other Beta-coronaviruses (Coutard et al., 2020). There is an 85% similarity between the receptor binding domain (RBD) of RaTG13 and SARS-CoV-2 sharing a sixth of critical amino acid

residues. Additionally, structural and nucleotide comparisons show that the RBD in SARS-CoV-2 is adapted to binding on human angiotensin-converting enzyme 2 (ACE2) receptors, a phenomenal that is also true for SARS-CoV (Wrapp et al., 2020).

A full length genome of SARS-CoV-2 was first generated by researchers at Yongzhen Zhang China (Wu et al., 2020). The genome's structure had the following order; a 5′-untranslated region (UTR)>replicase complex (orf 1ab) >structural proteins (Spike(S) >Envelope (E) >Membrane (M) >Nucleocapsid (N)) >3′-UTR and non-structural ORFs. On the virus are 6 accessory proteins, whose genes are encoded as ORF3a, ORF6, ORF7a, ORF7b, and ORF8 (Li et al., 2005; Oostra et al., 2007). Different coronaviruses have diverse variable numbers of extra ORFs between the Nucleocapsid and the Spike genes. The virus relies on the transcription regulator motif (TRS) situated at the 3' end of the RNA replication and recombination (Asim et al., 2020).

Further, Lu et al. (2020) report that, SARS-CoV-2 has a complete genome size of 29 kb (29825 bp-29903 bp) falling within the range given by Woo et al. (2007) and Lefkowitz et al., (2018), placing the genome size at 26.4 kb - 31.7 kb. The largest gene segment (ORF1ab gene) is made up of two Open Reading Frames (ORF1a and ORF1b). ORF1ab gene loci in the virus SARS-CoV-2 (251-21541 nt) alternate slightly with the start codon position unlike in MERS-CoV (279-21514nt) and SARS-CoV (265-21486 nt). Between the ORFs there are putative pseudoknot structures presided by a slippery sequence (UUUAAAC). Papain-like protease (PLpro) and 3 C-like protease (3CLpro) cleave to ORF1ab gene (Lu et al., 2020). This is similar to reports by Zhou et al. (2020) and Lu et al. (2020) that document that, ORF1ab gene of SARS-CoV-2 genome has 15–16 non-structural proteins (nsp) encoded at the consensus cleavage sites. However, Lu et al. (2020) further identify nsp 12 and 13 as the encoding gene for RNA-dependent RNA polymerase (RNA pol) and the helicase protein respectively. Additional information detects a single point mutation at the slippery sequence on this novel SARSCoV-2. Moreover, downstream of the ORF1ab (Upstream of S-Gene) is the haemagglutinin esterase (HE) gene in human infecting coronaviruses. There are three surface glycoproteins (spike (S), envelope (E) and membrane (M)) in coronaviruses (Lu et al., 2020).

Coronaviruses have type 1 membrane glycoproteins (S proteins) that form the pronounced "spikes". The S proteins are cleaved into two, the receptor binding S1 domain (high identity with bats) and cell membrane fusion S2 domains (low identity with bats). Therefore, there is high similarity between S-protein domains of SARS-CoV-2 and SARS-CoV that utilize angiotensin-converting enzyme 2 (ACE2) cellular receptor (Lu et al., 2020). However, key amino acid substitutions (at 501, 493, 439, 485 and 486 positions) were detected in the SARS-CoV-2 S-protein domain believed to be equally important in

SARS-CoV (Lu et al., 2020). The E and M genes (for transmembrane proteins) are conserved in coronaviruses.

The nucleocapsid (N) protein dictates the structure. It interacts with the viral RNA genome for packaging within the virus. This assembly signal is well established in SARS-CoV. The 3a protein of SARS-CoV controls virus release through transmembrane complex in channel proteins, apoptosis, Golgi fragmentation, and intracellular vesicles accumulation. Similar to SARS-CoV and MERS-CoV, SARS-CoV-2 has other small Open Reading Frames (ORF9, ORF13, ORF14, and ORF10) situated downstream of N gene (Marra et al., 2003). The function of the N gene and small ORFs of SARS-CoV-2 is not yet known (Asim et al., 2020). A reference sequence for SARS-CoV-2 has been coined from sequence of 63 isolated strains in order to promote viral detection, vaccine design, epidemic investigation, functional analysis, and drug efficacy evaluation. The size of the reference genome strain was 29870 bp (Changtai et al., 2020). Probe and PCR primer design for SARS-CoV-2 should not be based on individual sample sequences but rather a reference sequence to undo the false positives associated with that (Changtai et al., 2020).

