USING BIOINFORMATICS TO INVESTIGATE THE NOVEL SARS-COV-2 VARIANTS AND THEIR IMPACTS ON INFECTIVITY

SHEREEN F. ELKHOLY[1,2,*]

1. Biology Department, College of Science and Humanities, Prince Sattam bin Abdulaziz University, 11942 Alkharj, KSA

2. Department of Plant Molecular Biology, Plant Transformation and Biopharmaceuticals Lab, Agricultural Genetic Engineering Research Institute (AGERI), Agricultural Research Centre (ARC), Giza, Egypt

** Corresponding author: s.elkholy@psau.edu.sa - and - shereen.elkholy@yahoo.com*

Keywords: *Computational Biology, COVID -19, SARS-COV-2, SARS-CoV-2 variant*

A large family of viruses "Coronaviruses" cause illness like common cold and other severe diseases as: Severe Acute Respiratory Syndrome-SARS and Middle East Respiratory Syndrome-MERS. (World Health Organization, 2020). This virus was first recognized in the respiratory system of patients with pneumonia in Wuhan, Hubei Province, China, in December 2019, which was then referred to SARS-CoV-2. It is a β coronavirus or an enveloped, single-stranded, positive-sense RNA virus, that is widespread in humans and other mammals (Astuti, 2020). The virus causes the coronavirus disease 2019 (COVID-19), with common symptoms such as fever, cough, shortness of breath and fatigue (World Health Organization, 2020). Virus transmission within the community and antiviral treatments, may engender novel mutations, resulting virulent strains resistant to treatment and/or with higher rates of mortality (World Health Organization, 2020).

SARS-COV-2

A family of enveloped RNA viruses "Coronaviruses" broadly distributed in animals and humans cause chronic and acute diseases (Tanner *et al*., 2003). There are six coronavirus species cause human diseases, four cause common cold symptoms, SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV) cause serious respiratory disease (Tanner *et al*., 2003). SARS-CoV-2 shares many similarities with SARS-CoV, SARS-CoV-2 is closely related to two bat-derived SARS-like coronaviruses and SARS-CoV in the phylogenetic distance (Lu *et al*., 2020).

Both SARS-CoV-2 and SARS-CoV uses the same cellular receptor. Both SARS- CoV and MERS-CoV cause fatal

pneumonia and rapidly replicated, along with an elevation of pro inflammatory cytokines and inflammatory cell infiltration (Channappanavar *et al*., 2017). SARS-CoV infection is immune dysregulation, additionally the level of peak viremia (Channappanavar *et al*., 2016).

Covid-19

Covid-19, the disease caused by the emerging coronavirus, SARS-Cov-2. The new technologies are the way to go for the medical industry to control and monitor the spread of the COVID-19 pandemic (Vaishya *et al*., 2020).

The artificial intelligence (AI) helps fighting the virus *via* medical help, population screening, notification, and suggestions to control the infection. AI has the potential to improve the treatment, planning, being an evidence-based medical tool (Bai *et al*., 2020; Hu *et al*., 2020).

SARS-COV-2 variant

A sample isolation from pneumonia patients who were some of the workers in the Wuhan seafood market found that strains of SARS-CoV-2 had a length of 29.9 kb (Wu, 2020). The SARS-CoV-2 mainly structured 1; spike (S) glycoprotein, 2; small envelope (E) glycoprotein, 3; membrane (M) glycoprotein, 4; nucleocapsid (N) protein and several accessory proteins (Fig. 1) (Jiang *et al*., 2020).

In the outer portion, the S glycoprotein "trans membrane protein" is found with a molecular weight of about 150 kDa. The S protein structure homotrimers protruding in the viral surface and assists binding of envelope viruses in order to host cells via attraction with angiotensin-converting enzyme 2 (ACE2) articulated in lower respiratory tract cells. That glycoprotein mainly cleaved via the host cell furin-like protease into 2-sub units i.e., S1 and S2 (Fehr and Perlman, 2015).

Subunit S1 is responsible for the determination of the host virus range and cellular tropism with the receptor binding domain make-up while S2 functions to mediate virus fusion in transmitting host cells (Astuti, 2020).

Nucleocapsid "N protein" is the structural component of CoV found in the endoplasmic reticulum at Golgi region and structurally bound to the virus' nucleic acid. Protein (bound to RNA) is involved in the processes of viral replication cycle, viral genome, and the cellular response of host cells to viral infections (Tai *et al*., 2020). The N protein is heavily phosphorylated and cause structural changes improving the affinity for viral RNA. The membrane or M protein "structured protein" plays a role in determining the shape of the virus envelope. The envelope or E protein "smallest protein" in the SARS-CoV structure plays a role in the maturation and production of this virus (Astuti, 2020).

Nidovirus family includes SARS-CoV-2, contracted from bats and fellow humans. The ACE2 receptors facilitate viral entry into target cells like heart, lungs, kidneys, and gastrointestinal tract

(Rabi *et al.*, 2020). The CoV enter the host cell by the attachment of S glycoprotein to the ACE2 receptor (Tai, 2020) in the host cells (like type II pneumocytes of lungs). The S protein of SARS-CoV-2 receptors at 331 to 524 residues binds strongly to human ACE2 and ACE2 of bat, followed by fusion of the host cell and viral membrane. The (TMPRSS2) serine protease of the host cell (Walls *et al*., 2020) clear ACE2 and activate the receptor-attached spike-like, S proteins (Rabei and Alzoubi, 2020). Activation of the S proteins leads to conformational changes and allows the virus to enter the cells (Simmons, 2013). These proteins TMPRSS2 and ACE2 are prove of the entry of the virus (Astuti, 2020).

