HSI Health Sciences Investigations Journal

share

Send us an email: hsijournal@ug.edu.gh
Visit us: https://www.hsijournal.ug.edu.gh

ISSN Online 2704-4890 | ISSN Print 2720-7609

Online first publication

Original Research Article

HSI Journal (2023) Volume 4 (Issue 1):443-447. https://doi.org/10.46829/hsijournal.2023.6.4.1.443-447

Open Access

Human SARS CoV-2 spike protein mutations in West Africa

Samuel O OLALEKAN [1]*, Muinat M ADEYANJU [2], Ifabunmi O OSONUGA [1], Babatunde TAYO [3]

[1]Department of Physiology, Olabisi Onabanjo University, Sagamu Campus, Sagamu, Ogun State, Nigeria; [2] Department of Biochemistry, Olabisi Onabanjo University, Sagamu Campus, Sagamu, Ogun State, Nigeria; [3] Department of Medical Microbiology and Parasitology, Babcock University Teaching Hospital, Ilishan, Ogun State, Nigeria.

Received June 2022; Revised September 2022; Accepted January 2023

Abstract

Background: The COVID-19 pandemic was caused by the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), first detected in Wuhan, Hubei province, China in December 2019. The virus rapidly spread worldwide, with mutations in various parts of its genetic material affecting its transmissibility and infectivity.

Objective: This study addressed some of the mutations present in the human SARS-CoV-2 spike proteins relative to Wuhan-Hu-1 reference sequence from China, according to different countries from West Africa

Methods: The SARS-CoV-2 spike protein sequences were obtained from the National Center for Biotechnology Information virus database in the FASTA format on November 12, 2021. The multiple sequence alignment of the proteins was carried out by MAFFT version 7 online. The human SARS-CoV-2 spike protein sequences from selected West African countries were analyzed by comparing them with the reference SARS-CoV-2 protein sequence from Wuhan-Hu-1, China.

Results: Out of 148 spike protein sequences analyzed, 137 proteins had one or more mutations. A total of 486 mutations were observed corresponding to 47 distinct mutation sites. In the analysis of the spike proteins in the study, it was observed that the Receptor Binding Domain which is involved in the interactions with human angiotensin-converting enzyme-2 (ACE-2) receptor causing infection leading to the COVID-19 disease had 8 distinct mutation sites. The D614G mutation is the most common in the SARS-CoV-2 spike protein observed so far among all the West African countries examined in this study and thus the most predominant. In this study, we examined spike proteins not associated with mutations, the distribution of mutations in spike proteins, mutation density in different regions of the spike protein sequence, spike protein sequences with multiple mutations and the Human SARS-CoV-2 spike protein mutation in West Africa and implications for vaccination and drug development purposes.

Conclusion: The identified mutations in SARS-CoV-2 are significant for infection prevention, control, and public health interventions. Further studies are imperative to understand the mutations in the virus's spike proteins to guide vaccine development and antiviral drug designs. Investigations should also be conducted to determine the infectivity of emerging variants in West Africa and their response to vaccines and available drugs to address public health concerns on vaccination and drug design goals.

Keywords: Mutations, spike proteins, West Africa, SARS-CoV-2

Cite the publication as Olalekan SO, Adeyanju MM, Osonuga IO, Tayo B (2023) Human SARS CoV-2 spike protein mutations in West Africa. HSI Journal 4 (1):443-447. https://doi.org/10.46829/hsijournal.2023.6.4.1.443-447

INTRODUCTION

The epicentre of the ongoing Coronavirus disease 2019 (COVID-19) pandemic caused by the human severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) was first identified in the city of Wuhan-Hubei-1 province, China in mid-December 2019 [1]. Since then, the disease has spread throughout the world leading to several deaths. In late February 2020, Nigeria recorded the first known case of COVID-19 in West Africa and just within a month, several cases were recorded in all 17 countries in the region. As of 11 October 2021, the total confirmed cases in West Africa were 695,027, with 521,413 recoveries and

* *Corresponding author*
Email: samuel.olalekan@oouagoiwoye.edu.ng

share
Send us an email: hsijournal@ug.edu.gh
Visit us: https://www.hsijournal.ug.edu.gh

