The Bioscientist Journal Vol. 11 (1): 26 - 36, January, 2023
Available online at https://www.bioscientistjournal.com

OPEN ACCESS

Full Text of Paper Presented at the 2nd International Conference on Post-Covid-19 Pandemic Resilience: Role of the Biosciences, 10 – 11 January, 2023, Nnamdi Azikiwe University, Awka, Nigeria

MULTIPLE SEQUENCE ALIGNMENT TO DETERMINE MUTATION IN SPIKE PROTEIN OF DIFFERENT COVID-19VARIANTS

Salisu A[1]., Aminu T. M[1]., Abdulhadi Y[1]., Garba U[1]. & Manya R. N[1]
[1]*Department of Science Laboratory Technology, Jigawa State Polytechnic, 7040 Dutse, Jigawa state, Nigeria*
Correspondence: sahmed@jigpoly.edu.ng

ABSTRACT

The outbreak of the coronavirus disease in 2019 (COVID-19) was caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The disease has been rapidly spread around the globe affecting more than 180 countries. During the COVID-19 pandemic, SARS-CoV-2 has accumulated mutations throughout viral genes encoding the spike (S) glycoprotein D614G mutation became the predominant globally. The potential threats are that SARS-CoV-2 is gradually mutating during incessant transmission among humans. Although most mutations in the SARS-CoV-2 genome are expected to be either deleterious and swiftly purged or relatively neutral, a small proportion will affect functional properties and may alter infectivity, disease severity, or interactions with host immunity. Therefore, this research work is aimed to determine the different mutational changes and phylogenetic relationship among different SARS-CoV-2 variants of the covid-19 virus using a relevant bioinformatics software BioEdit and MEGA 6.0 (Molecular Evolutionary Genetics Analysis version 6.0). The result shows mutation in the positions 491, 621, and 688 in the SARS-CoV-2 spike protein. Such mutations in the S-protein are very critical and might be an associated risk factor of virus transmission. The phylogenetic study has classified the SARS-CoV-2 variants into different countries. This information might provide insight on the impacts on antigenicity and contextualizing them in the S protein structure and observed mutation frequencies in global sequences datasets.

Keywords*:* Amino acids, Covid-19, Mutation, Phylogenetic, SARS-CoV-2

INTRODUCTION

The global outbreak of the coronavirus disease in 2019 (COVID-19) was predominantly caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Wu *et al*., 2020) the disease has been rapidly spread in more than 180 countries worldwide. The virus has been identified with accumulated mutations throughout viral genes encoding the ORF1a, ORF1b, ORF3, ORF8, nucleocapsid (N), Spike proteins, etc. Different type of mutated strains such as beta coronavirus belonging to the *Coronaviridae* family. The first occurrence of this novel coronavirus was observed in Wuhan, China, in December 2019, which later spread globally via human-to-human contact transmission (Lu *et al*., 2020). However, WHO epidemiological update, reported that the virus has predominantly reported in many countries with 28.6 million total confirmed cases and 0.9 million deaths globally. Nevertheless, the onset of COVID-19 sent waves of panic across Nigeria, like in every other country. Due to globalization, the health risk of communicable diseases could be pandemic. The first cases of covid-19 in Nigeria were reported in Lagos State, an Italian citizen who works in Nigeria had returned from Italy (Liu *et al*., 2021). The second case was also confirmed, a Nigerian citizen from Ogun State who had contact with the Italian citizen. Studies confirmed that 139 cases and 2 deaths had been confirmed. The suspected cases that Nigeria were tracing, rose to 6,000 (ICTV, 2020). According to the NCDC, the training of the rapid response teams across the 36 states in Nigeria was concluded in December 2019.

However, NCDC further reported that a Coronavirus Group had been set up to activate its incident system to respond to any emergency. Additionally, the NCDC worked with 22 states in Nigeria to activate their emergency operations centers to manage and link up with the national incidence coordination centers. Although the government had strengthened the surveillance at the airport since January 2020, Nigeria recorded its COVID-19 index case that was imported from Italy, on February 27. This raised concerns about the effectiveness of airport surveillance and, by extension, the country's general preparedness. The index case (an Italian) had visited some other states of the federation before testing positive for COVID-19. The pre-COVID-19 preparedness was grossly inadequate.