The emergence of SARS-COV-2 variants

SARS-CoV-2 virus causes COVID-19 is stable over last fall. Currently, many variants are identified that indicate to change how the virus behaves. Viruses are designed to reproduce rapidly, but without 100% accuracy. The RNA viruses, "coronaviruses" are highly mutating as the proteins that copy RNA are error prone (Hou *et al*., 2020). The mutation (or mutations) in the “spike” protein of SARS-CoV-2 (of late 2019) allowed it to emerge from its animal host(s) and infect humans. The SARS-CoV-2 doesn’t mutate rapidly as others coronaviruses. During 1st wave of infections last spring, many variants were tracked but nothing of particular concern emerged. There is a high surge in infections worldwide since October, however, this virus has had millions of opportunities to mutate and evolve. Most of the changes (variants) have little or no effect; they can even hamper transmissibility, and quickly disappear. Few of these variants will actually enhance virus spread (e.g., in B.1.1.7, identified in England). If the variant is significantly more infectious, it can out-compete the other ones and become the most common found in humans over time.

In China variant of SARS-CoV-2 originated and spread around the world is called D614G (Hou *et al*., 2020; Zhou *et al*., 2021). This variant moved during both spring and summer, and minor changes enabled the tracking of different transmission routes, which functioned in the same ways. Many pharmaceutical companies used D614G’s spike protein sequence and developed the vaccines.

A variant found in the U.K., B.1.1.7 is more efficiently transmitted. This variant carries 18 lineage-defining mutations, one of them in the spike (S) gene known as N501Y affects binding to the human ACE2 receptor. That variant is more virulent causing severe disease with higher percentage of patients. The current vaccines protect against B.1.1.7, and increasing the pressure to vaccinate many people. The B.1.351 variant emerged in South Africa, and P1, arose in Brazil may be the cause of a highly concerning 2nd wave of infection in the city of Manaus. B.1.351" has the same N501Y S gene mutation as the B.1.1.7 variant, and two more in key sites of receptor binding domain of spike protein, and not recognized by the

immune systems of people previously infected with D614G. Vaccine reactivity is weak compare to B.1.351, but still effective. Studies are ongoing on vaccine and immune protection against the variant.

A long with many other lineage defining mutations, the P.1 has the N501Y mutation. Indicating that it spontaneously arose in three variants investigated for ease transmissibility, the N501Y provides advantage for the variants in which it's existing. The P.1 is emerged in Manaus, underwent a massive surge of COVID-19, with infection rates of up to 75%. P.1 was not identified prior to December 2020, however, so its properties and potential for immune evasion are just now being studied.

Bioinformatics

Bioinformatics can appear to be a challenging field as it's a combination of the complex science of Biology with the complex theory of computer (Holloway *et al*, 2020). Bioinformatics's tool helps to design, compare, and forecast the structural and functional characteristics of proteins and genes. Also, biology, molecular biology, in particular, is undergoing transformations with the growing awareness of the computational nature of many biological processes and that computational and statistical model can be used. Moreover, high-throughput data acquisition requires computational and statistical analysis at each stage (Gentlman *et al*, 2004). Thus, bioinformatics offer the potential of speeding up drug discovery with cost reduction (Alberto *et al*, 2006).

This study used computational analysis to compare the wild type of SARS-Cov-2 with new variants of SARS-Cov-2 all over the world and find out the difference between them in terms of mutations, changes in amino acids sequences, impact on viral load, infectivity and resistance by manipulation from the immunogenic responses.

Objectives

1- Comparison between the wild type of SARS-COV-2 and new variants of SARS-COV-2 from three different countries Pakistan, Brazil and Egypt.

2- Alignment between the old and the new viral strains and the percentage of identity and determine the location of the mutation and identify the type of mutations.

3- Identify the mutations of the surface glycoprotein of SARS- COV-2 and the 3D configurations of them and determining the sites of mutations.

4- Identify the difference between the D614G variant of the surface glycoprotein mutant and the wild-type.

5- The interaction between the wild type and the mutant (D614G) of S glycoprotein with the Fab fragment of the neutralizing antibody.

6- Building phylogenetic tree of 30 variants of SARS-COV-2 isolated from different countries all over the world.

MATERIALS AND METHODS

Materials

- ***NCBI***

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information. The NCBI houses a series of databases relevant to biotechnology and biomedicine and is an important resource for bioinformatics tools and services. Major databases include GenBank for DNA sequences and PubMed, a bibliographic database for biomedical literature (The New York Times 1997).

- *BLAST*

An algorithm used for calculating sequence similarity between biological sequences such as nucleotide sequences of DNA and amino acid sequences of proteins (Altschul *et al*., 1990). BLAST is a powerful tool for finding sequences similar to the query sequence within the same organism or in different organisms. It searches the query sequence on databases and servers and posts the results back to the person's browser in the chosen format. HTML is the default output format for NCBI's web-page. Results for NCBI-BLAST are presented in graphical format with all the hits found, a table with sequence identifiers for the hits having scoring related data, along with the alignments for the sequence of interest and the hits received with analogous BLAST scores (Madden, 2002).

- ***Protein Data Bank (PDB)***

The Protein Data Bank (PDB) is a database for the three-dimensional structural data of large biological molecules, such as proteins and nucleic acids (Consotium, 2019). The PDB is a key in areas of structural biology, such as structural genomics. Most major scientific journals and some funding agencies now require scientists to submit their structure data to the PDB. Many other databases use protein structures deposited in the PDB (Berman, 2008). The Protein Data Bank was announced in October 1971 in Nature New Biology as a joint venture between Cambridge Crystallographic Data Centre, UK and Brookhaven National Laboratory, US.

Method

- ***Computational analysis***

First, the study used the NCBI website to identify the wild type of SARS-COV-2 and three new variants of SARS-COV-2 in three different countries all over the world (Pakistan, Brazil and Egypt) Table (1).

Second, the study compared the wild type of SARS-COV-2 with the new variants of SARS-COV-2 for each country, the study used Blast website to make alignment of old and

new viral strains, identity ratio, mutation localization, and determination of the kind of mutation.

Third, the study used PDB to identify the 3D configurations of the S glycoprotein of SARS-COV-2 and its variants and determining the localization of mutations of the new variants of S glycoprotein.