Mutations on the SARS-CoV-2 genome

Studies by Changtai et al. (2020) document mutations at nt8782 of ORF1a, nt28144 of ORF 8, and nt29095 of the N section. No mutations were reported on E, 6, 7b regions but rather on ORFs 1a, S, 8, and N region (Changtai et al., 2020). The mutation variants in the study included 46 missense, 52 synonymous, 3 InDels and 14 non-coding alleles. Specifically, the common mutations were at ORF1ab of 13 samples with 8782C > T, ORF8 (28144T > C) 14 samples and N-Region (29095C > T) in 8 samples. ORF1ab recorded the highest number of variants at 24 on the Nucleocapsid 3(NSP3). Further, noncoding mutations were traced in 3′UTR or 5′UTR (Rozhgar et al., 2020). However, ORF8 protein of the SARS-CoV-2 virus lacks known vital region or motif. Total motif VLVVL (amino corrosive 75-79) was documented in ORF8b of SARS-CoV thought to lead to intracellular stress cascade pathways that promote NLRP3 (NOD-like receptor family pyrin region containing-3) (Shi et al., 2019). Additionally, multiple arrays of different ORF8 sequences suggest that L84 relatedness to 28144T > C (L84S) is not preserved (Koyama et al., 2020).

ORF10 consists of short protein peptide, 38 deposits in length. Koyama et al. (2020) depicted ORF10 in SARS-CoV-2 has no comparative proteins in NCBI repository. The study proposes adoption of this unique protein in PCR diagnostics for rapid disease distinction (Koyama et al., 2020). Though similar studies report new annotations of ORF1ab deposited in NCBI, NSP6 is the main contrast putative protein hence the NSP proteins were withheld (Koyama et al., 2020). Further 12 variations were referenced in NSP3 protein in ORF1ab suggesting a linkage between nsp3 association and coronavirus inception (Hurst et al., 2013) though they determined papain-like protease in NSP3 considered important in COVID-19 infection (Niemeyer et al., 2018). Considering the high rate of mutation in RNA viruses, more mutations are expected to appear in the viral genome and will be an important aspect of tracking the spread of SARS-CoV-2 (Grubaugh et al., 2020).

Phylogenesis of SARS-CoV-2 virus

Earlier studies used Pol or N gene as standard practice of phylogenetic tree construction, hence SARS-CoV was suggested to be a gammacoronavirus (Marra et al., 2003; Rota et al., 2003). However, extra amino-terminal domain analysis of the spike protein of the SARS-CoV discovered 19/20 cysteine residues spatially conserved in Betacoronaviruses (Rota et al., 2003).

Further, conserved residues were only five in number within the alpha and gammacorona viruses (Rota et al., 2003). Subsequent phylogenesis based on whole genome analysis aided concluding that SARS-CoV belongs to betacoronavirus, consequently, whole genome phylogenesis of SARS-CoV-2 based on RdRP gene and spike gene revealed its classification falls under genus betacoronavirus (Eickmann, 2003). The coronaviruses SARS-CoV-2 and Yunnan/RaTG13 share a common ancestry, though they vary in spike genes sizes (SARS-CoV-2:3822nt; RaTG13:3809nt). Similar studies report that SARS-CoV and MERS-CoV used intermediate hosts (civets and camels) to infect humans (Guan et al., 2003; Alagaili et al., 2014).