Despites, all the global measures against the incessant outbreak of the virus, the ability of the virus to mutate has increase the transmission of the disease among humans globally. Such mutations can occur in the spike (S) protein that binds to the ACE2 receptor and is cleaved by TMPRSS2. Nevertheless, whether S mutations affect SARS-CoV-2 cell entry remains unknown. The naturally occurring S mutations can reduce or enhance cell entry via ACE2 and TMPRSS2. A SARS-CoV-2 S-pseudotyped lentivirus exhibits substantially lower entry than that of SARS-CoV-2 S. Among S variants, the D614G mutant shows the highest cell entry, as supported by structural and binding analyses. Nevertheless, the D614G mutation does not affect neutralization by antisera against prototypic viruses. It has been observed that the D614G mutation increases cell entry by acquiring higher

affinity to ACE2 while maintaining neutralization susceptibility (Zuckerman *et al.,* 2021).

Previous studies confirmed that the mutations in the S protein are crucial because the S protein is key for the first step of viral transmission, i.e, entry into the cell by binding to the angiotensin-converting enzyme-2 (ACE2) receptor (Wu *et al*., 2020), followed by cleavage with transmembrane protease serine 2 (TMPRSS2), both of which are abundantly expressed in not only the airways, lungs, and nasal/ oral mucosa, but also the intestines. Conversely, it remains unclear whether S mutations affect SARS-CoV-2 cell entry. It has been investigated that ACE2/TMPRSS2 usage of SARS-CoV S and SARS-CoV-2 S, compare cell entry of natural SARS2-S variants, and focus on the D614G variant by investigating its structure. ACE2- binding affinity, and neutralization susceptibility. Molecular studies confirmed that the D614G mutation confers increased entry efficiency resulting from enhanced binding affinity for ACE2 with no influence on the antigenicity of the S protein. (Valine-to-Isoleucine at position 1216 and Leucine-to-Methionine at position 1233) amino acid differences in the CT and TM domains, respectively. Hence, the created a SARS2-S C1247A mutant and a chimeric SARS2-S harboring the TM/CT domains of SARS-S. All SARS2-S proteins showed comparable levels of cell entry and actual virion incorporation. These results suggest that the significantly lower rate of cell entry of the SARS2-S pseudovirus was not due to the incompatibility between SARS2-S and a lentiviral vector but rather to the intrinsic nature of the SARS2-S protein. To further assess differences between the SARS-S and SARS2-S proteins, there is need to address whether these S proteins might differ in their ability to utilize a given level of cell-surface ACE2 or TMPRSS2. Based on lentiviral infection of 293T cells expressing a high and constant level of ACE2 together with a range of expression levels of TMPRSS2 and vice versa. SARS2- S-mediated infection required higher levels of cell-surface TMPRSS2 expression than SARS-S to attain maximum levels of cell entry. Equally, the mutation in the spike protein of SARS-CoV-2 is still uncertain and the implication of the virus might affect SARS-CoV-2 cell entry. The spike protein mediates attachment of the virus to host cell- surface receptors and fusion between virus and cell membranes. It is also the principal target of neutralizing antibodies generated following infection by SARS-CoV-2 and is the SARS- CoV-2 component of both mRNA and adenovirus- based vaccines. Consequently, mutations that affect the antigenicity of the spike protein are of particular importance.

Therefore, the main objectives of this study were to identify mutations and their antigenic consequences, focusing on the spike protein among different global SARS-CoV-2 variants and to observed mutation frequencies in world genomic sequence.

MATERIAL AND METHODS

To determine the different mutational changes that may occur at different variants of covid-19 virus. Relevant bioinformatics software BioEdit (USA version.7.0.5.3; Therapeutics, Ibis) and MEGA 6.0 (Molecular Evolutionary Genetics Analysis version 6.0) was used to

determine the mutations and phylogenetic relationship between different SARS-CoV-2 variants.

