Variants in Saudi Arabian SARS-Cov-2 Genome; Causes and Similarity to Wuhan hCoV-19/Wuhan/WIV04/2019

A. Khuzaim Alzahrani

October 24, 2022

Variants in Saudi Arabian SARS-Cov-2 Genome; Causes and similarity to Wuhan strain hCoV-19/Wuhan/WIV04/2019

A. Khuzaim Alzahrani[1]

[1]Medical Laboratories Technology, Faculty of Applied Medical Sciences, Northern Border University, P.O. Box 1321, Arar 91431, Saudi Arabia; akaalz@nbu.edu.sa

Variants in Saudi Arabian SARS-Cov-2 Genome; Causes and similarity to Wuhan strain hCoV-19/Wuhan/WIV04/2019

Abstract

In addition to SARS-Cov-2, Saudi Arabia have already experienced similar sever Beta-coronavirus outbreak in 2003 and 2012. Like SARS and MERS-CoV, SARS-Cov-2 begins with a minor upper respiratory injury and progresses to serious respiratory illness. Infectious viruses, including SARS-Cov-2, use different strategies such as mutations which alter the virus phenotype in a way that confers its fitness advantage. These mutations play critical roles in establishing successful entry of the virus into its host cell. By the present study, and using a group of bioinformatics tools, we have analysed the genomic variation in a 53 SARS-CoV-2 strains recorded in Saudi Arabia and compared to Wuhan strain hCoV-19/Wuhan/WIV04/2019. A significant population expansion of SARS-CoV-2 transmission was recorded (Tajima’s D=-2.486, P<0.001; Fu'sFs=-4.658, P<0.001). In total, 147 polymorphic sites were detected within the 11 genes. Among them, there were 110 singleton variable sites and 37 parsimony informative sites. The changes in the evolutionary relationship between the 53 Saudi SARS-CoV19 variants, reflects that the analysed genotypes were clustered at different groups compared to Wuhan strain hCoV-19/Wuhan/WIV04/2019. Indeed, the recorded amino acids changes (aa) were with highest percentage on ORF1ab region (52 %), followed by the spike, nucleocapsid and finely the envelop (42%, 38% and 25 % respectively). These are the most common sites undergoing to an aa change, providing an insight of some important proteins of the COVID-2019 that are involved in the mechanism of viral entry and viral replication. Thus, understanding the main drivers for pathogen appearance, spreading, and supremacy on human defences is highly required as these mutations may adversely affect all efforts of drug repurposing.

Introduction

In 2019 the world was devastated by the outbreak of SARS-CoV-2 that started in in Wuhan, Hubei Province, China where patients developed pneumonia signs as well as a diffused alveolar injury, resulting in acute respiratory distress syndrome (ARDS). Saudi Arabia have already experienced similar sever Beta-coronavirus outbreak in 2012, a pair of Saudi Arabian nationals were diagnosed with a new coronavirus named the Middle East Respiratory Syndrome Coronavirus (MERS-CoV). However, the outbreak was closely monitored and dealt with in a professional way that reduces the mortality to 838 deaths among the reported cases (1). Like SARS, MERS-CoV infection begins with a minor upper respiratory injury and progresses to serious respiratory illness (2).

Both MERS-CoV and SARS-CoV lie within the β group of four genera of coronaviruses that include α, β, γ, and δ, however, only type α and β are known to infect humans (3). Although, both types normally originate in bats or rodent (4), a positive transition to humans in various incidents. SARS-CoV-2 has and adaptive genetic strategy similar to that of a new virus from space trying to fit into, and replicate within, the genetic-background and thus biochemistry of the host cells. Thus, both innate and adaptive immune responses of human body to coronaviruses resistance can be affected by these genetic variants (Mutations). These mutations that may affect virus adaption and fitness were rarely occur in comparison with tolerated low-effect or no-effect 'neutral' amino acid shifts (5). Thanks to the proofreading activity of Nsp12 exonuclease, these significant mutations are regularly eliminated, thus, most of the detected mutations in SARS-CoV-2 genomes are expected to be either neutral or mildly (6). However, these type of mutations alter the virus phenotype in a way that confers a fitness advantage, in at least some contexts. Indeed, only few months after the evolution of SARS-CoV-2 within the human population, before a fitness-enhancing mutations were confirmed to have arisen.