Phylogenetic networks were used by Forster et al. (2020), to document how the SARS-CoV-2 virus spreads across the globe based on three central variants differing by changes in amino acid , named A, B, and C. Type A and C were found in Europe and America while type B in East Asia (Figure 1). Typically, the coronavirus evolution rate is close to 10^{-4} substitute per bp annually, with a possible per replication cycle mutation (Su et al., 2016). Homology at amino acid and nucleotide levels in comparison to reference sequence is at 99.99% (Lu et al., 2019; Chan et al., 2019). The estimated "To Most Recent Common Ancestor" (TMRCA) dates and evolution degree for SARS-CoV-2 show high similarity on various clock models and coalescent tree before subjection to tip-dating technique and constrained-dating technique, but highly distinct over prior coalescent trees. Bayesian analyses plus tip-dating technique by a strict clock and coalescent tree (constant size) prior demonstrated that SARS-CoV-2 evolution rate is 1.24×10^{-3} substitutions per site annually (Li et al., 2020a, b, c), confirming the evolution rate in similar studies (Lu et al., 2020; Chan et al., 2019).
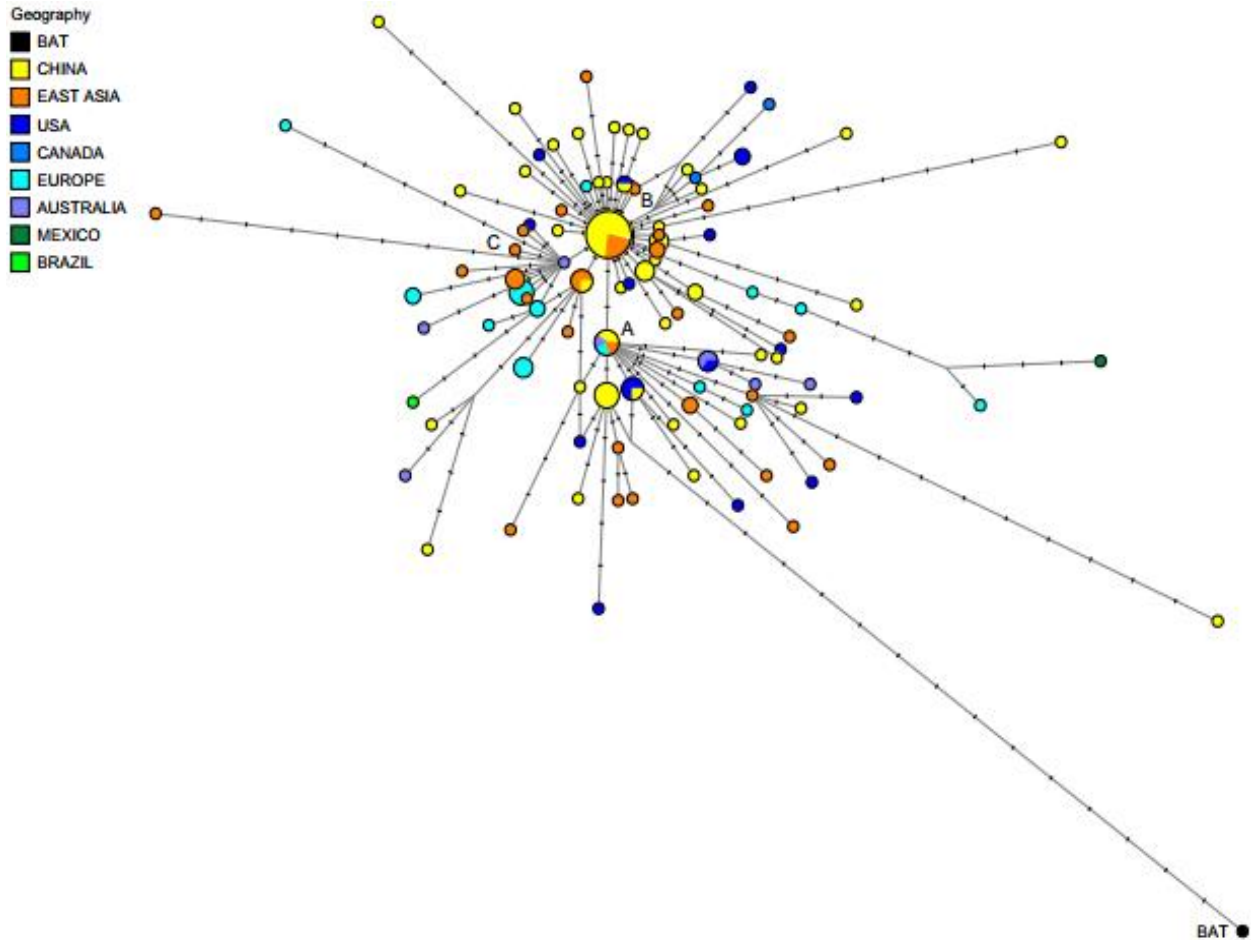
Figure 1. Phylogenetic networks of SARS-CoV-2 courtesy of Forster et al. (2020) (https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2004999117/-/DCSupplemental). A is the root in China. The spread is seen to diverge to B and C, on each, the researchers show notches representing a mutation on the nucleotide sequence.

