

Article

Single Amino Acid Variant (SAV) percentage and monomer modeling of spike protein of SARS-CoV-2 in Jordan

Walid Al-Zyoud 1*, Hazem Haddad 2 & Ramzi Foudeh 3

¹ Department of Biomedical Engineering, School of Applied Medical Sciences, German Jordanian University, Amman, Jordan; walid.alzyoud@gu.edu.jo

² Princess Haya Biotechnology Centre, Jordan University of Science and Technology, Irbid, Jordan; hazem_haddad1981@just.edu.jo

³ Jordan Society of Genetic Engineers (JSGE), Amman, Jordan; r.foudeh1312@gmail.com

* Correspondence: walid.alzyoud@gu.edu.jo

Abstract:

Spike protein is the surface glycoprotein of the severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) necessary for the entry of the virus via the transmembrane receptors of the human endothelial cells of the respiratory system for the virus to be engulfed causing COVID-19 disease after priming by type II transmembrane protease TMPRSS2 and then binding with the angiotensin-converting enzyme 2 (ACE2). Therefore, mutations and amino acid variants analysis are essential in understanding the mechanism of binding of spike protein with its receptor to have an insights on possibilities to design a peptide or nucleotide-based vaccine for COVID-19. Here, we employed Iterative Threading Assembly Refinement (I-TASSER) and Multiple Alignment using Fast Fourier Transform (MAFFT) to predict the three-dimensional monomer structure of spike protein of SARS-CoV-2 and to analyze the amino acid variants for protein sequences from GISAID database for samples collected from Jordan in a try to find an explanation for the low confirmed number of COVID-19 in Jordan. Our Protein Homology/analogy Recognition Engine V 2.0 (*Phyre2*) findings showed four single amino acid variants (SAV) found in 20 samples of SARS-CoV-2. What is equal to 5% of samples showed tyrosine deletion at Y144 located in the SARS-CoV-like_Spike_S1_NTD (N terminal domain), 62% showed aspartate substitution to glycine at D614G located in the SARS-CoV-2_Spike_S1_RBD (spike recognition binding site), 5% showed aspartate substitution to tyrosine at D1139Y and 5% showed glycine substitution to serine at G1167S both located in the Corona_S2 domain. The findings have shown lower mutational sensitivity in all variants that might not affect the function of spike glycoprotein except for D614G, which has the highest mutational sensitivity score (5 out of 9) indicating a higher likelihood to affect the function of the spike protein. This might suggest, in general, a reduced transmissibility of SARS-CoV-2 in Jordan.

Key words: COVID-19, SARS, spike, variants, & structure.

1. Introduction:

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) caused an outbreak in Wuhan city, China, at the beginning of December 2019 that rapidly spread across the country and to other nations around the world and characterized as a pandemic by the World Health Organization WHO [1]. The first case of SARS-CoV-2 was reported the ministry of health in Jordan on March 2, 2020 for a Jordanian citizen who returned from Italy. To the date of this report, there are 953 confirmed cases, 678 recovered and 9 deaths of COVID-19 in Jordan, according to the official web site launched by the Jordanian Ministry of health as a unified source of information about coronavirus (<https://corona.moh.gov.jo/en>).

SARS-CoV-2 has a positive, single-strand RNA genome that is over 29 kilobases in length, which belongs to one of the four genera of *Orthocoronaviridae*, the beta-coronavirus [2]. Moreover, SARS-CoV-2 encodes four major structural proteins, the envelope (E), membrane (M), nucleocapsid (N), and spike (S) proteins. Spike protein (approx. 180 kDa) is the surface glycoprotein of the severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2)[3]. Spike glycoprotein is necessary for the interaction of the virus with human cell receptors for a sequential combination of the viral encompass with the cell membrane to be engulfed and permit COVID-19 disease by binding with the angiotensin-converting enzyme 2 (ACE2) [4][5] after an evident activation by type II transmembrane protease TMPRSS2 [6].

Here, to understand the early steps of COVID-19 infection, we predicted a three-dimensional structure of the spike glycoprotein of SARS-CoV-2 from positive nasopharyngeal specimens collected in Jordan and sequenced by *Biolab Diagnostic Laboratories* (Jordan) & *Andersen lab at Scripps Research* (USA) who published sequences were retrieved from GISAID, a maintained global database based in Germany. The insight in this work is helpful for scientists to understand different molecular and cytological approaches involved in vaccine development for COVID-19.