Fourth, the study used the PDB site to find out the difference in the 3D configuration structures between the D614G variant of the surface glycoprotein and the wild type that affect its receptor binding domain and the extent of the interaction between the wild type and the mutant (D614G) of the glycoprotein S with the Fab fragment of the neutralizing antibody.

Finally, the study used the NCBI website to build a phylogenetic tree of 30 SARS-COV-2 variants isolated from different countries all over the world. A phylogenetic tree (also phylogeny or evolutionary-tree) is a branching diagram or a tree showing the evolutionary relationships among various biological species or other entities based upon similarities and differences in their physical or genetic characteristics. In a rooted phylogenetic tree, each node with descendants represents the inferred most recent common ancestor of those descendants and the edge lengths in some trees may be interpreted as time estimates.

RESULTS AND DISCUSSION

✓ RESULTS

Alignment between the wild type SARS-COV-2 and the new variants

The study used BLAST site to compare between the genomic sequences of the wild type and new variants that emerged in three different countries all over the world (Brazil, Pakistan, and Egypt) in order to identify the type of point mutations that was discovered and its location in, Brazil (Fig.2 and Table 2), Pakistan (Fig. 3 and Table 3) and Egypt (Fig. 4 and Table 4).

Identification of the mutations in the Surface glycoprotein of SARS- COV-2 and the 3D configurations of them

The study focused on the surface glycoprotein of SARS-COV-2 since it is the immunogenic part of the virus that binds to the host cell receptor (ACE2), mediating viral cell entry. The study used SARS-COV-2 gene data bank https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/scov2_snp to identify the 3D configurations of the S glycoprotein of SARS-COV-2 and its variants that detected all over the world and determine the genomic location of mutations of the new variants of S glycoprotein in addition to the codon change, protein change type and collection location (Table 5). There are differences

were found in 3D configurations between the wild type of S glycoprotein with PDB id: 6VXX and the three mutant variants that detected in different countries all over the world (Table 5).

The difference between the D614G variant of the surface glycoprotein mutant and the wild-type

RCSB Protein data bank I determined the main differences in structure and immunogenic features between the wild type surface glycoprotein with PDB id: 6VXX and the mutant variant D614G with PDB id: 6XS6 in addition to the interaction between the wild type and the mutant (D614G) of S glycoprotein with the Fab fragment of the neutralizing antibody (Table 5).

In the new variant of the S glycoprotein (D614G), Aspartic acid in 614 is substituted by Glycine. This substitution in the spike glycoprotein of SARS-CoV-2 strain is now the most prevalent form globally. The study found out the difference between the wild type and the mutant in terms of mutations and immunological features and their receptor binding domains which is a key part of the virus that is located on its' spike' protein that facilitates its entrance into host cells and lead to infection. The study found that the D614G variant exhibit alteration in the RBD that affects its pathogenicity, infectivity, reduced antibody binding and immune protection (Fig. 5).

Building phylogenetic tree of 30 variants of SARS-COV-2 isolated from different countries all over the world

The NCBI website was used to build a phylogenetic tree of 30 SARS-COV-2 variants isolated from different countries all over the world and their relationship with the wild type with accession number NC_045512.2 that was isolated from China (Fig. 6).

The rooted phylogenetic tree, every node with descendants condtitutes the inferred most recent common ancestor of those descendants, edge lengths may be interpreted as time estimates. The smaller the distances it's more related. Fortunately, I didn't detect any new variants in Saudi Arabia in 2021.

✓ DISCUSSION

In Wuhan (Dec. 2019) Chinese government reported a novel pneumonia-causing disease, and named severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) by the International Committee on Taxonomy of Viruses (ICTV), and spread across China and to many countries. The World Health Organization has named the illness caused by SARS-CoV-2 as coronavirus disease 2019 (COVID-19). Currently, three coronaviruses cause illness in humans: SARS-CoV-2, severe acute respiratory syndrome coronavirus (SARS-CoV), and Middle East respiratory syndrome coronavirus (MERS-CoV).

The study verified the genomic sequences of the selected viruses, and I also learnt about virus mutations and the extent of their effect. Despite the multiplicity of mutations and their abundance, it is single point mutations i.e base pair substitution and therefore did not significantly affect morphology of the virus radically. Structurally, SARS-CoV-2 has four main structural proteins including spike (S) glycoprotein, small envelope (E) glycoprotein, membrane (M) glycoprotein, and nucleocapsid (N) protein, and also several accessory proteins (Jiang *et al*., 2020).

The study focused on the surface glycoprotein of SARS-COV-2 since it is the immunogenic part of the virus that binds to the host cell receptor (ACE2), mediating viral cell entry. SARS-COV-2 gene data bank was used to identify the 3D configurations of the S glycoprotein of SARS-COV-2 and its variants that were detected all over the world and determine the genomic locations of mutations of the new variants of S glycoprotein in addition to the codon change, protein change type and collection location (Table 5).

This corona virus changes very slowly compared to similar viruses that cause influenza. In light of the population's relatively low immunity to it, with a scarcity of effective treatments, there is no pressure on the virus to try to adapt. So far it is doing a pretty good job of keeping its spread intact. The, D614G (prominent mutation) forms the spiny protrusions to penetrate the host cells, appeared Wuhan and possibly Italy. It is now observed in about 97 percent of samples worldwide.