Nucleotide Sequencing and Analysis of SARS-CoV-2 Different Variants

A total of 25 nucleotide sequences of Spike protein of different SARS-CoV-2 variants, was retrieved from the NCBI BLAST Gen Bank. Multiple sequence alignment was carried out from those sequences with the reference sequence of WUHAN NC_045512.2. The nucleotide sequences were also translated to amino acids and the mutation was observed based on the amino acids' substitutions (Salisu *et al*., 2021).

Construction of Phylogenetic Tree

Phylogenetic tree was constructed from the 25 nucleotide sequences retrieved from NCBI. All aligned sequences were exported to BioEdit (USA version.7.0.5.3; Therapeutics, Ibis) and the phylogenetic tree was constructed using MEGA 6.0 Distance- based neighbour joining with 1,000 Bootrap replicates used to evaluate the evolutionary distances.

DISCUSSION

The causative agent of the COVID-19 pandemic, SARS-CoV-2, is steadily mutating during continuous transmission among humans. Such mutations can occur in the spike (S) protein. In an attempt to determine the periodic mutations in the SARS-CoV-2 spike protein different sequences of SARS-CoV-2 spike protein was retrieved from NCBI and subjected to different Bioinformatics software's. From the results mutation was identified in the three (3) different positions Glutamate-491-Lysine, Aspatate-621-Glysine and Proline-688-Histidine. This mutational change observed in the spike protein of the SARS-CoV-2 variants has distinguished the virusesinto different geographical distribution. However, based on the molecular biology of the virus, the mutation exists between the reference strain (Wuhan-045512.2) indicate that there is amino acid substitution in position 421 from Glutamate to Lysine in the USA, UK, and Brazil strains this might further differentiate them base on the virus virulence nature. Furthermore, this trend of amino acids changes confirmed that the mutation in SARS-CoV-2 spike protein is consistent from the emergence of virus in 2019 to date.

It is critical to understand those mutations in the SARS-CoV-2 variants because the S protein play an important role in the virus internalization into the cell and binding angiotensin-converting enzyme-2 (ACE2) receptor1 and this might affect the process of diagnosis and therapeutics of the Covid-19 virus. However, this work is in line with wok done by (Zhang *et al*., 2021) whose reported that mutations in the S protein is crucial because the S protein is key for the first step of viral transmission, i.e., entry into the cell by binding to the angiotensin-converting enzyme-2 (ACE2) receptor1, followed by cleavage with transmembrane protease serine 2 (TMPRSS2).

Moreover, the result also confirmed mutation in position Aspatate-621-Glysisne and Proline-688-Histidine respectively. And the work is in corroboration with work done by Shang *et al*., 2020 whose reported mutation in position D614G mutation in the S protein.

