HIGH SCHOOL EDITION
Journal of Student Research
Volume 11 Issue 3 (2022)

Genomic Analysis of S protein in SARS-CoV-2 samples from Washington D.C

Nikila Swaminathan[1]

[1]Lowery Freshman Center/High School

ABSTRACT

COVID-19, otherwise known as the coronavirus, ranks as one of the seven world's worst epidemics since recorded history. As severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) evolves, it accumulates changes in the genetic code, commonly known as mutations. The greater spreadability of the SARS-CoV-2 variants is hypothesized to be driven by key mutations that are distinguished by higher mutation frequency and mutation density. This study investigates variants of the SARS-CoV-2 virus and identifies mutations that are of particular importance for future antibody, vaccine, and drug development. The human spike protein sequences from Washington D.C were analyzed by being compared to the reference SARS-COV-2 protein sequence from Wuhan-Hu-1. Out of 7012 sequences analyzed, a total of 2463 distinct mutations, at 1251 sites were observed. The mutations observed in the spike proteins are discussed in the context of mutation frequency, characteristics of mutations at different regions, number of mutations per sample, and mutation density at different regions. Different variants observed in the sequences, and how they have adapted from the existing variants such as Alpha, Delta, and Omicron are also discussed in this research paper. This study summarizes the mutation characteristics that influenced the virus' advantage and identifies the mutations to be considered for future antibody, vaccine, and drug development against the virus.

Introduction

SARS-COV-2 is the virus that causes the disease COVID-19. It has been sequenced often from patients around the world [AA]. Viruses mutate to adapt to various environmental factors, climate, and populations. Viruses infect the host and propagate inside the host cells using the spike (S) protein. Understanding the genetic variation with the characterization of the rates and patterns of mutation shall support knowing virus evolution.

As of the 2nd of Jun 2022, there are 84,315,762 total new SARS-CoV-2 cases, and about 1,002,993 deaths were reported in the USA [BB]. With the threat of new variants constantly hanging over our heads, it is important to be prepared and equipped to tackle the new variants.

The purpose of this project is to analyze and interpret DNA sequence data to understand how SARS-CoV-2 has evolved over time during an outbreak, to recognize how different mutations can affect the virus's spreading ability and its prevalence, to understand the change in characteristics of the virus that may lead to a new variant of concern (VOC) [CC], and Lastly to know whether a new vaccine is needed to deliver antigens matching the mutated virus.

Hypothesis

Mutation frequency (independent variable) and mutation density (independent variable) in a SARS-COV2 S protein sequence region can have an effect on the selective advantage of the virus (dependent variable).

Research questions

The study also tries to answer the below research questions

- How does the frequency of mutations and mutation density in different regions of the protein sequence influence the selective advantage for the virus?
- Any specific mutations that could be considered for vaccine or drug development?
- Any mutation predicted in the future variant that could be useful to create precautionary doses?
- When and where did the mutations occur? Has any of the mutations become prevalent over time?
- Does a type of mutation make the virus spread or break past the host's immunity more easily than the other?

Methods

Dataset

The study uses all available spike protein (surface glycoprotein) sequences of samples from the US/DC region from Mar'20 to May'22. 8265 surface glycoprotein amino acid sequences of SARS-CoV-2 related to the COVID-19 pandemic were retrieved from NCBI Virus Variation Resource repository.

Among the 8265 surface glycoprotein amino acid sequences, 186 were discarded due to incomplete sequence coverage (less than 99% of 1273). In addition, the alignment tool removed 1067 sequence(s) with more than 5.0% ambiguous letters. Hence, the remaining protein sequences taken for analysis are 7012 surface glycoprotein amino acid sequences of SARS-CoV-2 (Samples for this experiment).

Tools

To conduct the study, various open tools were used. UGENE tool [DD] was used to annotate protein sequences and identify mutations. NCBI alignment tool was tried but was not used because it was not convenient for a large number of sequences, hence I opted to use the MFFTT tool to align the sequences [EE]. Primarily I used the downloaded Python code that utilized the biopython library to identify the mutations that occurred in each sequence. Microsoft Excel was used to analyze the mutation data.

Procedure

Firstly, I collected SARS-COV2 protein sequences in the FASTA format from the NCBI virus database. Then aligned the sequences using MAFFT to get it in .aln format (clustal format). Wuhan-Hu-1, China YP_009724390 was used as the canonical/reference sequence for the analyses of SARS-CoV-2 S glycoprotein mutations. Using python scripts identified different mutations that occurred in the spike protein on the given sample.