SARS-CoV-2 consist of four main structural proteins- (i) Nucleocapsid protein (N), (ii) Spike protein (S), (iii) Membrane protein (M), and (iv) Envelope protein (E). E protein is the smallest of which ranging from 8–12 kDa, and is involved in a wide spectrum of functional repertoire(7). E protein includes three domains: (i) short hydrophilic N-terminus domain consisting of 7–12 amino acids, (ii) hydrophobic transmembrane domain that is about 25 amino acids long, and (iii) long hydrophilic C terminal region (8,9). Whereas, the S protein has two major domains, S1 which facilitate the attachment and binding to the host cell receptor, and S2 that mediates fusion to the host cell membrane (10). Both domains play critical roles in establishing successful entry of the virion into its host cell. SARS-CoV-2 is known to mutate rapidly especially, the Spike protein to escape host cell immune system and adapt with the host environment (11). Unlike S protein, the N gene with its 90% amino acid homology and less mutations over time is more conserved and stable (12–16). Both M and N proteins of many SARS-CoV-2 were reported to be the most abundant viral protein expressed during infection and a key protein in the assembly of both naked and enveloped virus particles(17). M protein plays a critical role in the intracellular formation of virus particles, for which S appears not to be required (18).

Currently, Saudi Arabia has recorded over 541,201 positive cases (5554 of which are still active) and 8458 deaths (as of Aug 20, 2021). Evaluate the possible levels of genetic diversity and polymorphism existing in SARS-CoV-2 genomes in Saudi Arabia. By the present study, 53 genomic sequences from Saudi Arabian studied cases were analysed. The sequences were retrieved from the National Centre for Biotechnology Information (NCBI). Compared to the reference sequence for SARS-CoV-2 (Wuhan strain hCoV-19/Wuhan/WIV04/2019) by homology and phylogenetic, tree analyses, mutations at different locations, and preliminary bioinformatics analyses were studied. Many attempts have been performed to study SARS-CoV-2 variants around the world since the breakout, most of them on a particular group of genomes using limited datasets. Whereas, major lineages didn’t propose until March 2020, these include the interesting proposal that identified the same major lineages (named A and B) and other sublineages (19). Our aims were to: (1) characterize genomic variations of SARS-CoV-2; (2) infer the evolutionary relationships of the Saudi strains; and (3) deduce the transmission history of Saudi SARS-CoV-2 with Wuhan strain.

Methods

Bioinformatics

The current work is an in silico study aimed to evaluate the genetic variations of SARS-CoV-2 in Saudi Arabia recorded strains. SARS-CoV-2 sequences were obtained from severe acute respiratory syndrome coronavirus 2 data hub NCBI (May/2021). The sequence of Wuhan strain hCoV-19/Wuhan/WIV04/2019 included as the reference sequence. A 53 sequence were aligned on MEGA X platform against the reference (20), using muscle algorithm. Evolutionary relationships: The Neighbor-Joining method was used to determine the evolutionary relationship in the form of a phylogenetic tree. The Maximum Composite Likelihood method was used to compute the evolutionary distances (21) and represent units of the number of base substitutions per site. This analysis involved two sets of 53 nucleotide sequences of SARS-2 from Saudi Arabia and the reference strain. All ambiguous positions have been removed for each sequence pair (pairwise deletion option). There were a total of 29,903 positions in the final dataset. Evolutionary studies were conducted in MEGA X software (20).

Statistical Analysis

Using the multiple alignment analysis (msa) package from R platform that provides a unified R/Bioconductor interface to the multiple sequence alignment algorithms including Muscle (22). Evolutionary analyses were conducted in MEGA X (20), using the aligned sequences that also used to measure both the transition/transversion rate ratios in current population.