The SARS-CoV-2 S protein variant D614G confers a replication advantage to SARS-CoV-2, such that it increases the likelihood of human-to-human transmission. Some research groups found that such an association exists (Furuyama *et al*., 2020; Korber *et al*., 2020 and Volz *et al*., 2021). Future prospective comparisons of D614G transmission to that of D614 seem unlikely given that D614G has gone to near fixation worldwide (Yurkovetskiy *et al*., 2020). The sequenced genomes of SARS-CoV-2 are only a narrow snapshot of the COVID-19 pandemic and archived samples sequenced may pinpoint the origin of D614G and/or better resolve the variant's trajectory. Indirect evidence that D614G is more infectious was provided by experiments with pseudotyped viruses showing that D614G transduces 3- to 9-fold more efficiently than does the ancestral S protein (Yurkovetskiy *et al*., 2020). Effect was on cellular targets, as colon epithelial cells and lung. Researches are running to compare the replication efficiency of D614G with D614 in the context of ~30,000 nucleotide SARS-CoV-2 genome, which technically difficult and potentially confounded by acquisition of unnatural, tissue-culture-adapted mutations during expansion in transformed cell lines and genome rescue, which done during similar assessments of Ebola virus variants (Marzi *et al*., 2018; Ruedas *et al*., 2017 and Wang *et al*., 2017). Consistent with the increased ability of D614G to infect cells in tissue culture, several stud-

ies suggest that D614G is associated with increased viral load in people with COVID-19 (Korber *et al*., 2020; McNamara *et al*., 2020 and Volz *et al*., 2021), although these studies quantitated SARS-CoV-2 RNA and did not measure infectious virus. Although, high infectivity of D614G in tissue culture, and high viral load in infected people, the severity of COVID-19 not detected in association with D614G infection (Korber *et al*., 2020; McNamara *et al*., 2020 and Volz *et al*., 2021). The study proved that the receptor binding domain of D614G variant of S glycoprotein is affected in such a way that its affinity for binding with ACE2 and the fab fragment of the neutralized antibody is reduced, this finding is matched with (Yurkovetskiy *et al*., 2020) and that enhances transmission, infectivity and is likely to be advantageous for immune evasion.

SUMMARY

The coronavirus pandemic swept across the world in 2020 and changed the pace, texture and nature of our lives. The causative agent belongs to the RNA Coronavirus, also called SARS-CoV-2. Viruses constantly change through mutations and variations, due to evolution and adaptation processes, have been observed worldwide. Most emerging mutations have insignificant impact on the virus spreading, combinations of mutations or some mutations could provide a selective advantage, like increased transmissibility, infectivity or the capability to avoid the host immune response. An emergent Aspartic acid identified in 614 is substituted by Glycine (Asp 614 →Gly) (D614G) in the spike glycoprotein of SARS-CoV-2 strains which is the prevalent form globally. The study provides a computational analysis to compare the wild type of SARS-Cov-2 with a new variant (D614G) and find out the difference between them in terms of mutations and immunological features and their receptor binding domains which is a key part of the virus that is located on its' spike protein that facilitates its entrance into host cells and lead to infection. The study found that the D614 G variant exhibit alteration in the RBD that affect its pathogenicity, infectivity and reduced antibody binding and immune protection and is likely to be advantageous for immune evasion. It is the primary target in the prevention and treatment of infections.

REFERENCES

Alberto Ambesi-impiombato and Dego d. Bernardo (2006). Computational Biology and Drug Discovery: From Single-Target to Network Drugs", Current Bioinformatics 1: 3-10.

Altschul S. F., Gish W., Miller W., Myers E.W. and Lipman D. J., (1990). Basic local alignment search tool. Journal of molecular biology, 215(3): 403-410.

Astuti I., (2020). Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response. Diabetes & Metabolic Syndrome:

Clinical Research & Reviews, 14(4): 407-412.

Bai H. X., Hsieh B., Xiong Z., Halsey K., Choi J. W., Tran T. M. L., Pan I., Shi L. B., Wang D. C., Mei J. and Jiang X. L., (2020). Performance of radiologists in differentiating COVID-19 from non-COVID-19 viral pneumonia at chest CT. Radiology, 296(2): E46-E54.

Berman H. M., (2008). The protein data bank: a historical perspective. Acta Crystallographica Section A: Foundations of Crystallography, 64(1): 88-95.

Channappanavar R., Fehr A. R., Vijay R., Mack M., Zhao J., Meyerholz D. K. and Perlman S., (2016). Dysregulated type I interferon and inflammatory monocyte-macrophage responses cause lethal pneumonia in SARS-CoV-infected mice. Cell Host & Microbe, 19(2): 181-193.

Channappanavar R. and Perlman S., (2017). Pathogenic human coronavirus infections: causes and consequences of cytokine storm and immunopathology. In Seminars in Immunopathology, Vol. 39(5): 529-539. Springer Berlin Heidelberg.

Fehr A. R. and Perlman S., (2015). Coronaviruses: an overview of their replication and pathogenesis. Coronaviruses: pp:1-23.

Furuyama T. N., Antoneli F., Carvalho I. M., Briones M. R. and Janini L. M., (2020). Temporal data series of COVID-19 epidemics in the USA, Asia and Europe suggests a selective sweep of SARS-CoV-2 Spike D614G variant. arXiv preprint arXiv:2006-11609.

Gentleman R. C., Carey V. J. and Bates D. M. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5, R80, Article number: R80.

Holloway Eric (2020). Tutorial: Bioinformatics Basics, Communications of the Blyth Institute 35-38.

Hou Y. J., Chiba S., Halfmann P., Ehre C., Kuroda M., Dinnon K. H., Leist S.R., Schäfer A., Nakajima N., Takahashi K. and Lee, R.E., (2020). SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. Science, 370(6523) :1464-1468.

Hu Z., Ge Q., Li S., Jin L. and Xiong M., (2020). Artificial intelligence forecasting of covid-19 in China. arXiv preprint arXiv: 2002-07112.

Jiang S., Hillyer C. and Du L., (2020). Neutralizing antibodies against SARS-CoV-2 and other human coronaviruses. Trends in Immunology, 41(5): 355-359.

Lu R., Zhao X., Li J., Niu P., Yang B., Wu H., Wang W., Song H., Huang B., Zhu N. and Bi Y., (2020). Genomic characterization and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet, 395(10224): 565-574.

Madden T., (2002). Chapter 16. The BLAST Sequence Analysis Tool. The NCBI Handbook.

Marzi A., Chadinah S., Haddock E., Feldmann F., Arndt N., Martellaro C., Scott D. P., Hanley P. W., Nyenswah T. G., Sow S. and Massaquoi M., (2018). Recently identified mutations in the Ebola virus-Makona genome do not alter pathogenicity in animal models. Cell Reports, 23(6): 1806-1816.