With the identified list of mutations and variants, the data analysis was done by using Excel to analyze which mutations became a Selective advantage or disadvantage to the virus (based on the amount of population each mutation spread to). In addition, mutation frequency, number of mutation, mutation density, mutation variability at a site and region analysis were undergone.

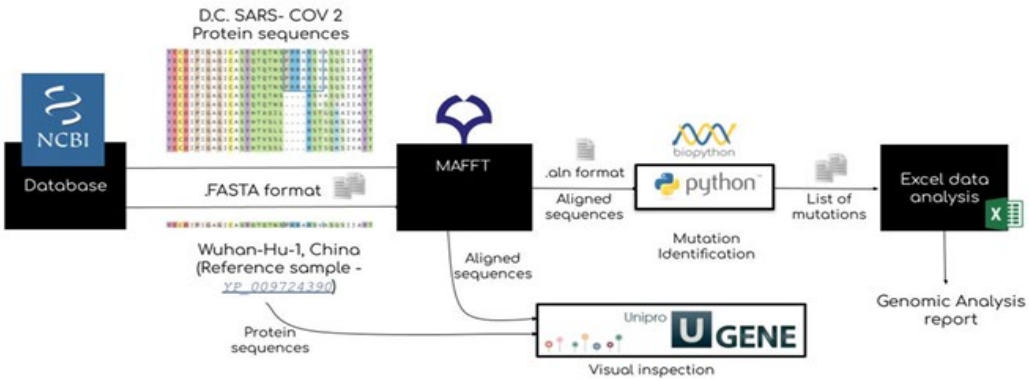

Figure 1. Process of the Experiment

Observation

The spike protein mutations that were observed in Washington D.C had 7012 (pre-Omicron 3450) Sequences, 1273 Positions/sites in the Spike protein sequence, 2464 (pre-Omicron 1453) Unique mutations, and 1251 (pre-Omicron 1032) positions/sites where the mutations have occurred.