Using DnaSP ver. 5.10.00, (23), genetic diversity was calculated applying the indices of haplotype diversity (Hd) (24), and pairwise estimates of nucleotide diversity (p) (Jukes and Cantor, 1996). Selection neutrality tests were estimated by Tajima'sD (25) and Fu and Li's D* and F* methods (26) In these analyses, the insertions/deletions (indels) were treated as missing data in each analysis. The program TCS ver. 1.06 (27), was used to construct the phylogenetic network of haplotypes. In this software, the indels were recorded as nucleotide substitutions. Both nucleotide substitution and maximum likelihood estimate of gamma parameter for site rates were conducted on MEGA X. the transition/transversion ratio (R) represent the ratio of the number of transitions to the number of transversions for the included sequences. The R value indicate the bias towards either transitional or transversional substitution.

Extraction of haplotypes

The haplotypes and sequence variations were extracted of the DNA matrix using the R base package (28), to compare sequences and extract the haplotype number, frequency. Unique haplotype sequences were calculated by ignoring common nucleotides between haplotypes. The haplotypes were also calculated using the DNAsp application test the ratio of the extracted haplotypes using the available tools (23), and haplotype Neighbor-Joining tree generated using PopART Population Analysis and genetics software (29).

Amino acid (aa) mutation analysis

The Aa mutations analysis was conducted on CoVsurver that is a research tool developed to aid the research community with the identification, analysis and interpretation of Aa changes in coronavirus genomes. The 53 SARS-2 from Saudi Arabia compared with the reference genome of Wuhan strain hCoV-19/Wuhan/WIV04/2019. The aa mutations and their positions on the viral genome were presented along their frequency change(s) in each structural models (30).

Result

SARS-CoV19 diversity analysis

The examined sequences consisted of the whole genome of 54 genotypes: 53 sequences for

SARS-Cov2 genotypes from Saudi Arabia and the genotype of Wuhan sequence. The total aligned sequence length was 29695 bp coding for 11 genes (Table 1). The GC content of the non-coding region (50.2 %) studied is higher than those of coding regions spacers (37% for ORF1ab, S, for ORF3a, and 38% for N and M, 36% for ORF6 and ORF7a, 40% ORF8 and 33% for E). In total, 147 polymorphic sites were detected within the 11 genes. Among them, there were 110 singleton variable sites and 37 parsimony informative sites.

As showed by table 2a,b, the mismatch distribution analysis of the 53 genomes exhibited a significant population expansion of SARS-CoV-2 transmission (Tajima's D= -2.486, P<0.001; Fu'sFs=-4.658, P<0.001). Indeed, the estimated value of the shape parameter for the discrete Gamma Distribution is 0.0500. Substitution pattern and rates were estimated under the Tamura-Nei (1993) model (+G) (31). A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories, [+*G*]). Mean evolutionary rates in these categories were 0.00, 0.00, 0.00, 0.03, 4.97 substitutions per site. The nucleotide frequencies are A = 29.88%, T = 32.14%, C = 18.35%, and G = 19.64%. For estimating ML values, a tree topology was automatically computed. The maximum Log likelihood for this computation was -41827.691 (Table 3**)**

The evolutionary relationship in the form of a phylogenetic tree showed that the 53 Saudi SARS-CoV19 variants are clustered at different groups, with MT820461 being the closest to Wuhan strain hCoV-19/Wuhan/WIV04/2019; the accession number of each sequence was cited according the NCBI, Figure showed great diversity indicating high level of changes in the viral genomic sequences compared to the out group (Fig.1). Indeed, a nucleotide frequencies analysis of current sequences that differ at the following rate (Table 2a,b): 29.89% (A), 32.12% (T), 18.35% (C), and 19.64% (G). Thus, the transition/transversion rate ratios followed these differences where k_1 = 3.247 (purines) and k_2 = 7.306 (pyrimidines). The overall transition/transversion bias is R = 2.485.

SARS-CoV19 Haplotype analysis

Haplotype network of COVID-19 genomes were used to reveal the genetic distance and evolution relationship between different Saudi strains COVID-19 haplotypes and Wuhan (Fig.2). With the development of time, there are more and more virus mutations and the haplotype network of virus genome will become more and more complex, which will lead to the unrecognizability of human eyes. Thus, we only selected the data of Saudi Arabia to display, corresponding to the previously mentioned sequences in the method. Nodes represent different groups that were used in the diversity analysis. The branch lengths to integer values so that they represent the number of mutations of COVID-19 belonging to this haplotype. The integers represent the distance between the two haplotypes, the more the integers, the farther the distance (the greater the difference). Current tree shows how sequences like MZ208928, MZ215961, MZ206430 and MZ149269 that had the highest rate of aa mutations are far from the reference as 15 to 26% of their genome has already changed.