McNamara R. P., Caro-Vegas C., Landis J. T., Moorad R., Pluta L. J., Eason A. B., Thompson C., Bailey A., Villamor F. C. S., Lange P. T. and Wong J. P., (2020). High-density amplicon sequencing identifies community spread and ongoing evolution of SARS-CoV-2 in the Southern United States. Cell Reports, 33(5): 108352.

Tanner J. A., Watt R. M., Chai Y. B., Lu L. Y., Lin M. C., Peiris J. M., Poon L. L., Kung H. F. and Huang J. D., (2003). The severe acute respiratory syndrome (SARS) coronavirus NTPase/helicase belongs to a distinct class of 5′ to 3′ viral helicases. Journal of Biological Chemistry, 278(41): 39578-39582.

Rabi F. A., Al Zoubi M. S., Kasasbeh G. A., Salameh D. M. and Al-Nasser A. D., (2020). SARS-CoV-2 and coronavirus disease 2019: what we know so far. Pathogens, 9(3): 231.

Ruedas J. B., Ladner J. T., Ettinger C. R., Gummuluru S., Palacios G. and Connor J. H., (2017). Spontaneous mutation at amino acid 544 of the Ebola virus glycoprotein potentiates virus entry and selection in tissue culture. Journal of Virology, 91(15): e00392-17.

Simmons G., Zmora P., Gierer S., Heurich A. and Pöhlmann S., (2013). Proteolytic activation of the SARS-coronavirus spike protein: Cutting enzymes at the cutting edge of antiviral research. Antiviral Research, 100: 605-614.

Tai W., He, L., Zhang X., Pu J., Voronin D., Jiang S., Zhou Y. and Du L., (2020). Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. Cellular & Molecular Immunology, 17(6): 613-620.

Vaishya R., Javaid M., Khan I. H. and Haleem A., (2020). Artificial Intelligence (AI) applications for COVID-19 pandemic. Diabetes &

Metabolic Syndrome: Clinical Research & Reviews, 14(4): 337-339.

Volz E., Hill V., McCrone J. T., Price A., Jorgensen D., O'Toole Á., Southgate J., Johnson R., Jackson B., Nascimento F. F. and Rey S. M., (2021). Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. Cell, 184(1): 64-75.

Walls A. C., Park Y. J., Tortorici M. A., Wall A., McGuire A. T. and Veesler D., (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell, 181(2): 281-292.

Wang M. K., Lim S. Y., Lee S. M. and Cunningham J. M., (2017). Biochemical basis for increased activity of Ebola glycoprotein in the 2013–16 epidemic. Cell Host & Microbe, 21(3): 367-375.

Wu F., Zhao S., Yu B., Chen Y. M., Wang W., Song Z. G., Hu Y., Tao Z. W., Tian J. H., Pei Y. Y. and Yuan M. L., (2020). A new coronavirus associated with human respiratory disease in China. Nature, 579 (7798): 265-269.

Yurkovetskiy L., Wang X., Pascal K. E., Tomkins-Tinch C., Nyalile T. P., Wang Y., Baum A., Diehl W. E., Dauphin A., Carbone C. and Veinotte K., (2020). Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant. Cell, 183(3): 739-751.

Zhou B., Thao T. T. N., Hoffmann D., Taddeo A., Ebert N., Labroussaa F., Pohlmann A., King J., Steiner S., Kelly, J. N. and Portmann J., (2021). SARS-CoV-2 spike D614G change enhances replication and transmission. Nature, 592 (7852): 122-127.

Table (1): Shows Accessions numbers, geo-locations, collection dates and length of each variant in base pairs.

Gene bank Accession No.	Geo Location:	Collection Date:	Molecular Type:	Length in base pairs:
NC_045512	China	2019-12	RNA	29903
MW617293	Pakistan	2021-01-25	RNA	29822
MW592707	Brazil	2021-02-11	RNA	29862
MW595907	Egypt	2021-01-03	RNA	29793

Table (2): types of mutations in Brazil.

Wild type (Subject)	RNA complment	Amino acid	New variant strain (Query)	RNA complement	Amino acid	effectiveness	Type of mutation
CGT	GCA	Ala	TGT	ACA	Thr	Active	Missense mutations
TAG	AUC	Ile	TAA	AUU	Ile	In-active	Silent mutations
GGG	CCC	Pro	ACG	UGC	Cys	Active	Missense mutations

Table (3): types of mutations in Pakistan.

Wild type (Subject)	RNA complement	Amino acid	New variant strain (Query)	RNA complement	Amino acid	effectiveness	Type of mutation
GCT	CGA	Arg	GTT	CAA	Gln	Active	Missense mutations
GGT	CCA	Pro	GAT	CUA	Leu	Active	Missense mutations
AAT	UUA	Leu	GAT	CUA	Leu	In-active	Silent mutations

Table (4): types of mutations in Egypt.

Wild type (Subject)	RNA complement	Amino acid	New variant strain (Query)	RNA complement	Amino acid	effectiveness	Type of mutation
TGA	ACU	Thr	CGA	GCU	Ala	Active	Missense mutation
AAA	UUU	Phe	AAG	UUC	Phe	In-active	Silent mutation
GTT	CAA	Gln	TTT	AAA	Lys	Active	Missense mutation
CAC	GUG	Val	CAT	GUA	Val	In-active	Silent mutation
TTC	AAG	Lys	TTT	AAA	Lys	In-acive	Silent mutation
GCC	CGG	Arg	GCT	CGA	Arg	In-active	Silent mutation
CTC	GAG	Glu	TTC	AAG	Lys	Active	Missense mutation

Table (4):Cont'

ATA	**UAU**	**Tyr**	**GAP**	**---**	**---**	**Active**	**Deletions non frameshift mutation**
CAT	**GUA**	**Val**	**GAP**	**---**	**---**	**Active**	**Deletions non frameshift mutation**
GTC	**CAG**	**Gln**	**GAP**	**---**	**---**	**Active**	**Deletions non frameshift mutation**
ACG	**UGC**	**Cys**	**ACT**	**UGA**	**STOP**	**Active**	**Nonsense mutation**
AAG	**UUC**	**Phe**	**AGG**	**UCC**	**Ser**	**Active**	**Missense mutation**
TTT	**AAA**	**Lys**	**TTC**	**AAG**	**Lys**	**In-active**	**Silent mutation**
GCG	**CGC**	**Arg**	**GTG**	**CAC**	**His**	**Active**	**Missense mutation**
AGT	**UCA**	**Ser**	**GAP**	**---**	**---**	**Active**	**Deletions non frameshift mutation**

Table (5): The variants of the Surface glycoprotein of SARS- COV-2 and the 3D configurations of them.