2. Materials and Methods:

Genomic sequence retrieval

A total of 20 whole-genome sequences of SARS-CoV-2 collected from Jordan were retrieved from GISAID database and analyzed at the amino sequence level of the spike glycoprotein. The database showed that the nasopharyngeal specimens were collected through March 2020 only with GISAID sequential accession number from EPI_ISL_429992 to EPI_ISL_4300015.

Iterative Threading ASSEmbly Refinement (I-TASSER)

To produce a predicted three-dimensional structure for the S-protein of SARS-CoV-2 collected in Jordan as a PDB file, a hierarchical approach to protein structure and function prediction known as I-TASSER server was used. The I-TASSER pipeline consists of three steps: 1) identification of models, 2) assembly of full-length structures, and 3) annotation of structure-based functions.

Submitting sequence in FASTA format and Multiple Alignment using Fast Fourier Transform

The FASTA formats of the spike gene were aligned (Appendix A), isolated and translated into 1273 amino acids from the whole genome 20 (Jordan) sequences plus 1 reference sequence (accession number YP_009724390.1) of the SARS-CoV-2 by using an open-source functions by the The University of Alcalá, Madrid, Spain at (<http://biomodel.uah.es/en/lab/cybertory/analysis/trans.htm>) and the BLAST function at the NCBI, a web-based service, in addition to Multiple Alignment using Fast Fourier Transform (MAFFT) [7] and viewed by Jalview [8] of Dundee University Scotland. Then the FASTA format of an amino acid sequence of spike protein was submitted to I-TASSER server to get protein structure and function prediction (see appendix for the submitted Sequence in FASTA format).

Single Amino Acid Variant (SAV) Phenotype, protein modeling, and mutation analysis
Four Amino Acid Variant (SAV) found from 20 spike glycoprotein submitted and retrieved by *Phyre2* server to predict mutational sensitivity [9–11].

Nomenclature sequence Amino Acid Variant (SAV) and annotation used the accession number
Surface glycoprotein [Severe acute respiratory syndrome coronavirus 2] with accession number YP_009724390.1 was used as a reference sequence to compare with, and it was downloaded from https://www.ncbi.nlm.nih.gov/protein/YP_009724390.1?report=fasta.

3. Results:

Predicted Secondary Structure

Initially, the I-TASSER was utilized to recognize the basic templates from the PDB by multiple threading approach LOMETS, with full-length atomic models produced by iterative fragment assembly simulations based on templates. Function insights of the targeted molecule are then obtained by rethreading the three-dimensional models via the BioLiP database of protein functions. Figure 1 shows the predicted secondary structure of the SARS-CoV-2 spike glycoprotein tested in Jordan.

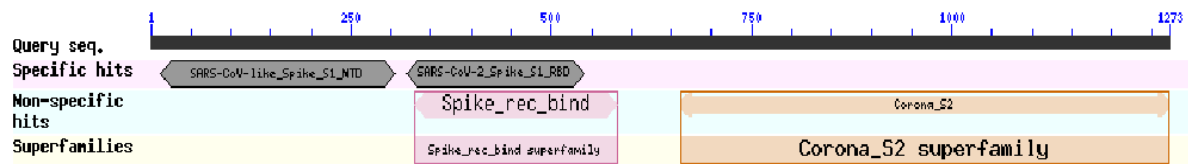


Figure 1: Spike glycoprotein predicted secondary structure included SARS-CoV-like_Spike_S1_NTD (N terminal domain), SARS-CoV-2_Spike_S1_RBD (spike recognition binding site) and Corona_S2 doamin.