SARS-CoV19 amino acid mutation analysis

Mutation is a predictable event in viruses, and SARSCoV-2 is no exception. During 2020, the rapid increase in COVID-19 cases in Saudi Arabia was associated with the emergence of a new variants that were identified by genomic sequencing (Figure 1). These variants are defined by multiple mutations, as showed by Table 4, including those in peak proteins like N501Y, A570D, D614G, P681H, T716I, S982A, and D1118H. Each mutation in the table starts with a letter indicating the effected gene like S for spike and NSP for ORF, e.g. the mutation (Spike_T716I) indicated that T (Threonine) amino acid in position 16th of Wuhan was replaced with an I (Isoleucine) in the query. Whereas, the mutation NSP6_G107del indicated a deletion of Glycine from the position 107 in the query compared to the reference. Other code indicates the following genes, N_ nucleocapsid, M_ membrane, E_ envelope. Indeed, current aa changes analysis showed most of the recorded mutations were on ORF1ab region with a frequency of 52%, followed by the spike, nucleocapsid and finely the envelop (42%, 38% and 25 % respectively).

Fig.1. The evolutionary history was inferred using the Neighbor-Joining method (32). The optimal tree is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (33).

Fig.2. Neighbor-Joining tree built using the haplotype extracted from an aligned sequence of 54 strains including Wuhan strain hCoV-19/Wuhan/WIV04/2019, the branch lengths are set to integer values so that they represent the number of mutations between sequences.

Discussion

The genome of SARS-CoV-2 evolution is relatively stable, thanks to the “proofreading” activity of Nsp12 exonuclease; SARS-Cov-2 mutation rate proved to be slow compared with other RNA viruses (34). Thus, the frequency of derived haplotypes (hfreq) due to new mutations observed on single haplotype backgrounds was shown to be moderate, and sequences carrying such haplotypes were usually restricted to single strain in many of the previous works. However, Broad genomic resources for SARS-CoV-1 played an important role that facilitate optimal drug and vaccine design, especially when coupled with knowledge of human proteome and immune interactions (35). Hence, drugs and vaccines should target relatively invariant, particularly constrained regions of the SARS-CoV-2 genome, to limit both drug resistance and vaccine evasion. Therefore, the ongoing monitoring of genomic variation in SARS-CoV-2 will improve our understanding of fundamental host-pathogen interactions which will be reflected on drug and vaccine design (34). Indeed, this was confirmed in the current study where the recorded SARS-CoV-2 strains in Saudi Arabia can easily be divided into four groups only two of which were close to Wuhan. This might reflect the significant population expansion of Saudi SARS-CoV-2 transmission that was confirmed by both Tajima’s D=-2.486, $P<0.001$; Fu'sFs. Although, differences in the recorded frequency of certain haplotypes in our case may change rapidly due to founder effects caused by local outbreaks. Previous studies showed these changes are rarely invoking selective advantage of SARS-CoV-2 strains (36). However, rapid changes in the frequency of certain haplotypes in different populations may result of founder effects caused by local outbreaks, and this not necessarily invoke selective advantage of such strains. Therefore, unguided assumption about the prevalence of any given SARS-CoV-2 strain indicates its advanced virulence should be avoided, some coronavirus mutations require further investigation due to their global spread (36).

An increased virulence of the coronavirus linked to the hallmark G superclade mutation, p.D614G substitution in the spike protein-A23403G (37,38). Changes in the aa like the p.D614G that was also recorded in the current work is known to facilitate the interaction with the receptor on the surface of human cells is highly possible, even though still not fully confirmed (38–40), other mutated Aa like V121D, V843F, and those in G proteins have the potentials to destabilize or alter the viral protein structure and functions. Amino acid substitution in V843F and G150C took place due to G → T transversion that was possibly introduced by Oxo-guanine generated from reactive oxygen species (ROS) (41,42). Conversely, C → T transition, the most frequent transition of SARS-CoV-2, improved the structural stability of PLPro (42). Besides, the C → T transition in the 5' UTR region influence the function of N and other NSP proteins (43). The positive selective pressure in these proteins could justify some clinical features of SARS-CoV-2 compared with SARS and Bat SARS-like CoV. These are the most common sites undergoing to an aa change, providing an insight of some important proteins of the COVID-2019 that are involved in the mechanism of viral entry and viral replication.