Protein Id	Genomic location	Codon Change	Protein change type	Collection location	3D-configurations
1-6VXX	**23157**	**No codon change**	**No change**	**All over the world**	
2-A520E	**23121**	**GAA >GCA**	**Non-synonymous**	Australia (141) - USA (18) - South Africa (6)	
3-D614G	23403	GGT > GAT	non_synonymous	Australia (13942) - USA (11175) - South Africa (406) - Russia (241) - Israel (194) - Brazil (144) - India (102) - Czech Republic (51) - France (37) - Philippines (22) - United Kingdom (20) - Netherlands (12) - Qatar (11) - Chile (8) - Mozambique (7) - Turkey (6) - Ecuador (4) - South Korea (4) - China (3) - Peru (2) - Germany (2) - Bangladesh (1) - Malaysia (1) - Croatia (1) - Morocco (1	

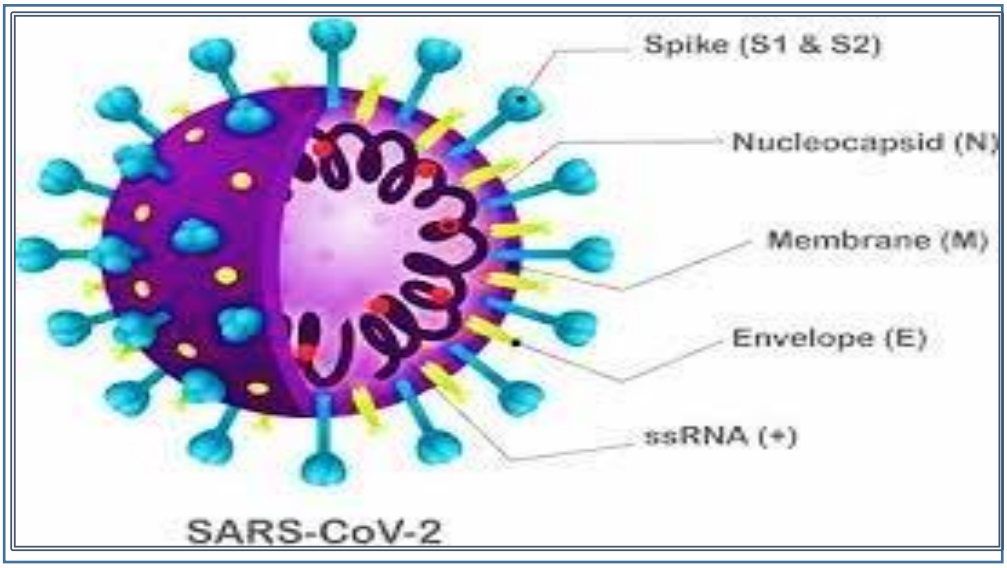

Fig.(1): Structure of SARS-CoV2 variant. Figure from "Antivirals Against Coronaviruses: Candidate Drugs for SARS-CoV-2 Treatment? " Santos *et al.* 2020. Open-access article licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) Copyright © 22020 Santos, Grosche, Bergamini, Sabino-Silva and Jardim.

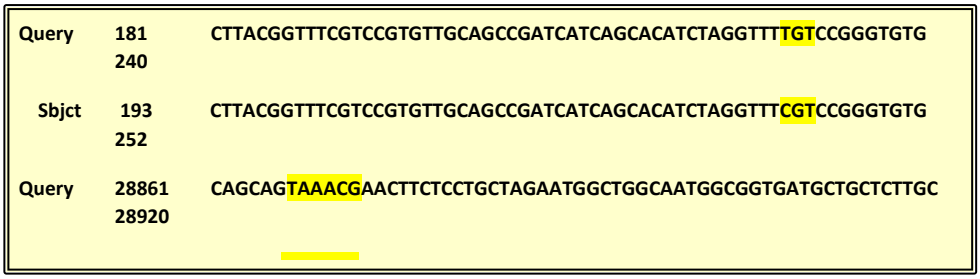

Fig. (2): Alignment between the wild type and new variant in Brazil to indicate the location of mutations.

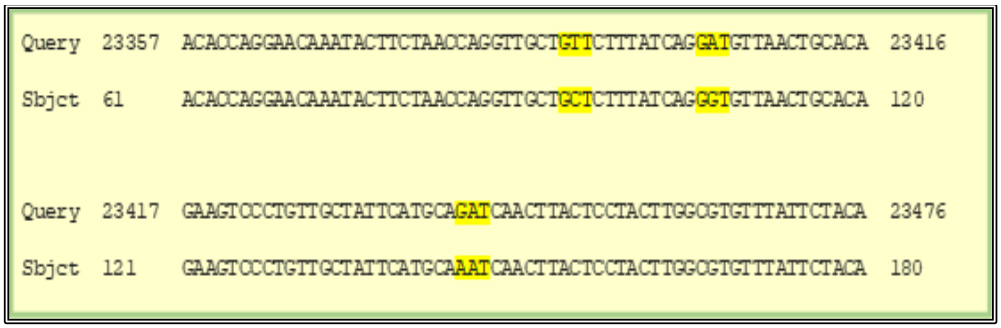

Fig. (3): Alignment between the wild type and new variant in Pakistan to indicate the location of mutations