10,355 total deaths [2]. The SARS-CoV-2 belongs to the family of Coronaviridae, a subfamily of *Orthocoronavirinae,* and the genus *Betacoronavirus* [3]. The spread of the disease is attributed to contact via respiratory droplets from breathing, talking, coughing, and sneezing. The SARS-CoV-2 is a positive-strand RNA virus whose 30-kilobase genome encodes four structural proteins: spike protein (S), small protein (E), matrix (M), and nucleocapsid (N) [4]. The spike protein is a homotrimer present on the surface of the coronavirus and is essential in the recognition of the human angiotensin-converting enzyme-2 (ACE-2) receptor [5]. This recognition precedes the fusion of the virus and it hosts cellular membranes for the transfer of the viral nucleocapsid into the host cells. The spike protein is a type I fusion protein that forms trimers on the surface of the virion. The structure consists of two subunits: the N-terminal S1, which is accountable for binding to receptors, and the C-terminal S2, which is located near the membrane and facilitates the process of membrane fusion [6,7]. The S1 subunit consists of $S1^{A}$, $S1^{B}$, $S1^{C,}$ and $S1^{D}$ domains. The $S1^{B}$ domain also called the receptor-binding domain (RBD) of the SARS-CoV-2 spike protein, interacts with the human ACE-2 receptor [6].

Mutations distal from the RDB influence the transmissibility of SARS-CoV-2. The structural elements within the S2 subunit comprise three long α-helices, multiple α-helical segments, extended twisted β-sheets, a membrane-spanning α-helix, and an intracellular cysteine-rich segment. There are four amino acid motif sequences located between the S1 and S2 subunits in SARS-CoV-2 at the furin-cleavage site [8]. In the S2 subunit, a second proteolytic cleavage site S20, upstream of the fusion peptide is present. Both these cleavage sites participate in the viral entry into host cells. In a study on the infectivity and reactivity to a panel of neutralizing antibodies and sera from convalescent patients, mutations and glycosylation site modifications were reported in human SARS CoV-2 spike proteins [9]. The D614G mutation is reported to be relatively more common and is known to increase the efficiency of causing infection [9]. This mutation appears to increase viral infectivity by favouring the open RBD conformational state, and it is now a major virus variant globally [11-13]. It is therefore imperative to study new mutations that are appearing. This study addressed some of the mutations present in the human SARS-CoV-2 spike proteins relative to the Wuhan-Hu-1 reference sequence from China, according to countries from West Africa.

MATERIALS AND METHODS

The SARS-CoV-2 spike protein sequences were obtained from the National Center for Biotechnology Information (NCBI) virus database in the FASTA format [14] on 12 November 2021. The multiple sequence alignments of the proteins were carried out by MAFFT online service with default parameters (https://mafft.cbrc.jp/alignment/server/) [15]. The human SARS-CoV-2 spike protein sequence from Wuhan-Hu-1, China (NCBI accession code: YP_009724390.1) [1] was used as a reference sequence to determine the mutations. The mutations were analyzed according to their presence within different regions of the spike protein sequences. This study examined mutations that were found in seven West African countries because there was no available data on mutations from other West African countries on the NCBI database. Specifically, the study focused on mutations in the spike protein in the following countries: Benin Republic, Gambia, Ghana, Guinea, Mali, Nigeria, and Sierra Leone.

RESULTS AND DISCUSSIONS

Distribution of mutations in spike proteins

The NCBI virus database contained the human SARS-CoV-2 spike proteins that were analyzed. These represented spike proteins from different countries in West Africa including Benin Republic (n = 12), Gambia (n = 6), Ghana (n = 59), Guinea (n = 11), Mali (n = 2), Nigeria (n = 2) and Sierra Leone (n = 56). The length of the protein sequences used in this study ranged from 1270 to 1273 amino acid residues. Overall, 148 protein sequences were examined in this study. One or more mutations were observed in 137 proteins. A total of 486 mutations were observed corresponding to 47 distinct mutation sites. These mutation sites were situated in the following positions: 5, 9, 12, 22, 69, 70, 76, 80, 90, 113, 138, 144, 152, 176, 182, 281, 385, 439, 452, 478, 484, 510, 511, 570, 580, 583, 614, 655, 675, 677, 681, 688, 689, 716, 769, 796, 946, 965, 982, 1078, 1118, 1144, 1155, 1157, 1219, 1228, 1235.