At 30kb, coronaviruses possess the largest genome that include structural and accessory genes, ample replicas, and other non-structural proteins (Nsps) (44). The majority of SARS-CoV-2 genome contains the ORF1a/b region that is translated into 2 polyproteins, pp1a (Nsp1-Nsp11) and pp1ab (Nsp1-Nsp16) (45). Whereas, the other 4 structural proteins that include envelope (E), matrix (M), nucleocapsid (N) phosphoprotein, and spike (S) function together with the viral RNA and Nsp1-16 to assist the replication of the virus within the host cell (46). Mutation in the endosome-associated-protein-like domain of the Nsp2 protein may explain way SARS-CoV-2 virus is more contagious than SARS. Whereas, the mutated Nsp3 proteins that are located near the phosphatase domain may result in a potential mechanism distinguishing COVID-2019 from other viruses like SARS (47). Our result confirmed mutated Aa that effect the V622F and V70F positions that are highly conserved in the viral protein, and G150C or G179C that are mutated in the Saudi strains might reduce the essential flexibility of NSP-3. These non-structural proteins (Nsps) including the Nsp3 protein which is the largest component in the replicase assembly, consists of an ADP-ribose phosphatase (ADRP) domain that is believed to play a key role in altering innate immunity (48). Indeed, Nsp3 has a confirmed phosphatase activity that was shown to removes the 1"phosphate group from Appr-1"-p in in vitro assays (49). Indeed, among the list of mutation we have obtained in this study was one of the most important mutations within the receptor-binding domain (N501Y) that facilitate bind to the human angiotensin-converting enzyme (ACE)2 receptor (50,51). N501Y identified among the list of mutations in the UK variant SARS-CoV-2 VUI (52). Other mutations that changes the spike protein which plays critical roles in receptor binding (S1) and fusion (S2), may alter the cellular tropism (53), were also common in the Saudi strains. The key mutations within spike protein may play a key role in the viral biotype switch (53). Mutations like spike_S982A located in an particularly well-conserved region of S2, a three-nucleotide mutation at position 28280 (nucleocapsid:D3L) which deteriorates the initiation context of ORF9b, and C27972T (ORF8:Q27*) which cuts and presumably inactivates ORF8 (54).

Conclusion

Current study may enhance our understanding of COVID-2019, especially during the ongoing pandemic where scientific community is trying to enrich knowledge about this new viral pathogen. This can be achieved by understanding the main drivers for pathogen appearance, spreading, and supremacy on human defences as these mutations could also adversely affect all efforts of drug repurposing. Thus, the binding mode of drugs may change in these cases.

List of abbreviations

SARS-CoV-2: coronavirus 2, MERS-CoV: Middle East respiratory syndrome coronavirus, Nsp: Non-structural proteins, ORF: Open reading frame, aa: Amino acid, G: Guanine, C: Cytosine, T: Thymine, bp: Base pair.

Consent for publication

Not applicable.

Availability of data and materials

Data are however available from the authors upon reasonable request and with permission of [third party name].

Competing interests

The authors declare that they have no competing interests.

Funding

This work was not funded.

Authors' contributions

Author read and approved the final manuscript.

Acknowledgements

The author extends his appreciation to the University of Northern Border for providing help and support during this work, the appreciation is also linked to Dr Arbi Guetat for his assistance during the preparations and writing of this work.

References

1. Rahman A, Sarkar A. Risk factors for fatal Middle East respiratory syndrome coronavirus infections in Saudi Arabia: Analysis of the WHO Line list, 2013–2018. Am J Public Health. 2019;

2. Memish ZA, Zumla AI, Al-Hakeem RF, Al-Rabeeah AA, Stephens GM. Family Cluster of Middle East Respiratory Syndrome Coronavirus Infections. N Engl J Med. 2013;

3. Palacios Cruz M, Santos E, Velázquez Cervantes MA, León Juárez M. COVID-19, a worldwide public health emergency. Revista Clinica Espanola. 2021.