METHODOLOGY

We downloaded 55 Spike protein sequences from NCBI repository (https://www.ncbi.nlm.nih.gov). The sequences were aligned on BioEdit (Hall, 1999) using MUSCLE. The aligned file was exported onto MEGA-X (Kumar et al., 2018), converted to MEGA file format and using Neighbor Joining algorithm (Saitou and Nei (1987); we constructed a phylogenetic tree.

RESULTS

The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The optimal tree with the sum of branch length = 17.18547576 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (50 replicates) is shown next to the branches (Felsenstein, 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary

distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in the units of the number of amino acid substitutions per site. This analysis involved 55 amino acid sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 1358 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Kumar et al., 2018).

The tree had five clusters. Yunnan, Wuhan and USA SARS-COV-2 coronaviruses clustered together in the first category together with other SARS-CoV viruses from the regions. BetaCov viruses clustered in the second category and closer to other BetaCov identities in the third and fourth clusters. MERS CoV viruses clustered in the last group together with members from England, Egypt, UAE and Moscow. The SARS-CoV-2 cluster suggests a common ancestry with SARS-CoV, with some even closer to the Rousettus bat coronavirus in the first cluster (Figure 2). The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987).
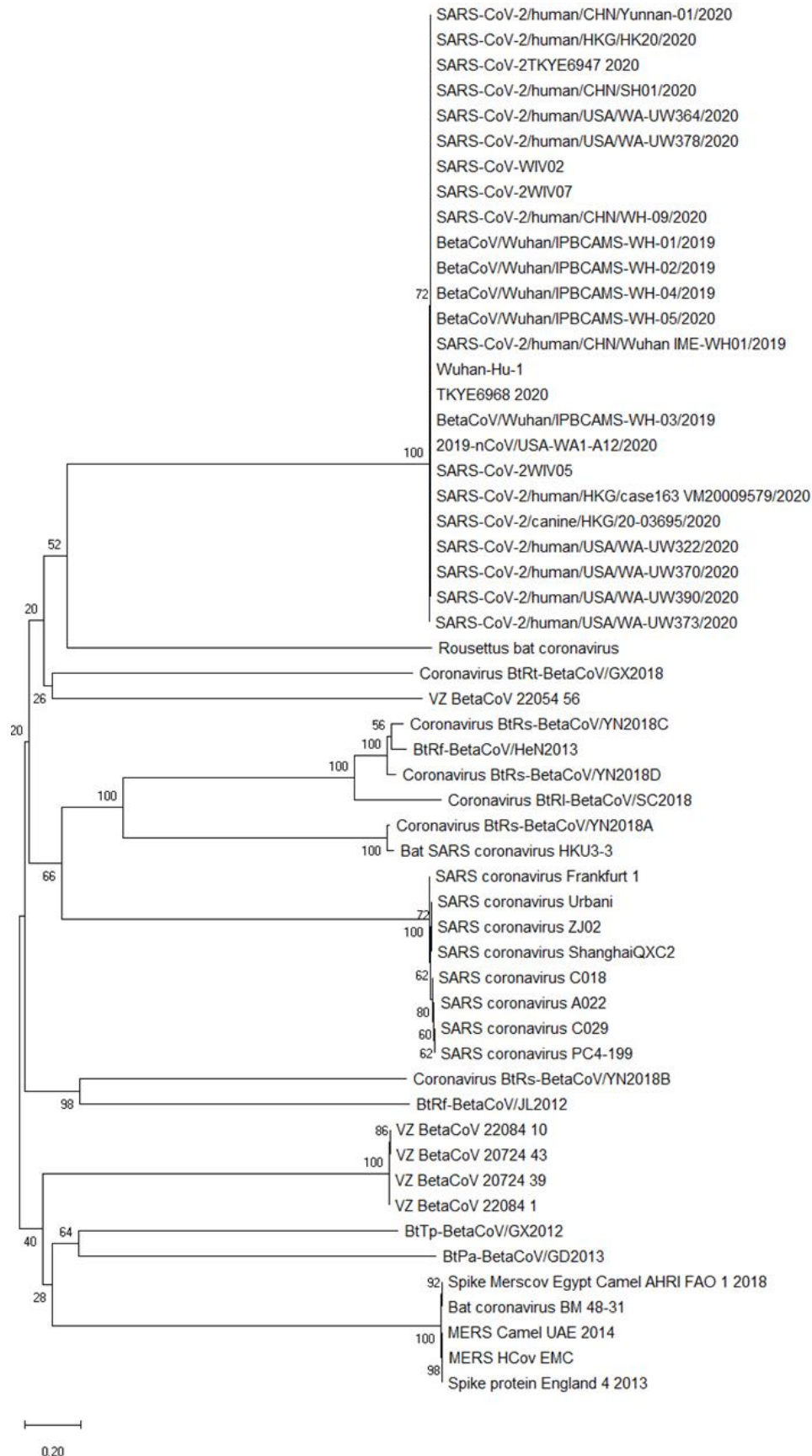
Figure 2. Evolutionary relationships of taxa.