The distribution of the total number of mutations in spike proteins analyzed from the different West African countries is represented in Table 1. Nearly 92.5% (n = 137/148) of the spike protein sequences were associated with mutations. The list of the mutation sites along with the total number and position of mutations observed at individual mutation sites is presented in Table 2. The top 4 mutation sites according to the total number of occurrences were D614G (n = 133), E484K (n = 71), G796Y(n = 70), and W152L (n = 66). Table 2 also shows the distribution of mutations in the different regions of the human SARS-CoV-2 spike proteins in West Africa.

Table 1: Distribution of human SARS-Cov-2 spike proteins in some West African Countries

Country	Number of spike proteins	Number of distinct mutations	Number of mutations
Benin Republic	12	4	11
Gambia	6	1	5
Ghana	59	39	193
Guinea	11	21	72
Mali	2	1	1
Nigeria	2	1	2
Sierra Leone	56	22	202
Total	148	89	486

share
Send us an email: hsijournal@ug.edu.gh
Visit us: https://www.hsijournal.ug.edu.gh

Mutation density in the spike protein sequence

Mutations were distributed in almost all regions of the protein. The $S1^D$ (positions 594 to 674) that comprises the D614G mutation was the most predominant and was observed in 133 of the 136 mutations in this region of the spike protein. The analysis in this study is in agreement with the high frequency of D614G mutations in the spike protein as reported previously [10]. The downward helix (positions 738 to 782) that comprise the G769V mutation was observed in 70 out of the 71 mutations in this region of the spike protein. Likewise, the S1A domain (postions 1-302) that comprises the W152L mutation was observed in 66 out of the 120 mutations in this region of the spike protein. Figure 1 displays the evaluation of mutation density as a function of the number of mutations observed over the sequence length for various regions in the spike protein. The protease cleavage site in the spike protein has the highest mutation density. The mutations at this site in the spike protein may be of benefit to the virus in proteolytic cleavage by a large number of host enzymes. Of

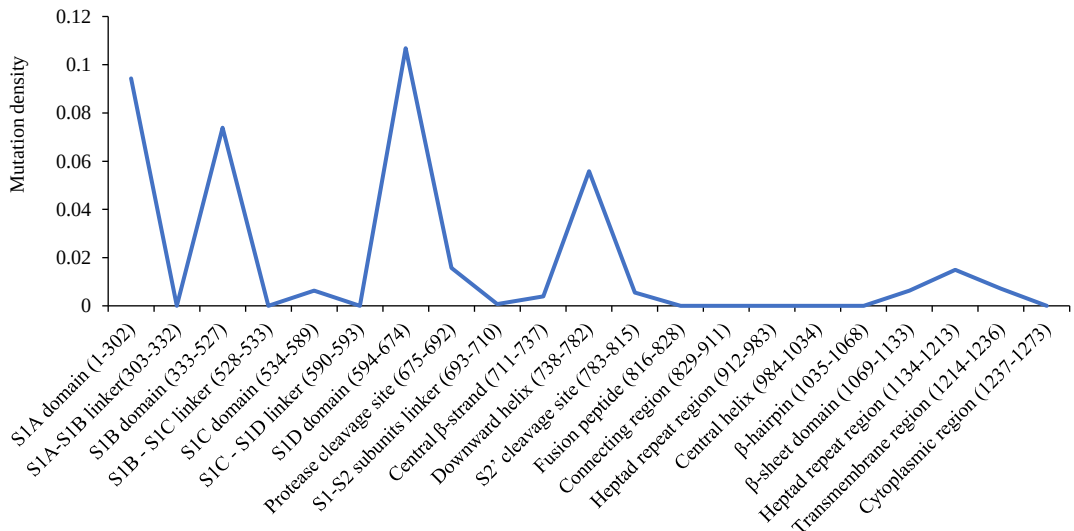

Figure 1: Mutation density in various human SARS-CoV-2 spike protein regions

Table 2: Distribution of mutations in the different regions of human SARS-CoV-2 spike proteins