4. Corman VM, Muth D, Niemeyer D, Drosten C. Hosts and Sources of Endemic Human Coronaviruses. In: Advances in Virus Research. 2018.

5. Frost SDW, Magalis BR, Kosakovsky Pond SL. Neutral theory and rapidly evolving viral pathogens. Mol Biol Evol. 2018;

6. Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, et al. SARS-CoV-2 variants, spike mutations and immune escape. Nature Reviews Microbiology. 2021.

7. Schoeman D, Fielding BC. Coronavirus envelope protein: Current knowledge. Virology Journal. 2019.

8. Corse E, Machamer CE. Infectious Bronchitis Virus E Protein Is Targeted to the Golgi Complex and Directs Release of Virus-Like Particles. J Virol. 2000;

9. Surya W, Li Y, Torres J. Structural model of the SARS coronavirus E channel in LMPG micelles. Biochim Biophys Acta - Biomembr. 2018;

10. Chu DKW, Pan Y, Cheng SMS, Hui KPY, Krishnan P, Liu Y, et al. Molecular Diagnosis of a Novel Coronavirus (2019-nCoV) Causing an Outbreak of Pneumonia. Clin Chem. 2020;

11. Saha P, Banerjee AK, Tripathi PP, Srivastava AK, Ray U. A virus that has gone viral: Amino acid mutation in S protein of Indian isolate of Coronavirus COVID-19 might impact receptor binding, and thus, infectivity. Biosci Rep. 2020;

12. Holmes K V., Enjuanes L. The SARS coronavirus: A postgenomic era. Science. 2003.

13. Marra MA, Jones SJM, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YSN, et al. The genome sequence of the SARS-associated coronavirus. Science (80-). 2003;

14. Zhu Y, Liu M, Zhao W, Zhang J, Zhang X, Wang K, et al. Isolation of virus from a SARS patient and genome-wide analysis of genetic mutations related to pathogenesis and epidemiology from 47 SARS-CoV isolates. Virus Genes. 2005;

15. Pyrc K, Berkhout B, van der Hoek L. Identification of new human coronaviruses. Expert Review of Anti-Infective Therapy. 2007.

16. Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. Cell Host Microbe. 2020;

17. Kuo L, Masters PS. Genetic Evidence for a Structural Interaction between the Carboxy Termini of the Membrane and Nucleocapsid Proteins of Mouse Hepatitis Virus. J Virol. 2002;

18. He R, Leeson A, Ballantine M, Andonov A, Baker L, Dobie F, et al. Characterization of protein-protein interactions between the nucleocapsid protein and membrane protein of the SARS coronavirus. Virus Res. 2004;

19. Rambaut A, Loman N, Pybus O, Barclay W, Barrett J, Carabelli A, et al. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. Virological.org. 2020;

20. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms | Molecular Biology and Evolution | Oxford Academic. Mol Biol Evol. 2018;

21. Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci U S A. 2004;

22. Bodenhofer U, Bonatesta E, Horejš-Kainrath C, Hochreiter S. Msa: An R package for multiple sequence alignment. Bioinformatics. 2015;

23. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 2017;

24. Nei M, Tajima F. Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. Genetics. 1983;

25. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;

26. Wang Y, Li T, Li Y, Bj??rn LO, Rosendahl S, Olsson PA, et al. Community dynamics of arbuscular mycorrhizal fungi in high-input and intensively irrigated rice cultivation systems. Appl Environ Microbiol. 2015;81(8):2958–65.

27. Clement M, Posada D, Crandall KA. TCS: A computer program to estimate gene genealogies. Mol Ecol. 2000;

28. R Core Team. R. R: a Language and Environment for Statistical Computing.[internet].

http://www.R-project.org/. 2018.