RESULT

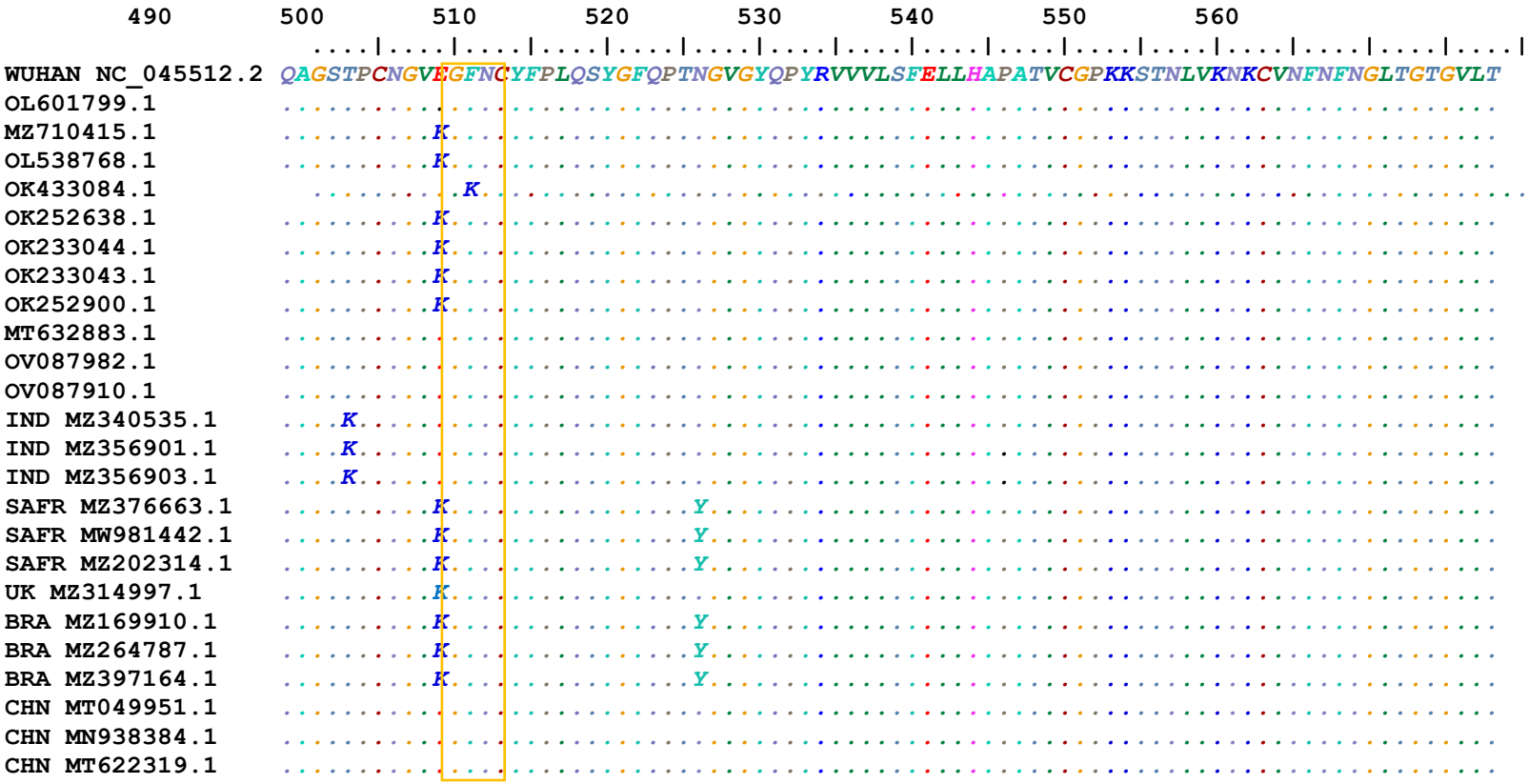

Figure 1a: Multiple sequence alignment of Covid-19 Spike protein amino acids sequences of the global SARS-CoV-2 variants to determine the pattern of mutation in the Spike protein sequences. A dot (.) indicate position where the sequences a identical to other variants, the colour indicate SARS-CoV-2 variants, China variants (yellow), UK variant (orange), Indian variants (green), UK variants (dark blue), Brazil variants (purple), China variants (grey) and USA (green)