Regions of Spike proteins	Total number of mutations	Number of distinct mutation types	Position of distinct mutation types
S1A domain (1-302)	120	23	L5F, P9S, S12C, L18F, T22I, H69del, V70del, T76N, T76I,D80Y, V90L, K113N, D138H, D138Y, G142D, Y144del, S151I, W152F, W152L, E154K, L176F, K182R, E281V
S1A-S1B linker (303-332)	-	-	-
S1B domain (333-527)	94	8	T478K, L452R, T385I, E484K, V511G, N501Y, V510T, N439K
S1B - S1C linker (528-533)			
S1C domain (534-589)	8	3	A570D, E583D, E580Q
S1C - S1D linker (590-593)	-	-	-
S1D domain (594-674)	136	3	D614G, A653V, H655Y
Protease cleavage site (675-692)	20	6	Q677H, Q675H, A688V, S689R, S689L, P681H
S1-S2 subunits linker (693-710)	1	1	S698L
Central β-strand (711-737)	5	1	T716I
Downward helix (738-782)	71	2	T761I, G769V
S2' cleavage site (783-815)	7	2	D796Y, D796H
Fusion peptide (816-828)	-	-	-
Connecting region (829-911)	-	-	-
Heptad repeat region (912-983)	9	3	Q965H, S982A, G946V
Central helix (984-1034)	-	-	-
β-hairpin (1035-1068)	-	-	-
β-sheet domain (1069-1133)	8	2	D1118H, A1078V
Heptad repeat region (1134-1213)	19	4	K1157M, E1202Q, Y1155F, E1144Q
Transmembrane region (1214-1236)	9	4	C1235F, G1219C, G1219V, V1228L
Cytoplasmic region (1237-1273)	-	-	-

share
Send us an email: hsijournal@ug.edu.gh
Visit us: https://www.hsijournal.ug.edu.gh

note, the N-terminal domain ($S1^A$) is also another region where mutations have accumulated relatively more in number compared with the rest of the spike protein.

Table 3: Mutation sites and mutation types observed in human SARS-CoV-2 spike proteins in West African countries.

West African Countries	Mutation types present in SARS-CoV-2 spike protein
Benin republic	V90L, E281V, E583D, D614G
Gambia	D614G
Ghana	L5F, L18F, T22I, T76I, D80Y, K113N, D138H, S151I, W152F, W152L, E154K, L176F, T385I, N439K, L452R, E484K, N501Y, V510T, E583Q, D614G, A653V, H655Y, Q675H, Q677H, A688V, S689R, S698L, G769V, D796H, D796Y, G946V, S982A, A1078V, D1118H, Y1155F, E1202Q, G1219C, G1219V, V1228L
Guinea	P9S, S12C, H69del, V70del, I76N, D138Y, Y144-, K182R, L452R, T478K, N501Y, V511G, A570D, D614G, P681H, T716I, D796Y, Q965H, S982A, D1118H, C1235F
Mali	D614G
Nigeria	D614G
Sierra Leone	T76I, D80Y, G142D, S151I, W152F, W152L, E154K, K182R, L452R, T478K, E484K, E580Q, D614G, Q677H, P681H, G769V, D796Y, D1118H, E1144Q, K1157M, E1202Q, C1235F

Human SARS-CoV-2 spike protein mutation

Accordingly, the total number of mutation sites observed were: Benin Republic (n = 11), Gambia (n = 5), Ghana (n = 193), Guinea (n = 72), Mali (n = 1), Nigeria (n = 2) and Sierra Leone (n = 202). It is clear from our study that the human SARS-CoV-2 spike protein undergoes mutations at multiple sites and there can be more than one mutation type associated with a mutation site. The D614G is the only mutation that has so far been commonly observed among the spike proteins from all the countries in West Africa. The 137 spike proteins that mutated had anywhere from 1 to 13 mutations. The spike proteins containing eight or more mutations in the same protein are discussed below.