The optimal tree with the sum of branch length = 17.18547576 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (50 replicates) is shown next to the branches (Felsenstein, 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in the units of the number of amino acid substitutions per site. This analysis involved 55 amino acid sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 1358 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Kumar et al., 2018).

DISCUSSION

As more partial and whole genome sequences of SARS-CoV-2 continue being harnessed and deposited in the repository, analyses of these RNA virus will continue encountering more mutations as evident in studies of Forster et al. (2020) and Grubaugh et al. (2020). The evolution of the virus using varied tools agrees on a common ancestry regardless of the type of sequence or loci of the gene from which the parentage is drawn. Most importantly, the structure of the virus is now definite as researchers in the recent studies agree with the perceived arrangement of genes on the viral genome, their functions and viral evolution-turnover rate. However, it will be better if more studies would be conducted on gene clusters with unknown functions (N gene and small ORFs), to open up the structural functions further. Unlike this study that is limited to partial genomes, comparative studies using whole genomes and increased sample size are proposed.

Conclusion

As detection kits continue to fill the market, it will be important if their suitability is based on the most potent scientific criterion that would give consistent results for positive or negative tests. Some studies suggest the use of a tailor made reference genome (Changtai et al., 2020) while others base diagnosis on initial genome of the prior viruses. Others such as Koyama et al. (2020) suggest the use of unique proteins (ORF10) for PCR diagnosis. This is in the wake of more countries expressing concerns over flawed tests where some kits give inconsistent results that stir up mistrust. As we follow through towards vaccine development, and tracking the infection patterns, use of modern technology for both diagnosis and treatment should not be an afterthought. Protocol and due process should be followed through when it comes to

these developments as we look forward to flattening the COVID-19 curve.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

REFERENCES

Alagaili AN, Briese T, Mishra N, Kapoor V, Sameroff SC, de Wit E, Munster VJ, Hensley LE, Zalmout IS, Kapoor A, Epstein JH (2014). Middle East respiratory syndrome coronavirus infection in dromedary camels in Saudi Arabia. MBio 5(4):e01482–14.

Asim B, Uttaran B, Alok Kumar C, Devendra NT, Hasina B, Shanta D (2020). Emergence of Novel Coronavirus and COVID-19: Whether to stay or die out? Critical Reviews in Microbiology doi10.1080/1040841X.2020.1739001.

Chan JFW, Kok KH, Zhu Z, Chu H, To KK-W, Yuan S, Yuen K-Y (2020). Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerging Microbes and Infections 9:221-236.

Chan JFW, Lau SKP, To KKW, Cheng VCC, Woo PCY, Yuen K (2015). Middle East respiratory syndrome coronavirus: Another zoonotic betacoronavirus causing SARS-like disease. Clinical Microbiology Reviews 28:465-522.

Changtai W, Zhongping L, Zixiang C, Xin H, Mengyuan X, Tengfei H, Zhenhua Z (2020). The establishment of reference sequence for SARS-CoV-2 and variation analysis. Journal of Medical Virology 92:667-674.

Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Yu T (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. Lancet 395:507-513.

Chu DKW, Poon LLM, Gomaa MM, Shehata MM, Perera RA, Zeid DA, El Rifay AS, Siu LY, Guan Y, Webby RJ, Ali MA (2014). MERS coronaviruses in dromedary camels, Egypt. Emerging Infectious Diseases 20:1049-1053.

Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG, Decroly E (2020). The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. Antiviral Research 176:104742.