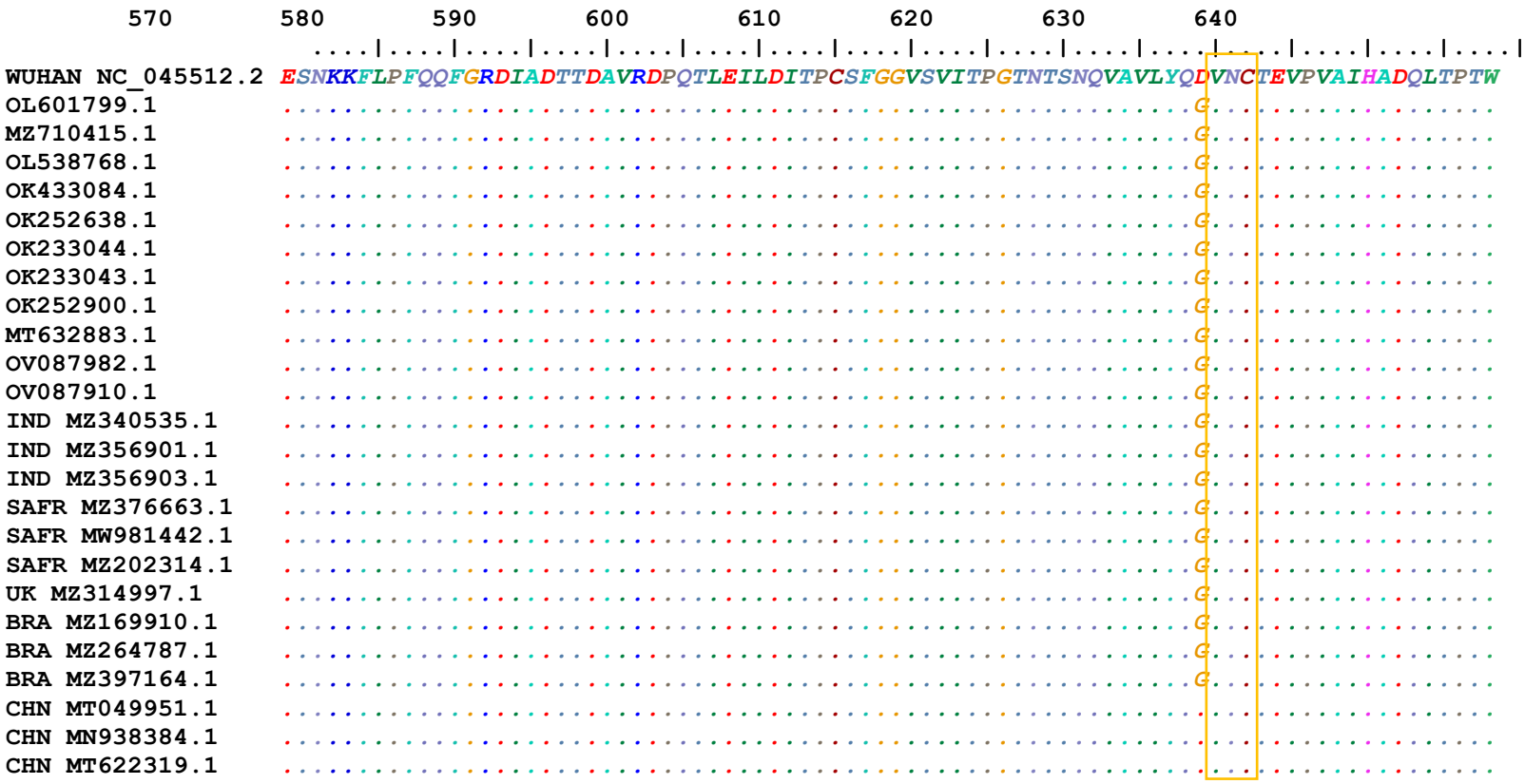

Figure 1b: Multiple sequence alignment of Covid-19 Spike protein amino acids sequences of the global SARS-CoV-2 variants to determine the pattern of mutation in the Spike protein sequences. A dot (.) indicate position where the sequences a identical to other variants, the colour indicate SARS-CoV-2 variants, China variants (yellow), UK variant (orange), Indian variants (green), UK variants (dark blue), Brazil variants (purple), China variants (grey) and USA (green)