29. Leigh JW, Bryant D. POPART: Full-feature software for haplotype network construction. Methods Ecol Evol. 2015;

30. Singer J, Gifford R, Cotten M, Robertson D. CoV-GLUE: A Web Application for Tracking SARS-CoV-2 Genomic Variation. Preprints. 2020;

31. Tamura K, Nei M. Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial-DNA in Humans and Chimpanzees. Mol Biol Evol. 1993;10(3):512–26.

32. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol [Internet]. 1987;4(4):406–25. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3447015

33. Felsenstein J. Confidence-Limits on Phylogenies - an Approach Using the Bootstrap. Evolution (N Y). 1985;39(4):783–91.

34. van Dorp L, Acman M, Richard D, Shaw LP, Ford CE, Ormond L, et al. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. Infect Genet Evol. 2020;

35. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, O'Meara MJ, et al. A SARS-CoV-2-human protein-protein interaction map reveals drug targets and potential drug-repurposing. bioRxiv. 2020;

36. Hryhorowicz S, Ustaszewski A, Kaczmarek-Ryś M, Lis E, Witt M, Pławski A, et al. European context of the diversity and phylogenetic position of SARS-CoV-2 sequences from Polish COVID-19 patients. J Appl Genet. 2021;

37. Brufsky A. Distinct viral clades of SARS-CoV-2: Implications for modeling of viral spread. Journal of Medical Virology. 2020.

38. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell. 2020;

39. Zhang N, Wang L, Deng X, Liang R, Su M, He C, et al. Recent advances in the detection of respiratory virus infection in humans. Journal of Medical Virology. 2020.

40. Volz E, Hill V, McCrone JT, Price A, Jorgensen D, O'Toole Á, et al. Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. Cell. 2021;

41. Gojobori T, Li WH, Graur D. Patterns of nucleotide substitution in pseudogenes and functional genes. J Mol Evol. 1982;

42. Lyons DM, Lauring AS. Evidence for the selective basis of transition-to-transversion substitution bias in two RNA viruses. Mol Biol Evol. 2017;

43. Yang D, Leibowitz JL. The structure and functions of coronavirus genomic 3' and 5' ends. Virus Research. 2015.

44. Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, et al. Severe acute respiratory syndrome-related coronavirus: The species and its viruses – a statement of the Coronavirus Study Group. bioRxiv. 2020;

45. Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. Nature Reviews Microbiology. 2019.

46. Khan MT, Irfan M, Ahsan H, Ahmed A, Kaushik AC, Khan AS, et al. Structures of SARS-CoV-2 RNA-Binding Proteins and Therapeutic Targets. Intervirology. 2021.

47. Angeletti S, Benvenuto D, Bianchi M, Giovanetti M, Pascarella S, Ciccozzi M. COVID-2019: The role of the nsp2 and nsp3 in its pathogenesis. J Med Virol. 2020;

48. Pandey A. Reappraisal of Trifluperidol against NSP-3 protein : Potential therapeutic for COVID-19. Res Sq. 2020;

49. Saikatendu KS, Joseph JS, Subramanian V, Clayton T, Griffith M, Moy K, et al. Structural basis of severe acute respiratory syndrome coronavirus ADP-ribose-1″-phosphate dephosphorylation by a conserved domain of nsP3. Structure. 2005;

50. Wise J. Covid-19: New coronavirus variant is identified in UK. BMJ. 2020;

51. Singh J, Ehtesham NZ, Rahman SA, Hasnain SE. Structure-function investigation of a new VUI-202012/01 SARS-CoV-2 variant. bioRxiv. 2021;

52. Acosta España JD, Arnao Noboa AV, Villacís Ramos IP, Avila Espinoza RE, Davalos de Castro ÁF. SARS-CoV-2 variant VUI 202012/01 (B.1.1.7) in a pediatric patient: first case report in Ecuador. Rev Ecuat Pediatr. 2021;

53. Licitra BN, Millet JK, Regan AD, Hamilton BS, Rinaldi VD, Duhamel GE, et al. Mutation in spike protein cleavage site and pathogenesis of feline coronavirus. Emerg Infect Dis. 2013;

54. Jungreis I, Sealfon R, Kellis M. SARS-CoV-2 gene content and COVID-19 mutation impact by comparing 44 Sarbecovirus genomes. Nat Commun. 2021;