The final predicted model of the monomer of spike protein by I-TASSER

For each target, I-TASSER simulations called decoys generate an extensive collection of structural conformations. I-TASSER uses the SPICKER to cluster all architectural structures based on the pair-sided similarity and records up to 5 models corresponding to the five largest structural clusters. The reliability of each model is evaluated quantitatively by a C-score based on the value of threaded prototype alignments and the parameters of convergence of structural mounting simulations. C-score is usually [-5, 2], where a higher-value C-score means a more positive and vice versa scale. Following the association observed between these attributes, the TM-score and RMSD are calculated using the C and the protein frequency. Since the group size classes the top 5 models, in some situations, a higher C-score is possible for the lower-ranking models. While the first model is better in most cases, as seen in our research (Fig. 4), lower-level models can also be better than higher-level models. If the I-TASSER simulations converge, less than 5 clusters can have been generated; it usually shows that because of the converged simulations, the models have good quality.

Very recently, the experimental cryo-EM structures of the glycoprotein of SARS-CoV-2 have been reported by two groups; Wrapp et al. and Walls et al. and the structures are available in the Protein Data Bank, e.g., 6vsb, 6vxx [12] & [3]. While the deposited structures are not full-length and not solved by crystallography but cryo-EM, the resolution of those structures is close to the crystal structures. Zhang’s group (zhanglab.ccmb.med.umich.edu/COVID-19/) has modeled other structural models of SARS-CoV-2, including the surface glycoprotein, with I-TASSAR. We found no difference between the recently reported experimental structures and the *in silico* models we generated, even with the new single amino variants.



Figure 4 : (a) The final I-TASSER predicted spike glycoprotein spike protein monomer model, the PDB file is provided in the supplementary data, and (b) the trimer spike glycoprotein as elucidated by Cryo-EM; PDB:6vxx [12].

The top two proteins structurally close to the spike glycoprotein in the Protein Data Bank (as identified by TM-align) are listed in Table 1. In Table 2 the top five hits of closest Enzyme Commission (EC) numbers and active sites are listed.

Table 1: Proteins structurally close to the spike glycoprotein in the Protein Data Bank (as identified by TM-align).

Rank	PDB Hit	TM-score	RMSD ^a	IDEN ^a	Cov
1	5x58A	0.827	0.41	0.751	0.828
2	6nzkA	0.741	4.62	0.276	0.840

Protein rankings are based on the structural alignment TM score in the PDB library between the query template and known structures. **RMSD^a** the RMSD among structurally aligned residues of TM-align; **IDEN^a** is the structurally related region's percentage sequence identity; **Cov** reflects the alignment range of the TM-alignment and is proportional to the sum by the length of query protein of structurally aligned residues. **5x58A**: Prefusion structure of SARS-CoV spike glycoprotein, conformation 1 (viral protein); **6nzkA**: Structural basis for human coronavirus attachment to sialic acid receptors (viral protein).

Table 2: Enzyme Commission (EC) numbers and active sites

Rank	Cscore ^{EC}	PDB Hit	TM-score	RMSD ^a	IDEN ^a	Cov	EC Number	Active Site Residues
1	0.082	1ileA	0.205	8.54	0.041	0.296	6.1.1.5	NA
2	0.082	1k32A	0.198	9.14	0.046	0.295	3.4.21.-	NA
3	0.082	3eqIM	0.201	8.65	0.046	0.288	2.7.7.6	NA
4	0.081	2pdaA	0.184	10.13	0.027	0.295	1.2.7.1	NA
5	0.081	1ej6A	0.190	10.21	0.016	0.309	2.7.7.50	NA

1ileA: Isoleucyl-tRNA synthetase (aminoacyl-tRNA synthetase); **1k32A**: Crystal structure of the tricorn protease (hydrolase); **3eqIM**: Crystal structure of the *T. Thermophilus* RNA polymerase holoenzyme in complex with antibiotic myxopyronin (transferase); **2pdaA**: Crystal structure of the complex between pyruvate-ferredoxin oxidoreductase from *Desulfovibrio africanus* and pyruvate (oxidoreductase); **1ej6A**: Reovirus core (virus).

One powerful way of multiple sequence alignment is the Multiple Alignment using Fast Fourier Transform (MAFFT) as shown in figure 5 below [8]. Figure 6 (a) shows the location of single amino acid variants del Y144 & D614G on spike protein as they have the highest mutation sensitivity scores and (b) shows the mutation sensitivity histograms for all four single amino acid variants on spike protein; Y 144 D 614, D 1139, and G 1167 calculated according to [10].