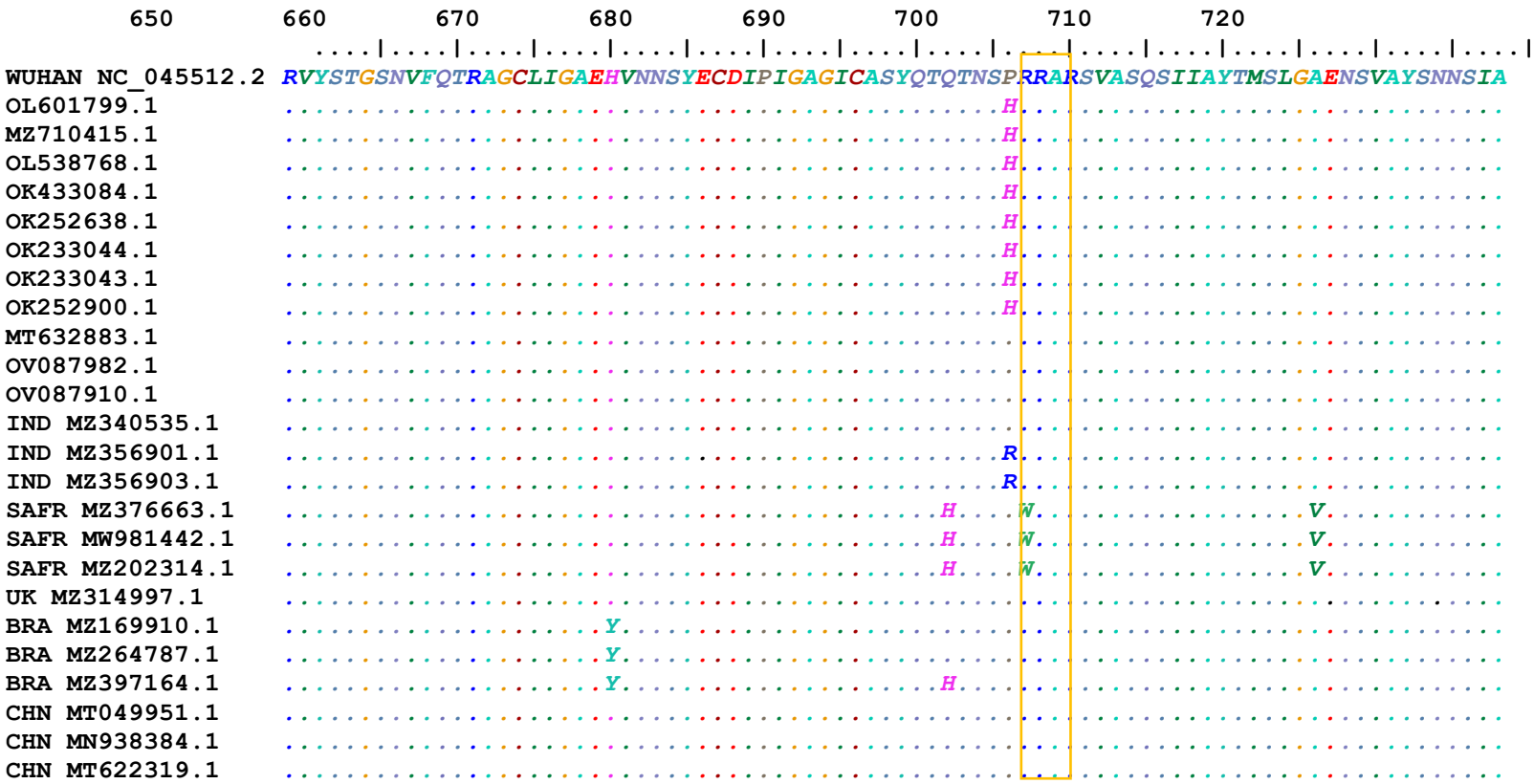

Figure 1c: Multiple sequence alignment of Covid-19 Spike protein amino acids sequences of the global SARS-CoV-2 variants to determine the pattern of mutation in the Spike protein sequences. A dot (.) indicate position where the sequences a identical to other variants, the colour indicate SARS-CoV-2 variants, China variants (yellow), UK variant (orange), Indian variants (green), UK variants (dark blue), Brazil variants (purple), China variants (grey) and USA (green)

Table 3.2: Amino acids substitutions in the Spike protein of SARS-CoV-2 variants retrieved from the NCBI GenBank