Cui J, Li F, Shi ZL (2019). Origin and evolution of pathogenic coronaviruses. Nature Reviews Microbiology 17(3):181-192.

Cyranoski D (2020). Did pangolins spread the China coronavirus to people?. Nature doi:10.1038/d41586-020-00364-2.

Eickmann M (2003). Phylogeny of the SARS Coronavirus. Science 302(5650):1504b–11505.

Felsenstein J (1985). Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783-791.

Forster P, Lucy F, Colin R, Michael F (2020). Phylogenetic network analysis of SARS-CoV-2 genomes. doi:10.1073/pnas.2004999117.

Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA (2020). Severe acute respiratory syndrome-related coronavirus: The species and its viruses – A statement of the Coronavirus Study Group. bioRxiv doi:10.1101/2020.02.07.937862.

Grubaugh ND, Petrone ME, Holmes EC (2020). We shouldn't worry when a virus mutates during disease outbreaks. Nature Microbiology 5:529–530.

Guan WJ, Ni Z, Hu Y, Liang W, Ou CQ, He Jx, Liu L, Shan H, Lei CL, Hui DSC, Du B, Li LJ, Zeng G, Kowk-Yung Y, Ru-chong C, Chun-li T, Tao W, Ping-yan C, Jie X, Shi-yue L, Jin-lin W, Zi-jing L, Yi-xiang P, Li W, Yong L, Ya-hua H, Peng P, Jian-ming W, Ji-yang L, Zhong C, Gang L, Zhi-jian Z, Shao-qin Q, Jie L, Chang-jiang Y, Shao-yong Z, Nan-shan Z (2019). Clinical characteristics of 2019 novel coronavirus infection in China. medRxiv. 2020:20020974. https://doi.org/10.1101/2020.02.06.20020974.

Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, Luo SW, Li PH, Zhang LJ, Guan YJ, Butt KM (2003). Isolation and characterization of viruses related to the SARS coronavirus from animals in Southern China. Science 302(5643):276-278.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41:95-98.

Huang C, Wang Y, Li X, Ren, L., Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395:497-506.

Hurst KR, Koetzner CA, Masters PS (2013). Characterization of a critical interaction between the coronavirus nucleocapsid protein and nonstructural protein 3 of the viral replicase-transcriptase complex. Journal of Virology 87(16):9159–9172.

Koyama T, Platt D, Parida L (2020). Variant analysis of COVID-19 genomes. Bull. World Health Organ. https://doi.org/10.2471/BLT.20.253591.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution 35:1547-1549.

Lam TTY, Shum MHH, Zhu HC, Tong YG, Ni XB, Liao YS, Wei W, Cheung WYM, Li WJ, Li LF, Leung GM, Holmes EC, Hu YL, Guan Y (2020). Identification of 2019-nCoV related coronaviruses in Malayan pangolins in southern China. bioRxiv, doi:10.1101/2020.02.13.945485.

Lau SKP, Woo PCY, Li KSM, Huang Y, Tsoi HW, Wong BH, Wong SS, Leung SY, Chan KH, Yuen KY (2005). Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proceedings of the National Academy of Sciences of the United States of America 102:14040-14045.

Lefkowitz EJ, Dempsey DM, Hendrickson RC, Orton RJ, Siddell SG, Smith DB (2018). Virus taxonomy: The database of the International Committee on Taxonomy of Viruses (ICTV). Nucleic Acids Research 46(D1):D708–D717.

Li W, Shi Z, Yu M, Wuze R, Craig S, Jonathan HE, Hanzhong W, Gary C, Zhihong H, Huajun Z, Jianhong Z, Jennifer M, Hume F, Peter D, Bryan TE, Shuyi Z, Lin-Fa W (2005). Bats are natural reservoirs of SARS-like coronaviruses. Science 310:676-679.

Li X, Junjie Za, Qiang Z, Qing N, Yi L, Brian TF, Antoine C (2020a). Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV-2. Journal of Medical Virology doi:10.1002/jmv.25731.

Li X, Wang W, Zhao X, Zai J, Zhao Q, Li Y, Chaillon A (2020b). Transmission dynamics and evolutionary history of 2019-nCoV. Journal of Medical Virology. https://doi.org/10.1002/jmv.25701

Li X, Zai J, Wang X, Li Y (2020c). Potential of large 'first generation' human-to-human transmission of 2019-nCoV. Journal of Medical Virology 92:448-454.