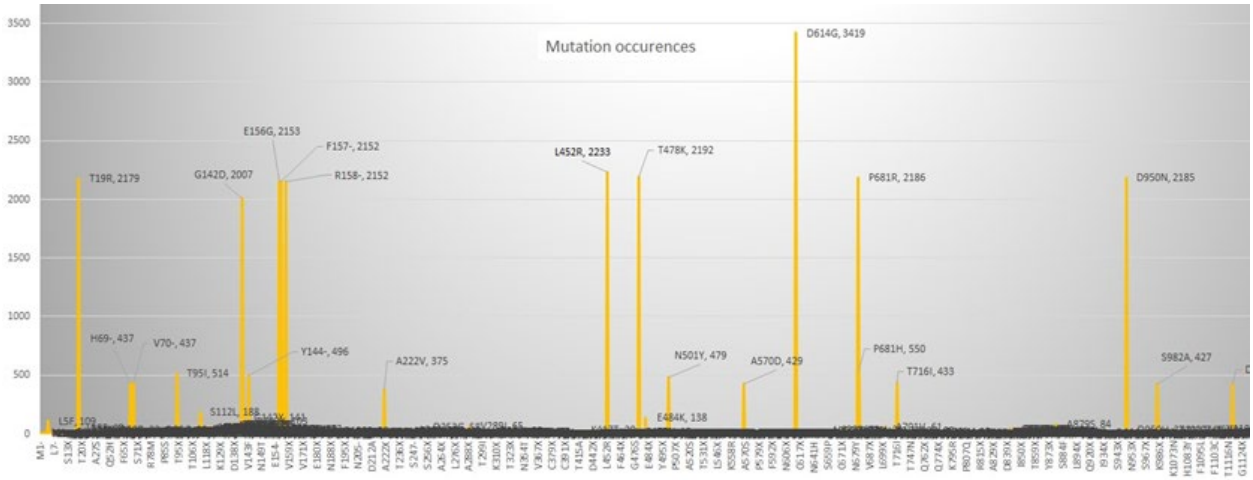

Figure 2. pre-Omicron COVID cases in D.C samples until Nov'21

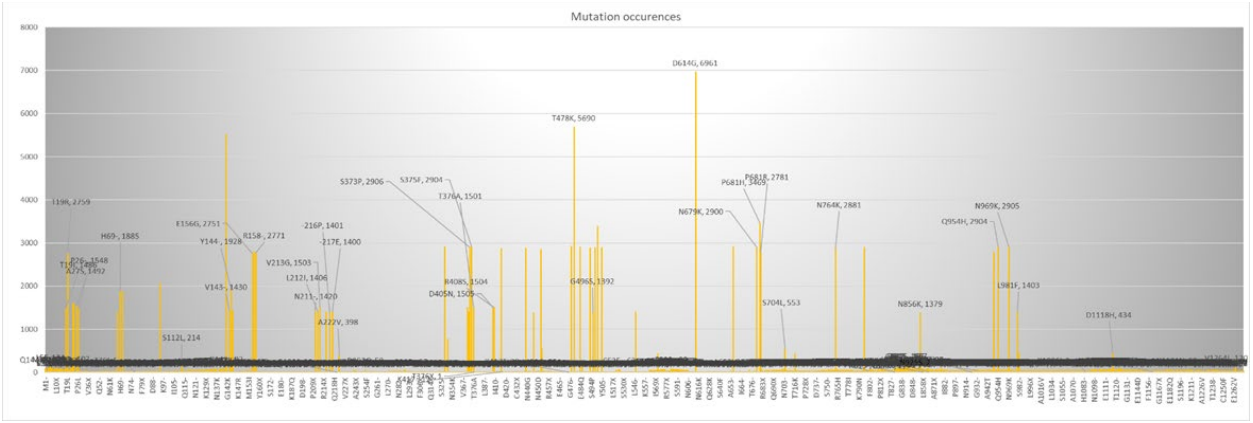

Figure 3. COVID cases Including Omicron until May 22

D614G mutation occurred in>99% of Washington D.C samples followed by T478K with ~82% and G142D with ~78%. D614G is in the S2 Domain of the S1 subunit which was reported to enhance SARS-CoV-2 infectivity.

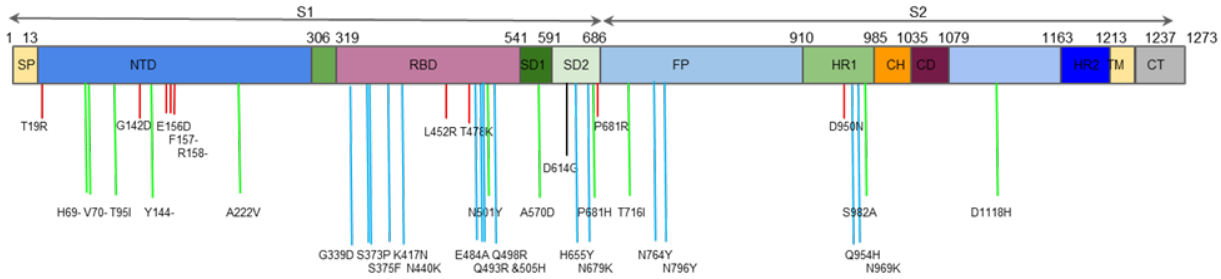

Figure 4. Notable mutations of different regions

Fig 4 depicts each mutation that occurred in the samples in the respective regions of the spike protein. The different colors (green, red, blue, and black) denote 4 clusters of prominent mutations. Blue is a new cluster observed with omicron cases. During pre-omicron, many notable mutations have occurred in the N-terminal domain but during the omicron period, the number of new mutations in the RBD region is more than in other regions. Overall, 57 noticeable variations/mutations were cited.

- 23 mutations (T19R, H69Del, V70Del, T95I, G142D, Y144Del, Y145Del, E156G, F157Del, and R158Del + (in omicron cases) T19I, L24Del, P25Del, P26Del, A27S, V143Del, N211Del, L212I, V213G, -215E, -216P, -217E) were located at the N-terminal domain (NTD). It was only 3 more compared to the pre-omicron period - before Nov'21.
- 22 mutations (L452R, T478K, E484K, N501Y + G339D, R346K, S371F, S371L, S373P, S375F, T376A, D405N, R408S, K417N, N440K, G446S, S477N, E484A, Q493R, G496S, Q498R, Y505H) were found at the receptor-binding domain (RBD), and especially 12 were in RBM; these mutations help the virus bind more tightly to human cells [1].
- Less than five variations in all other regions of the S protein.