S/N	Accession Number	Country	Amino acid (aa) substitution and position		
			491	688	621
1	**WUHAN***	CHINA	E	D	P
2	**OL601799.1**	USA	**E**	**G**	**H**
3	**MZ710415.1**	USA	K	G	H
4	**OL538768.1**	USA	K	G	H
5	**OK433084.1**	USA	K	G	H
6	**OK252638.1**	USA	K	G	H
7	**OK233044.1**	USA	K	G	H
8	**OK233043.1**	USA	K	G	H
9	**OK252900.1**	USA	K	G	H
10	**MT632883.1**	USA	E	G	P
11	**OV087982.1**	USA	E	G	P
12	**OV087910.1**	USA	E	G	P
13	**MZ340535.1**	INDIA	E	G	P
14	**MZ356901.1**	INDIA	E	G	R
15	**MZ356903.1**	INDIA	E	G	R
16	**MZ376663.1**	UK	K	A	P
17	**MW981442.1**	UK	K	G	P
18	**MZ202314.1**	USA	K	G	P
19	**MZ314997.1**	UK	E	G	P
20	**MZ169910.1**	BRAZIL	K	G	P
21	**MZ264787.1**	BRAZIL	K	G	P
22	**MZ397164.1**	BRAZIL	K	G	P
23	**MT049951.1**	CHINA	E	D	P
24	**MN938384.1**	CHINA	E	D	P
25	**MT622319.1**	CHINA	E	D	P