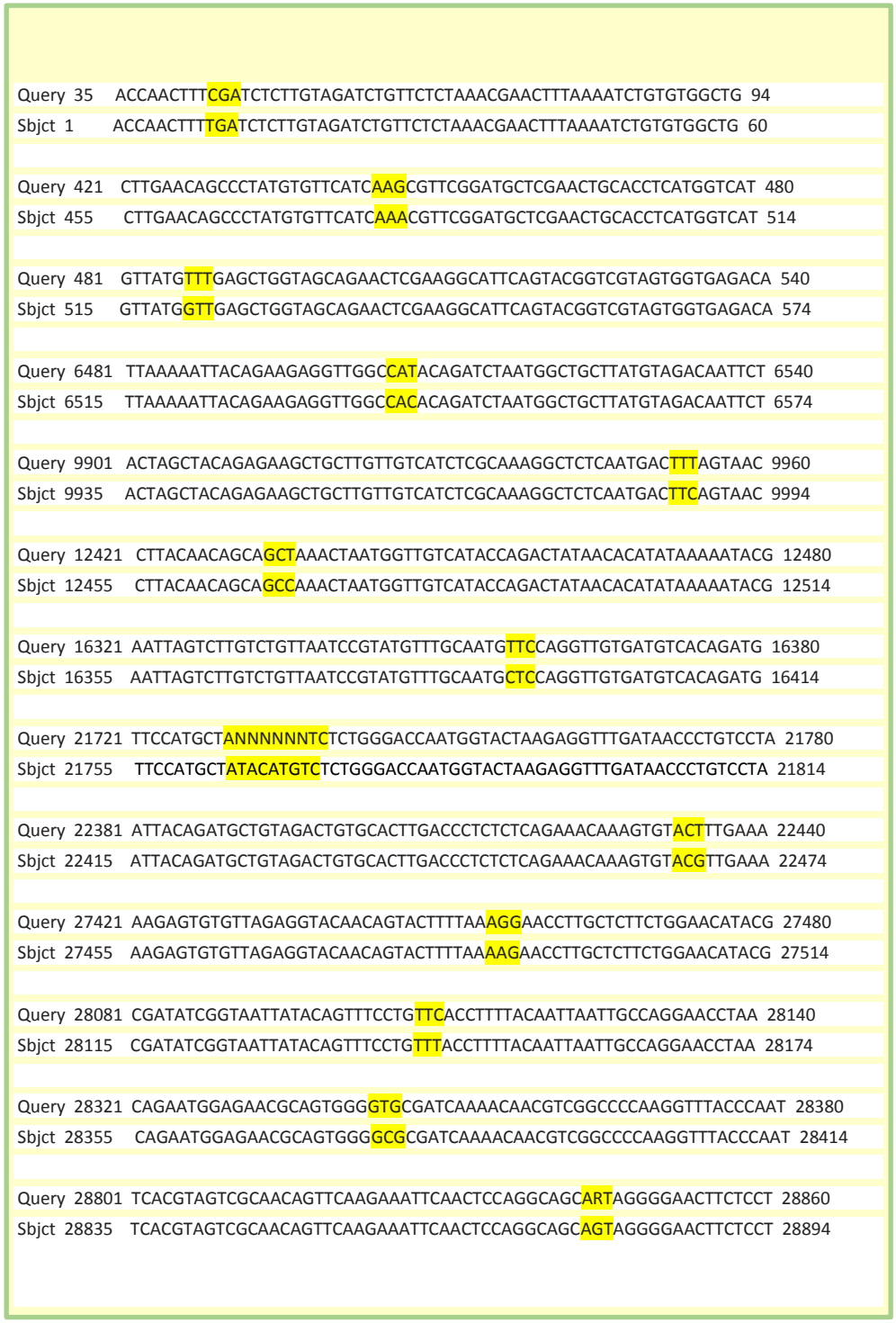

Fig. (4): Alignment between the wild type and new variant in Egypt to indicate the location of mutations.