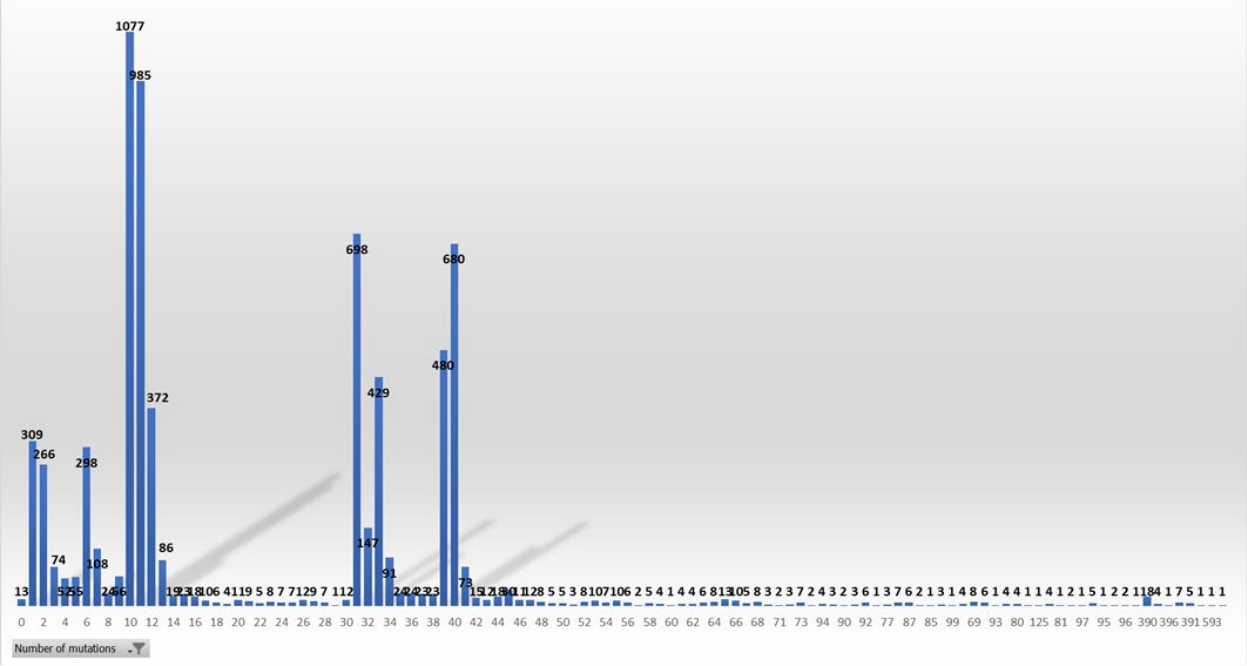

Figure 5. Distribution of occurrence number of mutations

From the above figure 5, we observe that most of the samples (>80%) had about 1 to 40 mutations(s) per sample. On average, 60 mutations per sample were observed. Compared to the 11 mutations pre-omicron (before Nov 21).

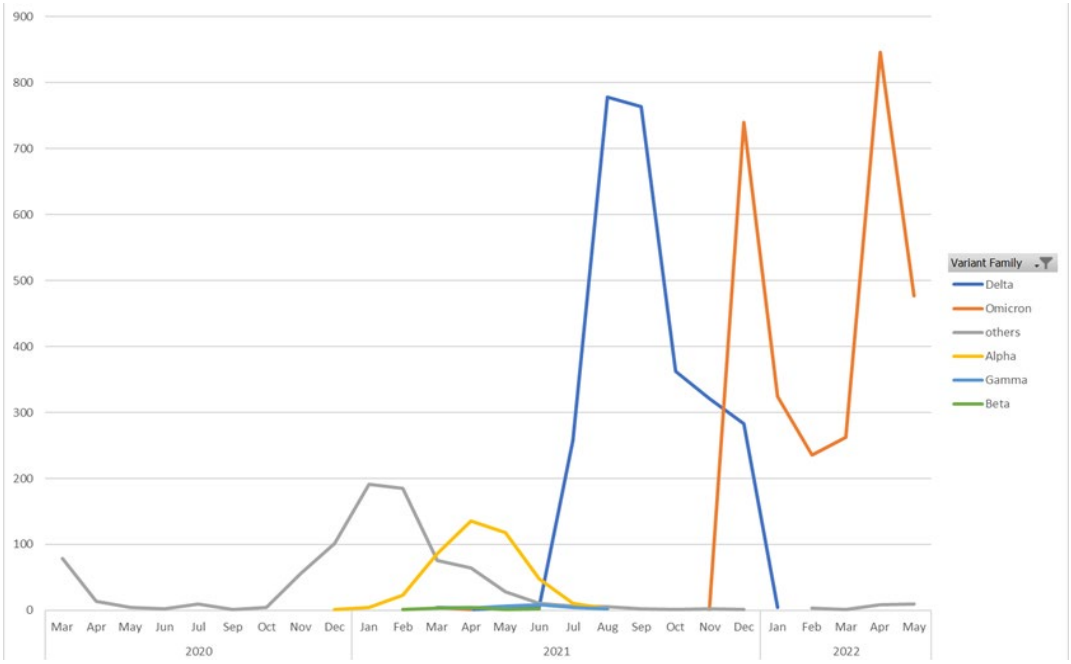

Figure 6. Occurrence of each variant over time

With respect to the variants analyzed, there were 960 variants, out of which 7 were prominent as seen in the below table 1.