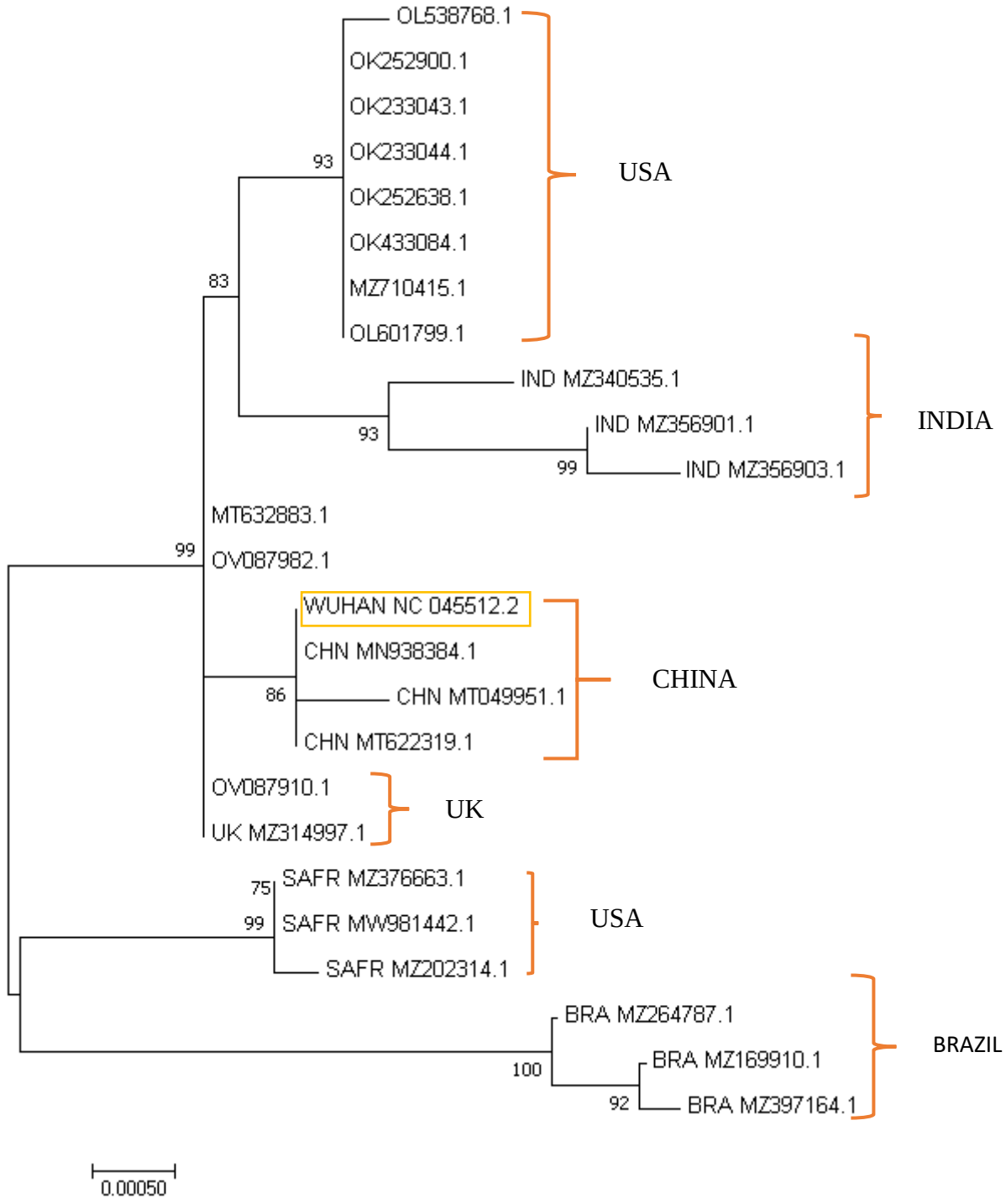

Figure 2: Phylogenetic analysis of Spike protein of SARS-CoV-2nucleotide sequence based on the 25 references isolates are labeled. At major nodes are indicated bootstrap values. Wuhan isolates in red box

Phylogenetically based on the 25 nucleotide sequences of the SARS-CoV-2 spike protein the result has categorized them into five different continents (China, India, UK, USA A and Brazil). The reference sequence was closely related to UK variants (86%) and is distantly related to USA variants. Interestingly USA and India variant a closely related.

CONCLUSION

Based on the bioinformatics study, mutation was observed in the position 491, 621 and 688 in the SARS-CoV-2 spike protein. Such mutation in the S-protein is very critical and might be an associated risk factors of virus transmission.

The phylogenetic study has classified the SARS-CoV-2 variants into different countries. This information might provide the researchers on the evolutionary relationship that exist between global SARS-CoV-2 variants strains.

Recommendation

In order to reduce the spread of the SARS-CoV-2 variants. intense molecular study has to be employed in order to understanding the factors responsible for mutation in the virus. Researchers currently working on vaccine development may use stable nucleotide sequences in the spike protein of the SARS-CoV-2.

REFERENCE

International Committee on Taxonomy of Viruses Executive Committee. The New Scope of Virus Taxonomy: Partitioning the Virosphere into 15 Hierarchical Ranks. *Nat Microbiol*;5(5):668-674

Liu, H., Zhang, Q., Wei, P., Chen, Z., Aviszus, K., Yang, J., Downing, W., Peterson, S., Jiang, C., Liang, B., Reynoso, L., Downey, G.P., Frankel, S.K., Kappler, J., Marrack, P and Zhang G. (2021). The Basis of a more Contagious 501Y.V1 variant of SARS-COV-2. BioRxiv ; [PMC free article] [PubMed]

Lu, I.N., Muller, C. P and F. Q. He. (2020). Virus Res. **283**, 197963 https://doi.org/10.1016/j.Viruses.

Salisu, A., Abdul Razak, M., Mohd H B., Abdul Rahman O., Aini, I and Nurulfiza, M, I.(2021).Molecular Markers and Phylogenetic Analysis of UPMT27, a Field Isolate of the Malaysian Fowl Adenovirus Associated with Inclusion Body Hepatitis. *Pertanika Journal of Science & Technology* 29 (1); 547 – 563

Wu, K., Werner, P., Moliva, J.I., Koch, M., Choi, A., Stewart-Jones, G.B.E., Bennett, H., Boyoglu-Barnum,S., Shi, W., Graham, B.S., Carfi, A., Corbett, K.S., Seder, R.A and Edwards, D.K. (2021) :mRNA-1273 Vaccine Induces Neutralizing Antibodies against Spike Mutants from Global SARS-CoV-2 Variants. *bioRxiv*. .01.25.42

Wu, Fan, Su, Zhao, Yu, Bin, Chen, Yan-Mei, Wen, Wang, Song, Zhi-Gang, Hu, Yi, Tao, Zhao-Wu. (2020). A new coronavirus associated with human respiratory disease in China. Nature, 579 (7798), 265–269.

Zhang W, Davis BD, Chen SS, Sincuir Martinez JM, Plummer JT, Vail E.

Emergence of a Novel SARS-CoV-2 Variant in Southern California. *JAMA*;325(13):1324-1326.

Zuckerman NS, Bucris E, Drori Y, Erster O, Sofer D, Pando R. (2021). Genomic Variation and Epidemiology of SARS-CoV-2 Importation and Early Circulation in Israel. *PLoS ONE* 16(3)