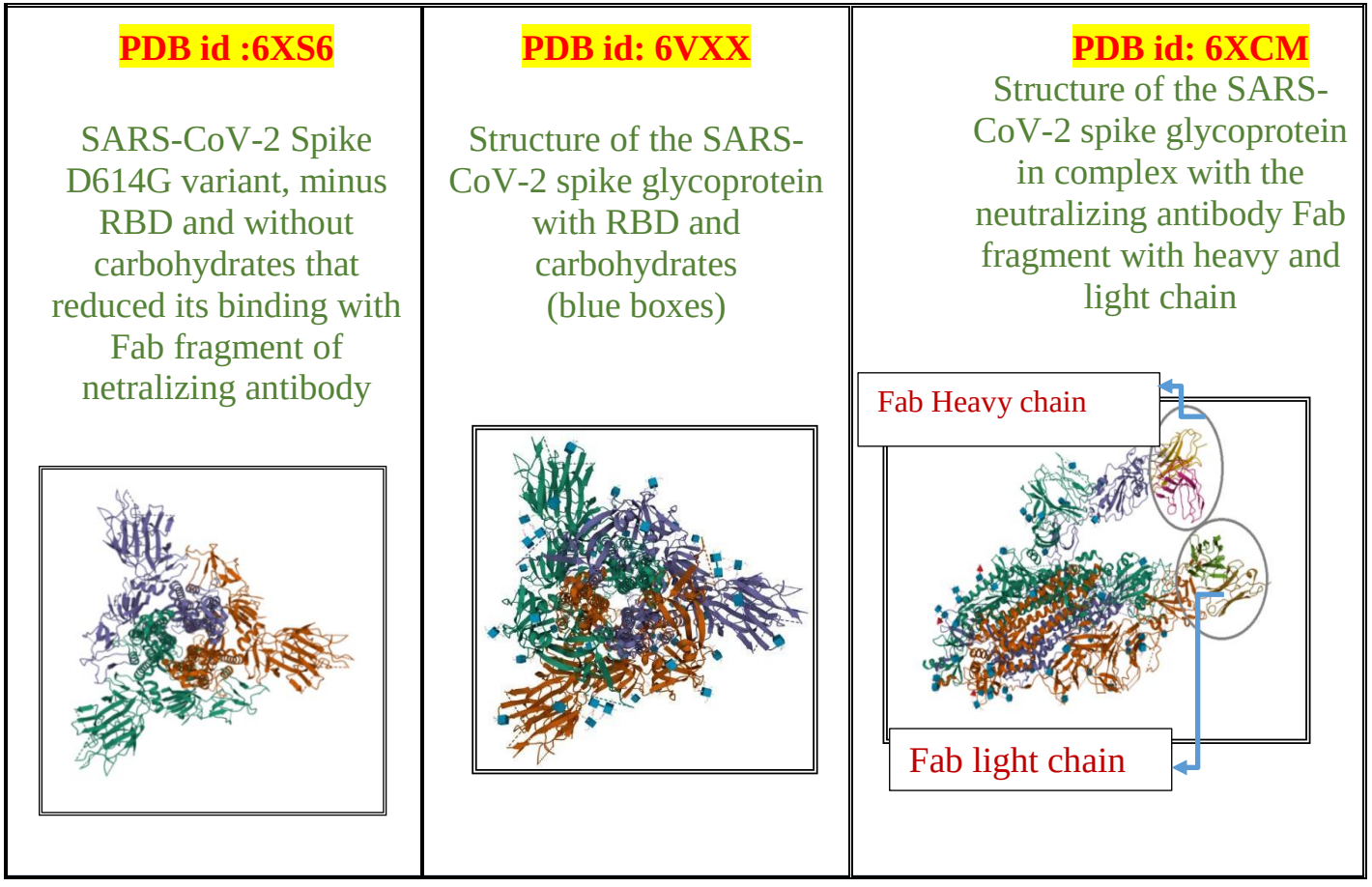

Fig (5): The difference in 3D configuration between the wild type of SARS-COV-2 spike glycoprotein and the D614G variant.

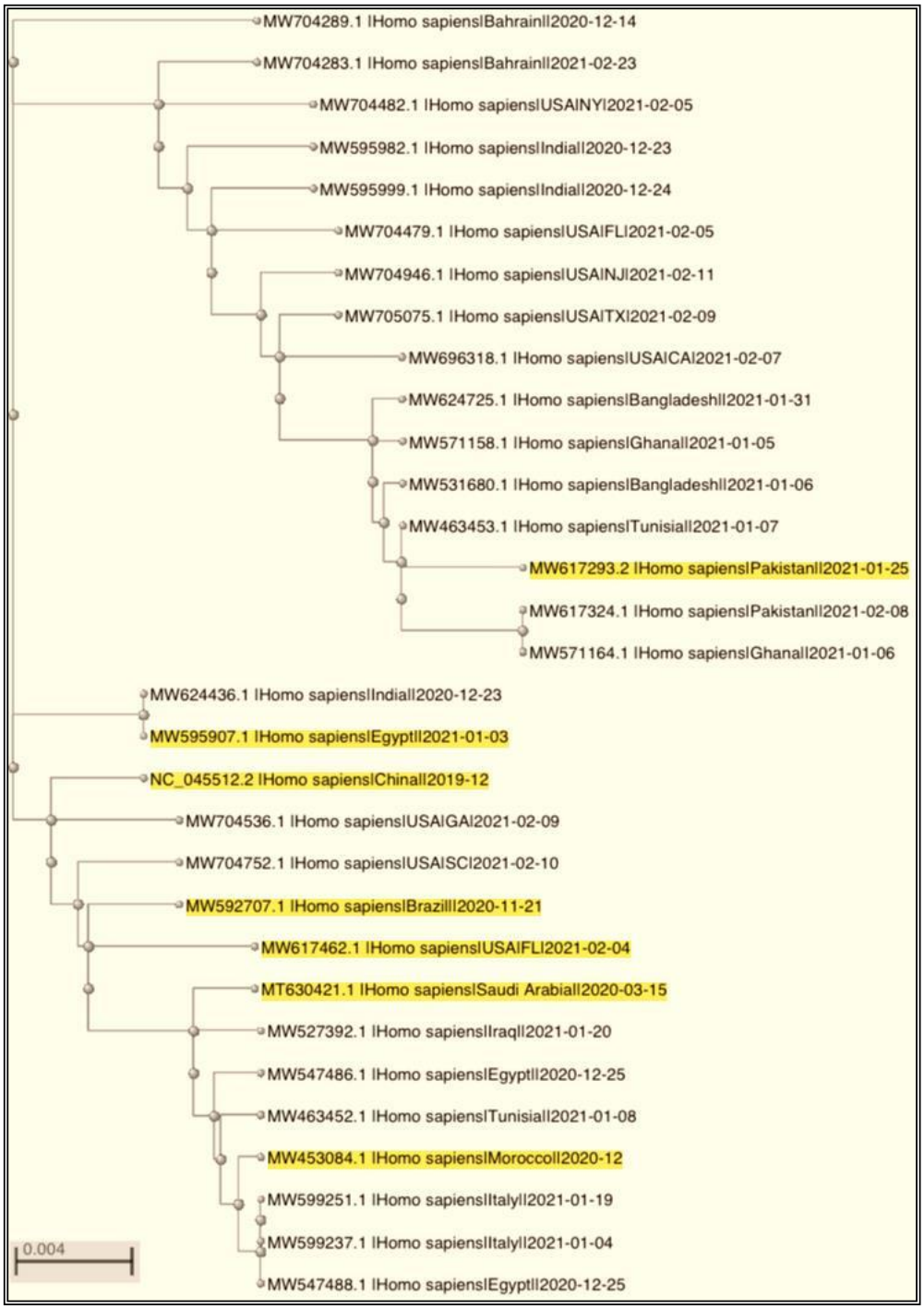

Fig. (6): Phylogenetic tree of 30 variants of SARS-COV-2 isolated from different countries all over the world and their relationship with the wild type..