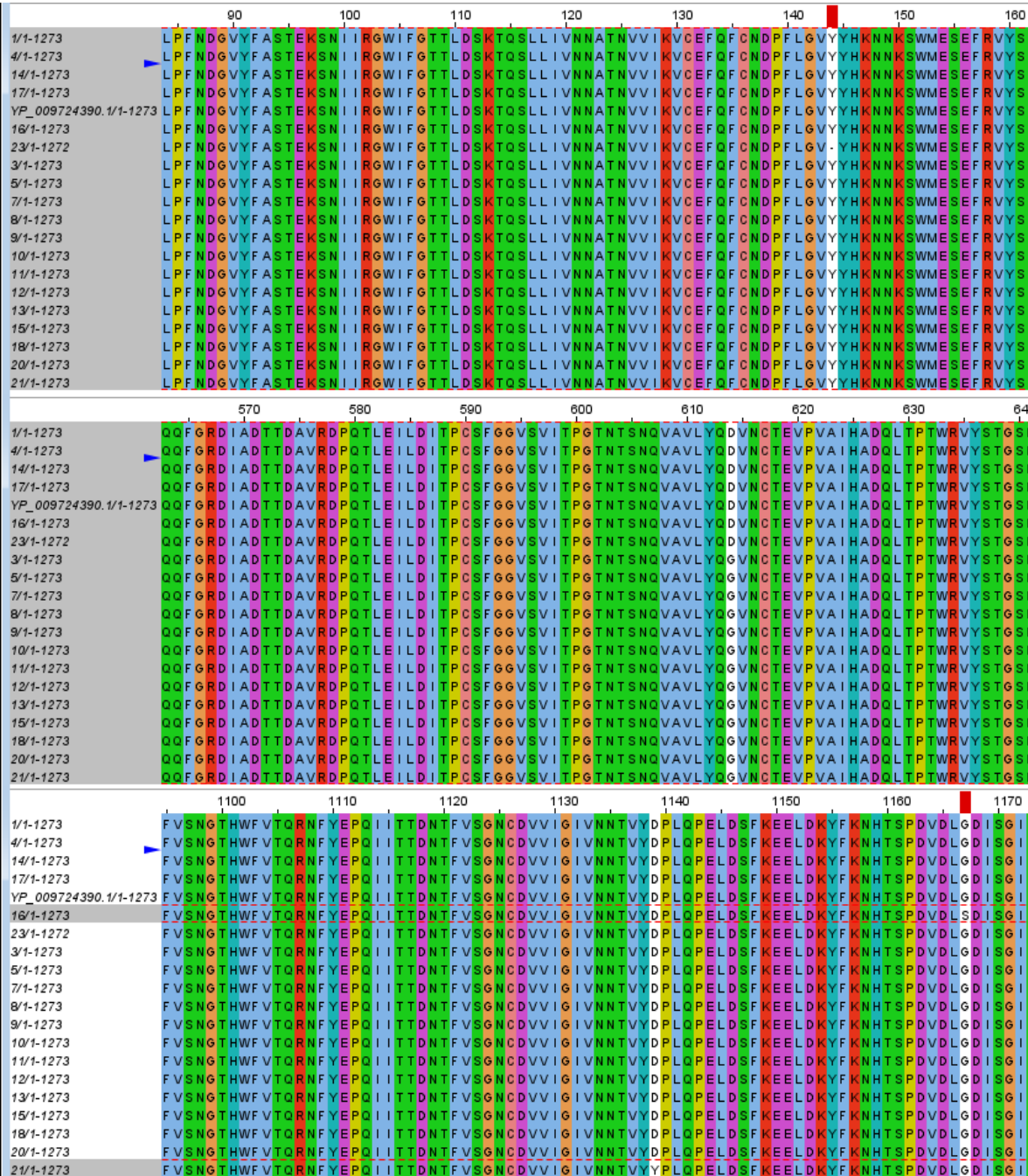


Figure 5 : Multiple Sequence Alignment showing Amino Acid Variant (SAV) viewed by *Jalview*. The blue arrow and the red box are just cursers; the white columns are the location of the SAV along the full length of amino acid sequence of spike protein for 20 SARS-CoV-2 samples from Jordan. YP-009724390 is the reference sequence of spike protein of from Wuhan. The number 1273 is the total number of amino acid in the sequence.