Lu R,Zhao X, Li J,Niu P, Yang B, Honglong W, Wenling W, Hao S, Baoying H, Na Z, Yuhai B, Xuejun M, Faxian Z, Liang W, Tao H, Hong Z, Zhenhong H, Weimin Z, Li Z, Jing C, Yao M, Ji W, Yang L, Jianying Y, Zhihao X, Jinmin M, William JL, Dayan W, Wenbo X, Edward CH, George FG,Guizhen W, Weijun C, Weifeng S, Wenjie T (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. The Lancet 395(10224):565–574.

Marra MA, Jones SJM, Astell CR, Holt RA, Brooks-Wilson A, Yaron SNB, Jaswinder K, Jennifer KA, Sarah AB1, Susanna YC, Alison C1, Shaun MC, Doug F, Noreen G, Obi LG, Stephen RL, Michael M, Helen M, Stephen BM, Pawan KP, Anca SP, Gordon AR, Jacqueline ES, Asim S, Duane ES, Jeff MS, George SY, Francis P, Anton A, Harvey A, Nathalie B, Kathy B, Timothy FB, Donnie B, Martin C, Michael D, Lisa F, Ramon F, Michael G, Michael G, Allen S, Steven J, Heinz F, Adrienne M, Amin K, Yan L, Susan N, Ute S, Graham AT, Shaun T, Robert V, Diane W, Brynn W, Robert CB, Mel K, Martin P, Danuta MS, Chris U, Rachel LR (2003). The genome sequence of the SARS-associated coronavirus. Science 300(5624):1399–1404.

Müller MA, Corman VM, Jores J, Meyer B, Younan M, Liljander A, Bosch BJ, Lattwein E, Hilali M, Musa BE, Bornstein S, Drosten C (2014). MERS coronavirus neutralizing antibodies in camels, Eastern Africa, 1983-1997. Emerging Infectious Diseases 20:2093-2095.

NCBI Repository (National Center for Biotechnology Information, U.S). National Library of Medicine 8600 Rockville Pike, Bethesda MD, 20894 USA. Available at https://www.ncbi.nlm.nih.gov.

Niemeyer D, Mösbauer K, Klein EM, Andrea S, Robert CM, Anna MM, Ronald D,Susan CB, Christian D, Marcel AM (2018). The papain-like protease determines a virulence trait that varies among members of the SARS-coronavirus species. PLoS Pathology 14(9):e1007296.

Oostra M, de Haan CA, Rottier PJ (2007). The 29-nucleotide deletion present in human but not in animal severe acute respiratory syndrome coronaviruses disrupts the functional expression of open reading frame 8. Journal of Virology 81:13876-13888.

Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, Peñaranda S, Bankamp B, Maher K, Chen M-H, Tong S, Tamin A, Lowe L, Frace M, DeRisi JL, Chen Q, Wang D, Erdman DD, Peret TCT, Burns C, Ksiazek TG, Rollin PE, Sanchez A, Liffick S, Holloway B, Limor J, McCaustland K, Olsen-Rasmussen M, Fouchier R, Günther S, Osterhaus ADME, Drosten C, Pallansch MA, Anderson LJ, Bellini WJ (2003). Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science 300(5624):1394-1399.

Rozhgar AK, Muhamad S, Mehmet O (2020). Genomic characterization of a novel SARS-CoV-2. https://doi.org/10.1016/j.genrep.2020.100682.

Saitou N, Nei M (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4:406-425.

Shi CS, Nabar NR, Huang NN, Kehrl JH (2019). SARS-Coronavirus Open Reading Frame-8b triggers intracellular stress pathways and activates NLRP3 inflammasomes. Cell Death Discovery 5;101.