The spike protein identified by the NCBI accession code QXT61053.1 corresponds to SARS-CoV-2 isolated from Guinea, Africa, on March 17, 2021, and contains 13 mutations. The mutations S12C, H69del, V70del, T76N, Y144del, N501Y, A570D, D614G, P681H, T716I, Q965H, S982A, D1118H are located at different positions that lie scattered in the spike protein with three deletions at positions 69, 70, and 144. The spike protein with NCBI accession code QXT61047.1 with 10 mutations was isolated on March 10, 2021, from Guinea with mutations P9S, H69del, V70del, Y144del, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H. Both proteins with NCBI accession codes: QXT61053.1 and QXT61047.1 have deletions at 69, 70, and 144 corresponding to the $S1^A$ domain with an RBD domain mutation at position N501Y. There is also a mutation corresponding to the protease cleavage site P681H. Table 3 indicates the presence or absence of spike protein mutation(s) within and between the various West African countries included in this study. The RBD in spike protein interacts with the host ACE-2 receptor to cause the novel coronavirus infection leading to COVID-19 disease [7, 8]. The RBD of human SARS COV-2 spike proteins from the different West African countries is associated with eight distinct mutation sites. The mutations are located at positions 385, 439, 452, 478, 484, 501, 510, and 511. Accordingly, the total number of distinct mutation sites observed in the RBD were Benin Republic (n = 0), Gambia (n = 0), Ghana (n = 6), Guinea (n = 4), Mali (n = 0), Nigeria (n = 0) and Sierra Leone (n = 3). The D614G + L5F mutation was present in two spike proteins from Ghana (NCBI accession codes QRG27565.1 and QRG27541.1) and is associated with increased infectivity. Two additional spike proteins, D614G and Q675H mutants, originating from Ghana (identified by NCBI accession codes QRG27589.1 and QRN78419.1) have been associated with lower levels of infectivity. Two spike protein mutations that have been found to have decreased sensitivity to neutralizing monoclonal antibodies are N439K and L452R. The N439K mutation is present in spike proteins with NCBI accession code QRN78443.1 from Ghana and has been linked to decreased sensitivity to convalescent sera [9]. The L452R mutation is present in spike proteins with NCBI accession codes QRN78251.1, QRN78203.1, QRN78287.1, and QSW63813.1 from Ghana and QUX03825.1 from Togo.

Spike proteins not associated with mutations

Of the total 148 proteins examined in this study, 11 spike proteins did not show mutations relative to the human SARS-CoV-2 spike protein reference sequence from Wuhan-Hu-1, China. These include the spike proteins from Benin Republic (with NCBI accession numbers QKX45633.1, QKX45621.1, QKX45609.1, and QKX45597.1), Gambia (with accession number QUI20763.1), Ghana (with NCBI accession number QNE11577.1), Mali (with NCBI accession number QQO98466.1), and Sierra Leone (with NCBI accession numbers QMX85013.1, QMX85001.1, QMX85109.1, and QMX85073.1).

Conclusion

The SARS-CoV-2 virus gains access to host cells through spike proteins on its surface which help the virus to recognize and then bind to the human ACE-2 receptor. Vaccine productions and drug designs have therefore targeted the binding site of the SARS-CoV-2 virus at the point of protein-protein interaction. Alterations in the shape and charge of proteins around the receptor binding areas of the spike protein will have public health implications such as the reduced response of the SARS-CoV-2 virus to vaccination and drug design efforts. This study identified eight distinct mutation sites in the RBD, which interacts with human ACE-2 receptors and causes COVID-19 infections. The mutations were located at positions 385,

439, 452, 478, 484, 501, 510, and 511. The presence of the mutations identified in the SARS-CoV-2 has significance in infection prevention and control as well as public health interventions. Further implications of the mutations identified include those related to testing and drug design and administration. It is therefore imperative to consider further studies to better understand mutations in the spike proteins of the human SARS-CoV-2 virus to guide vaccine development, infection prevention, and antiviral drug designs. Further studies should also be carried out to investigate the infectivity of emerging variants of the SARS-CoV-2 virus in West Africa and how they respond to vaccines and available drugs targeted toward the virus.

DECLARATIONS

Ethical considerations
Ethics approval was not needed for the study as it analyzed non-identifiable patient data obtained from the publicly available NCBI virus dataset at https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/.

Consent to publish
All authors agreed to the content of the final paper.

Funding
This research was conducted with personal funds, with no external sponsorship or financial support.

Competing Interests
No potential conflict of interest was reported by the authors.

Author contributions
SOO conceptualized, designed, collected, analyzed data, and drafted the manuscript. MMA contributed to the study design, provided feedback, and assisted with revisions. IOO helped in the design, data collection, and provided feedback. BT contributed to data analysis, provided feedback, and assisted with revisions. All authors approved the final manuscript.