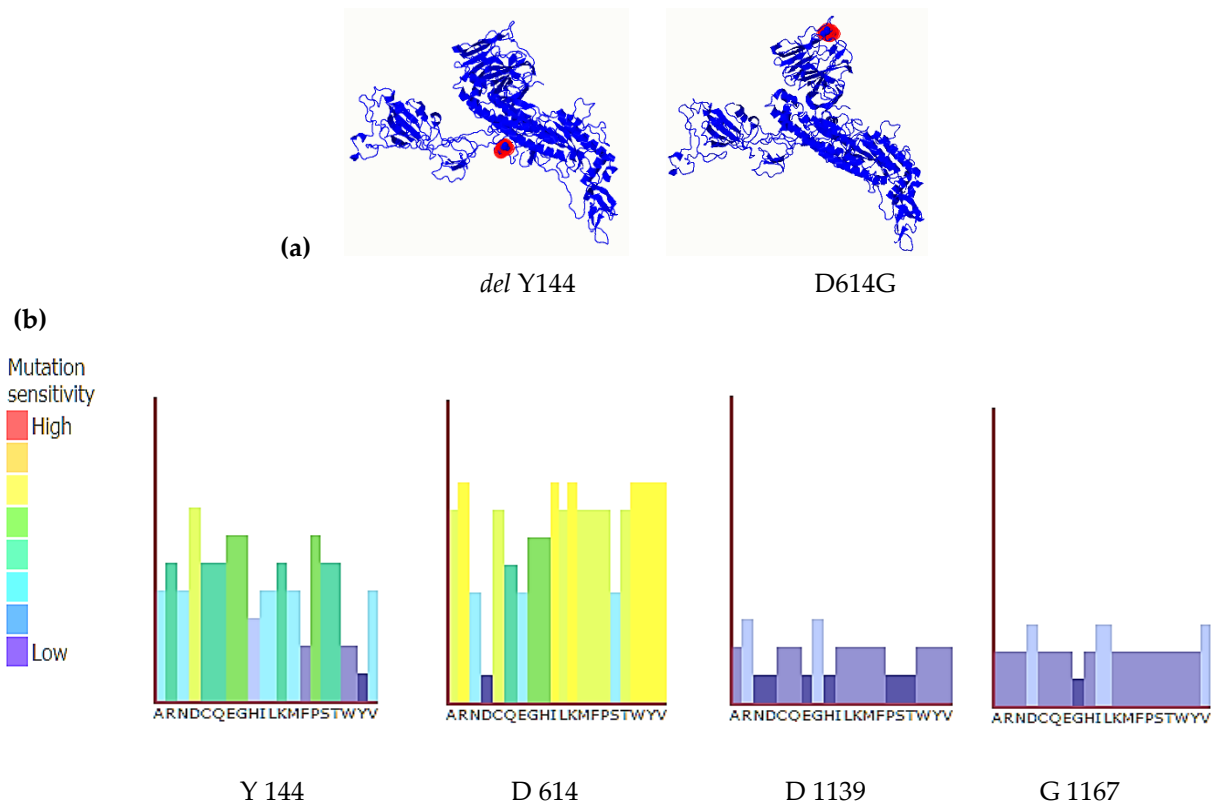


Figure 6: (Phyre2) Mutational sensitivity (a) the location of single amino acid variants del Y144 & D614G on spike protein as they have the highest mutation sensitivity scores. (b) The mutation sensitivity histograms for all four single amino acid variants on spike protein.

Table 3 shows that the D614G has the highest percentages and mutation sensitivity in the spike protein in the Jordanian population. This mutation showed the substitution of aspartate; a bulky amino acid to glycine; the simplest amino acid ever.