Song HD, Tu CC, Zhang GW, Sheng-Yue W, Kui Z, Lian-Cheng L, Qiu-Xia C, Yu-Wei G, Hui-Qiong Z, Hua X, Hua-Jun Z, Shur-Wern WC, Feng C, Chun-Ming P, Hua X, Sai-Juan C, Hui-Ming L, Duan-Hua Z, Yu-Fei L, Jian-Feng H, Peng-Zhe Q, Ling-Hui L, YQR, Wen-Jia L, Ye-Dong Y, Larry A, Ming W, Rui-Heng X, Xin-Wei W, Huan-Ying Z, Jin-Ding C, Guodong L, Yang G, Ming L, Ling F, Li-Yun J, Hui L, Fang C, Biao D, Li-Juan H, Jin-Yan L, Suxiang T, Xiangang K, Lin D, Pei H, Hua T, Andrea B, Xiao-Jing Y, Ottavia S, Zong-Ming G, Hai-Yan P, Wei-Zhong H, Jean-Claude M, Arnaud F, Antoine D, Neri N, Yi-Xue L Chung-I W, Guo-Ping Z (2005). Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. Proceedings of the National Academy of Sciences of the United States of America 102:2430-2435. https://doi.org/10.1073/pnas.0409608102.

Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, Liu W, Bi Y, Gao GF (2016). Epidemiology, genetic recombination, and pathogenesis of coronaviruses. Trends Microbiology 24:490-502.

Wang M, Yan M, Xu H,Liang W, Kan B, Zheng B, Chen H, Zheng H, Xu Y, Zhang E, Wang H, Ye J, Li G, Li M, Cui Z, Liu Y-F, Guo R-T, Liu X-N, Zhan L-H, Zhou D-H, Zhao A, Hai R, Yu D, Guan Y, Xu J (2005). SARS-CoV infection in a restaurant from palm civet. Emerging Infectious Diseases 11:1860-1865.

Woo CYP, Wang M, Lau SKP, Huifang X, Rosana WSP, Rongtong G, Beatrice HL W, Kai G, Hoi-Wah T, Yi H, Kenneth SML, Carol SFL, Kwok-Hung C, Bo-Jian Z, Kwok-Yung Y (2007). Comparative analysis of twelve genomes of three novel group 2c and group 2d coronaviruses reveals unique group and subgroup features. Journal of Virology 81(4):1574-1585.

World Health Organization (WHO) (2020) Novel Coronavirus (COVID-19) Situational Reports. Available at https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/ (30th April 2020, date last accessed).

Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, Graham BS, McLellan JS (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 367:1260-1263.

Wu F, Zhao S, Bin Y, Yan-Mei C, Wen W, Zhi-Gang S, Yi H, Zhao-Wu T, Jun-Hua T, Yuan-Yuan P, Ming-Li Y, Yu-Ling Z, Fa-Hui D, Yi L, Qi-Min W, Jiao-Jiao Z, Lin X, Edward CH, Yong-Zhen Z (2020). A new coronavirus associated with human respiratory disease in China. Nature, doi:10.1038/s41586-020-2008-3.

Xiao K, Zhai J, Yaoyu F, Niu Z, Xu Z, Jie-Jian Z, Na L, Yaqiong G, Xiaobing L, Xuejuan S, Zhipeng Z, Fanfan S, Wanyi H, Yu L, Ziding Z, Rui-Ai C, Ya-Jiang W, Shi-Ming P, Mian H, Wei-Jun X, Qin-Hui C, Fang-Hui H, Yahong L, Wu C, Lihua X, Yongyi S (2020). Isolation

and characterization of 2019-nCoV-like coronavirus from malayan pangolins. bioRxiv doi:10.1101/2020.02.17.951335.

Zhang T, Wu QF, Zhang ZG (2020). Pangolin homology associated with 2019-nCoV. bioRxiv, doi:10.1101/2020.02.19.950253.

Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, Si H-R, Zhu Y, Li B, Huang C-L, Chen H-D, Chen J, Luo Y, Guo H, Jiang R-D, Liu M-Q, Chen Y, Shen X-R, Wang X, Zheng X-S, Zhao X, Chen Q-J, Deng F, Liu L-L, Yan B, Zhan F-X, Wang Y-Y, Xiao G-F, Shi Z-L (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature, doi:10.1038/s41586-020-2012-7.

Zhu N, Zhang D, Wenling W, Xingwang L, Bo Y, Jingdong S, Xiang Z, Baoying H, Weifeng S, Roujian L,Peihua N, Faxian Z, Xuejun M, Dayan W, Wenbo X, Guizhen W, George FG, Wenjie T (2020). A novel coronavirus from patients with pneumonia in China, 2019. The New England Journal of Medicine 382(8):727–733.

Zuckerkandl E, Pauling L (1965). Evolutionary divergence and convergence in proteins. Edited in Evolving Genes and Proteins by Bryson V, Vogel HJ. Academic Press, New York. pp. 97-166.