Acknowledgements
The authors thank the research assistants at the Department of Physiology, Olabisi Onabanjo University, Sagamu Campus, Sagamu, Ogun State, Nigeria, for their valuable contributions. We acknowledge the reviewers for their helpful suggestions that enhanced the manuscript's quality.

Availability of data
Data for this work is available upon reasonable request from the corresponding author.

REFERENCES

1. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ (2020) A new coronavirus associated with human respiratory disease in China. Nature 579:265-269. https://doi.org/10.1038/s41586-020-2008-3
2. Sahel and West Africa club Secretariat. Tackling the Coronavirus (COVID-19): West African perspectives. https://www.oecd.org/swac/coronavirus-west-africa/ [Accessed 5 May 2023]
3. Taxonomy Browser Severe acute respiratory syndrome-related coronavirus. https://www.ncbi.nlm.nih.gov/taxonomy/694009 [Accessed 5 May 2023]
4. Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, Yuen KY (2020) Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbes Infect 9:221–236. https://doi.org/10.1080/22221751.2020.1719902
5. Zhang H, Penninger JM, Li Y, Zhong N, Slutsky AS (2020) Angiotensin-converting enzyme 2 (ACE-2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. Intensive Care Med 46:586-590. https://doi.org/10.1007/s00134-020-05985-9
6. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, and Veesler D (2020) Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell 181: 281–292. https://doi.org/10.1016/j.cell.2020.02.058
7. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, Graham BS, McLellan JS (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 367: 1260–1263. https://doi.org/10.1126/science.abb2507
8. Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, Guo L, Guo R, Chen T, Hu J, Xiang Z, Mu Z, Chen X, Chen J, Hu K, Jin Q, Wang J, Qian Z (2020) Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nat Commun 11:1620. https://doi.org/10.1038/s41467-020-15562-9
9. Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, Zhao C, Zhang Q, Liu H, Nie L, Qin H, Wang M, Lu Q, Li X, Sun Q, Liu J, Zhang L, Li X, Huang W, Wang Y (2020) The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. Cell 182:1284-1294. https://doi.org/10.1016/j.cell.2020.07.012
10. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi EE, Bhattacharya T, Foley B, Hastie KM, Parker MD, Partridge DG, Evans CM, Freeman TM, de Silva TI (2020) Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 Virus. Cell 182:812-827. https://doi.org/10.1016/j.cell.2020.06.043
11. Weissman D, Alameh MG, de Silva T, Collini P, Hornsby H, Brown R, LaBranche CC, Edwards RJ, Sutherland L, Santra S, Mansouri K, Gobeil S, McDanal C, Pardi N, Hengartner N, Lin PJC, Tam Y, Shaw PA, Lewis MG, Boesler C, Şahin U, Acharya P, Haynes BF, Korber B, Montefiori DC (2020) D614G spike mutation increases SARS CoV-2 susceptibility to neutralization. Cell Host Microbe 29:23-31. https://doi.org/10.1016/j.chom.2020. 12.015
12. Mansbach RA, Chakraborty S, Nguyen K, Montefiori DC, Korber B, Gnanakaran S (2021) The SARS-CoV-2 spike cariant D614G favors an open conformational state. Sci Adv 7:eabf3671. https://doi.org/10.1126/sciadv.abf3671
13. Yurkovetskiy L, Wang X, Pascal KE, Tomkins-Tinch C, Nyalile TP, Wang Y, Baum A, Diehl WE, Dauphin A, Carbone C, Veinotte K, Egri SB, Schaffner SF, Lemieux JE, Munro JB, Rafique A, Barve A, Sabeti PC, Kyratsous CA, Dudkina NV, Shen K, Luban J. (2020) Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant. Cell 183(3):739–751. https://doi.org/10.1016/j.cell.2020.09.032
14. NCBI Linrary of Medicine. Quick Access to SARS-CoV-2 Data https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/ [Accessed 5 May 2023]
15. Multiple alignment program for amino acid or nucleotide sequences. https://mafft.cbrc.jp/alignment/server/ [Accessed 5 May 2023]

share
Send us an email: hsijournal@ug.edu.gh
Visit us: https://www.hsijournal.ug.edu.gh