Table 3: SAV in spike-protein of SRAS-CoV-2; percentage and mutational sensitivity scores in Jordanian population

SAV in spike-protein of SRAS-CoV-2	SAV Percentage in Jordanian population	(Phyre2) Mutational sensitivity Score out of (9=high mutation sensitivity, 0=low)
Deletion at Y144	5%	3
D614G	62%	5
D1139Y	5%	1
G1167S	5%	1

4. Discussion:

In this study, we used the spike gene sequences from 20 whole-genome sequences of SARS-CoV-2 collected from Jordan retrieved from GISAID database and analyzed at the amino sequence level of the spike glycoprotein versus a reference sequence of the surface glycoprotein [Severe acute respiratory syndrome coronavirus 2; (SARS-CoV-2)] own the accession number YP_009724390.1 of Wuhan. Our findings showed that, the molecules which were structurally close to the spike glycoprotein from the Enzyme Commission (EC) numbers and active sites included Isoleucyl-tRNA synthetase which provide the protein the ability to synthesize tRNA, Crystal structure of the tricorn protease (hydrolase) which provide the protein the ability to hydrolyse the host proteins need for viral entry; Crystal structure of the *T. Thermophilus* RNA polymerase holoenzyme (transferase) which provide the protein the ability to synthesize the viral RNA; Crystal structure of the complex between pyruvate-ferredoxin oxidoreductase

from *Desulfovibrio africanus* and pyruvate (oxidoreductase); and Reovirus core (virus) all might explain the ability of SARS-CoV-2 in getting inside the human host cells.

Four Amino Acid Variants (SAV) founded in 20 samples of SARS-CoV-2 were discovered in this study with 5% of samples showed tyrosine deletion at Y144 located in the SARS-CoV-like_Spike_S1_NTD (N terminal domain), 62% showed aspartate substitution to glycine at D614G located in the SARS-CoV-2_Spike_S1_RBD (spike recognition binding site), 5% showed aspartate substitution to tyrosine at D1139Y and 5% showed glycine substitution to serine at G1167S both located in the Corona_S2 domain. From a biochemistry point of view, the D614G mutation showed the substitution of aspartate; a bulky amino acid to glycine; the simplest amino acid ever, which might suggest a reduced new affinity between spike protein and its receptors. The findings have shown lower mutational sensitivity in all variants that might not affect the function of spike glycoprotein except for D614G, which has the highest mutational sensitivity score (5 out of 9) indicating a higher likelihood to affect the function of the spike protein. This might suggest, in general, a reduced transmissibility of SARS-CoV-2 in Jordan. The D614G substitution was previously reported as a dominant mutation in Europe [9]. The I-TASSER predicted monomer three-dimensional structures of spike protein of SARS-CoV-2 had similar stability structures for all of the four Amino Acid Variant (SAV) when we compared the reference sequence of the spike glycoprotein YP_009724390.1 (SARS-CoV-2) with FASTA sequences of spike glycoproteins from Jordanian population; no change on the three-dimensional structure was noticed.

The generated three-dimensional monomer structure of the spike protein of SARS-CoV-2 is consistent with a perfusion conformation structure reported in the literature [3]. Like any *in silico* study, this study is limited with the capabilities of utilized servers and algorithms to it is highly dependent on the initial templates used for calculations, so if the initial template scoring is not good enough then this might affect the final output files.

5. Conclusion:

This is the first study of its kind in the Middle East to predict the monomer three-dimensional structure of the spike glycoprotein from SARS-CoV-2 of Jordanian specimens. In addition, we reported four amino acid variants, which might explain the current low number of COVID-19 cases, 953 confirmed cases, and 678 recovered and 9 deaths. However, the highest frequency mutation in our study, with 62% of samples showed aspartate substitution to glycine at D614G is consistent with other reports for samples collected in Europe at the same time of our samples collection, March 2020. In this study, we consider the mutation D614G as the dominant local mutation in Jordan. We believe that the reported four amino acid variants especially the tyrosine deletion at Y144 located in the SARS-CoV-like_Spike_S1_NTD (N terminal domain) and the aspartate substitution to glycine at D614G located in the SARS-CoV-2_Spike_S1_RBD (spike recognition binding site) have collectively reduced the spike protein affinity of SARS-CoV-2 with ACE2 receptors in the Jordanian population. It is highly recommended to keep monitoring the mutation rate of SARS-CoV-2 in Jordan in monthly bases with higher number of samples to fulfil a statistical power. Some of the low percentage appeared mutations e.g. 5% might be increased if the population size is higher.

Author Contributions: “Conceptualization, W.A-Z., R.F & H.H.; methodology, W.A-Z & H.H.; software, W.A-Z & H.H.; validation, W.A-Z & H.H. and R.F.; formal analysis, W.A-Z & H.H.; investigation, W.A-Z & H.H.; resources, R.F.; data curation, W.A-Z & H.H.; writing—original draft preparation, W.A-Z.; writing—review and editing, W.A-Z & H.H.; visualization, W.A-Z & H.H.; supervision, R.F.; project administration, R.F.; funding acquisition, W.A-Z & H.H. All authors have read and agreed to the published version of the manuscript.”

Funding: “This research received no external funding”

Acknowledgments: The authors acknowledge Biolab Diagnostic Laboratories (Jordan) & Andersen lab at Scripps Research (USA) who published sequences were retrieved from GISAID, a maintained global database based in Germany.

Conflicts of Interest: “The authors declare no conflict of interest.”

Appendix A:

Submitted Sequence in FASTA format

del. Y144 Sample 23

>hCoV-19/Jordan/SR-042/2020|EPI_ISL_430000|2020-03-30

MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFWHAIHVS GTNGTKRFDNPVLPF
NDGVYFASTEKSNIRGWIFGTTLD SKTQSL LIVN NATNVVIK VCEFCNDPFLGVYHKNNKSWMESEFRVYSSANNCTFEYVS
QPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHRSYLT PGDSSSGWT
AGAAAYYVGYLQPRTFLLKYNENGTTITDAVDCALDPLSETKCTLSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVF NAT
RFASVYAWNRRKRISNCVADYSVL YNSASFSTFKCYGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNYKLPDDFT
GCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFPQTNGVGYQPYRVVLSF
ELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP
GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNV FQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRA
RSVASQSIIAYTMSLGAENSVAYSNN SIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTEC SNLLQYGSFCTQLNRALTGIAV
EQDKNTQEVF A QVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLIC AQKFNG
TVLPPLLTDEMI AQYTSALLAGTITSGWTFGAGAA LQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFN SAIGKIQDSLSTAS
ALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSR LDKVEAEVQIDRLITGRLQSLQTYVTQQ LIRAAEIRASANLAATK
MSECVLGQSKRVDFCGKGYHLM SFPQSAPHGVVFLHV TYVPAQEKNF TTA PAICH D GKAHFPREGVFVSN GTHW FVTQRNFY
EPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPD VDLGDISGINASVVNIQKEIDRLNEVAKNLNES
LIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCC SCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT

D1139Y Sample 21

>hCoV-19/Jordan/SR-033/2020|EPI_ISL_429993|2020-03-16

MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFWHAIHVS GTNGTKRFDNPVLPF
NDGVYFASTEKSNIRGWIFGTTLD SKTQSL LIVN NATNVVIK VCEFCNDPFLGVYHKNNKSWMESEFRVYSSANNCTFEYV
SQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHRSYLT PGDSSSGW
TAGAAAYYVGYLQPRTFLLKYNENGTTITDAVDCALDPLSETKCTLSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNA
TRFASVYAWNRRKRISNCVADYSVL YNSASFSTFKCYGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNYKLPDDF
TGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFPQTNGVGYQPYRVVVL
SFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVIT
PGTNTSNQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTGSNV FQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRR
ARSVASQSIIAYTMSLGAENSVAYSNN SIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTEC SNLLQYGSFCTQLNRALTGIA
VEQDKNTQEVF A QVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLIC AQKFNG
LTVLPPLLTDEMI AQYTSALLAGTITSGWTFGAGAA LQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFN SAIGKIQDSLSTAS
SALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSR LDKVEAEVQIDRLITGRLQSLQTYVTQQ LIRAAEIRASANLAAT
KMSECVLGQSKRVDFCGKGYHLM SFPQSAPHGVVFLHV TYVPAQEKNF TTA PAICH D GKAHFPREGVFVSN GTHW FVTQRNF
YEPQIITTDNTFVSGNCDVVIGIVNNTVYYP LQPELDSFKEELDKYFKNHTSPD VDLGDISGINASVVNIQKEIDRLNEVAKNLNE
SLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCC SCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT

G1167S Sample 16

>hCoV-19/Jordan/SR-039/2020|EPI_ISL_429998|2020-03-28

MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFWHAIHVS GTNGTKRFDNPVLPF
NDGVYFASTEKSNIRGWIFGTTLD SKTQSL LIVN NATNVVIK VCEFCNDPFLGVYHKNNKSWMESEFRVYSSANNCTFEYV
SQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHRSYLT PGDSSSGW
TAGAAAYYVGYLQPRTFLLKYNENGTTITDAVDCALDPLSETKCTLSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNA
TRFASVYAWNRRKRISNCVADYSVL YNSASFSTFKCYGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNYKLPDDF
TGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFPQTNGVGYQPYRVVVL
SFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVIT
PGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNV FQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRR
ARSVASQSIIAYTMSLGAENSVAYSNN SIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTEC SNLLQYGSFCTQLNRALTGIA
VEQDKNTQEVF A QVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLIC AQKFNG
LTVLPPLLTDEMI AQYTSALLAGTITSGWTFGAGAA LQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFN SAIGKIQDSLSTAS
SALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSR LDKVEAEVQIDRLITGRLQSLQTYVTQQ LIRAAEIRASANLAAT
KMSECVLGQSKRVDFCGKGYHLM SFPQSAPHGVVFLHV TYVPAQEKNF TTA PAICH D GKAHFPREGVFVSN GTHW FVTQRNF
YEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPD VDLSDISGINASVVNIQKEIDRLNEVAKNLNES
LIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCC SCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT

D614G Sample 3,5,7,8,9,10,11,12,13,15,18,20 and 21

Example:

>hCoV-19/Jordan/SR-036/2020|EPI_ISL_429996|2020-03-23

MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPF
NDGVYFASTEKSNIIIRGWIFGTTLDSTQSLIVNNATNVVIKVFCEFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYV
SQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHRSYLT PGDSSSGW
TAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNA
TRFASVYAWNKRKISNCVADYSVLVNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPDDF
TGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVL
SFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVIT
PGTNTSNQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRR
ARSVASQSIIAYTMSLGAENSVAYSNNISAIPTNFTISVTTEILPVSMKTSTVDCTMYICGDSTECNLLLQYGSFCTQLNRALTGIA
VEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSPKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNG
LTVLPPLLTDemiaQYTSALLAGTITSGWTFGAGAAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTA
SALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLRAAEIRASANLAAT
KMSECVLGQSKRVDFCGKGYHLMSPQSAFHGVVFLHVTYVPAQEKNFTTAPAICHGDKAHFPREGVVFVSNNGTHWFTQRFN
YEPQIIITDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVLGDISGINASVVNIQKEIDRLNEVAKNLNE
SLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT

References:

1. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020 Mar 12;579(7798):265–9.
2. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*. 2020 Feb 22;395(10224):565–74.
3. Daniel Wrapp, Nianshuang Wang, Kizzmekia S. Corbett, Jory A. Goldsmith C-LH, Olubukola Abiona, Barney S. Graham JSM. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. - PubMed - NCBI. *Science*. 2020.
4. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*. 2020 Mar 30;1–6.
5. Zhang H, Penninger JM, Li Y, Zhong N, Slutsky AS. Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. *Intensive Care Med*. 2020 Apr 1;46(4):586–90.
6. Glowacka I, Bertram S, Muller MA, Allen P, Soilleux E, Pfefferle S, et al. Evidence that TMPRSS2 Activates the Severe Acute Respiratory Syndrome Coronavirus Spike Protein for Membrane Fusion and Reduces Viral Control by the Humoral Immune Response. *J Virol*. 2011 May 1;85(9):4122–34.
7. Abio Madeira F', Mi Park Y, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Web Serv issue Publ online*. 2019;47.
8. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;25(9):1189–91.
9. Angyal A, Brown RL, Carrilero L, Green LR, Groves DC, Johnson KJ, et al. Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2 on behalf of the Sheffield COVID-19 Genomics Group#, LaBranche CC2, and Montefiori DC2.
10. Yates CM, Filippis I, Kelley LA, Sternberg MJE. SuSPect: Enhanced prediction of single amino acid variant (SAV) phenotype using network features. *J Mol Biol*. 2014 Jul 15;426(14):2692–701.
11. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015 Jun 30;10(6):845–58.
12. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veasler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*. 2020 Apr 16;181(2):281-292.e6.