Table 1. Mutation frequency and density of different regions of pre-omicron period

Mutations seen in the variant	Variant name
T19R G142D E156G F157- R158- L452R T478K D614G P681R D950N	Delta
T19I, L24-, P25-, P26-, A27S, G142D, V213G, G339D, S371F, S373P, S375F, T376A, D405N, R408S, K417N, N440K, S477N, T478K, E484A, Q493R, Q498R, N501Y, Y505H, D614G, H655Y, N679K, P681H, N764K, D796Y, Q954H, N969K	Omicron
A67V, H69-, V70-, T95I, G142D, V143-, Y144-, Y145-, N211-, L212I, -215E, -216P, -217E, G339D, R346K, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, L981F	Omicron
A67V, H69-, V70-, T95I, G142D, V143-, Y144-, Y145-, N211-, L212I, -215E, -216P, -217E, G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, L981F	Omicron
T19I, L24-, P25-, P26-, A27S, G142D, V213G, G339D, S371F, S373P, S375F, T376A, D405N, R408S, K417N, N440K, L452Q, S477N, T478K, E484A, Q493R, Q498R, N501Y, Y505H, D614G, H655Y, N679K, P681H, S704L, N764K, D796Y, Q954H, N969K, N764K, D796Y, N856K, Q954H, N969K	Omicron
T19R, T95I, G142D, E156G, F157-, R158-, L452R, T478K, D614G, P681R, D950N	Delta

In Top 7, 6 variants were with ten or more mutations. D614G, a point mutation [2] is one of the top variants observed. (with >200 repetitions). Very limited number of Alpha, Beta, and Gamma variants were observed.

The mutation density (MD) is evaluated as a function of the number of mutations observed over the sequence length corresponding to different regions in the spike protein [3]. Mutation frequency (MF) is evaluated as the number of total mutations observed over the number of samples. MD and MF in diff regions of S protein observed in the samples are tabulated below.

Table 2. Mutation frequency and density of different regions of pre-omicron period

Regions	Sequence length	Number of distinct mutation	Total number of mutations	Mutation Density	Mutation Frequency
signal peptide (N-terminus)	13	31	216	**2.38**	0.06
N-terminal domain	292	456	18864	**1.56**	**5.47**
receptor-binding domain (RBD)	223	224	6023	1.00	**1.75**
SD1 & SD2	145	168	7295	1.16	**2.11**
fusion peptide (FP)	47	60	210	1.28	0.06
connecting region	77	93	947	1.21	0.27
heptapeptide repeat sequence 1 (HR1)	73	76	2808	1.04	0.81
central helix (CH)	50	12	51	0.24	0.01
connector domain (CD)	85	129	989	**1.52**	0.29
heptapeptide repeat sequence 2 (HR2)	51	57	182	1.12	0.05
transmembrane domain (TM)	25	19	44	0.76	0.01
cytoplasm domain	37	22	162	0.59	0.05

Table 3. Mutation frequency and density of different regions of cases including omicron

Regions	Sequence length	Number of distinct mutation	Total number of mutations	Mutation Density	Mutation Frequency
signal peptide (N-terminus)	13	28	1328	2.15	0.38
N-terminal domain	292	742	53832	**2.54**	**15.60**
receptor-binding domain (RBD)	223	401	54791	**1.80**	**15.88**
SD1 & SD2	145	279	24601	1.92	**7.13**
fusion peptide (FP)	47	85	5295	1.81	1.53
connecting region	77	137	6797	1.78	1.97
heptapeptide repeat sequence 1 (HR1)	73	116	13486	1.59	3.91
central helix (CH)	50	64	1944	1.28	0.56
connector domain (CD)	85	155	5023	1.82	1.46
heptapeptide repeat sequence 2 (HR2)	51	75	2117	1.47	0.61
transmembrane domain (TM)	25	40	1006	1.60	0.29
cytoplasm domain	37	66	1627	1.78	0.47

As seen from Table 3, the N-Terminal region in the spike protein is associated with the maximum mutation density. RBD had the highest frequency followed by the N-terminal domain (NTD) and SD1 & SD2. While comparing the pre-Omicron period (depicted in Table 1), there is a huge leap in the mutation frequency and density of RBD (from 1 to 1.8 in density and 1.75 to 15.88 in frequency).

Table 4. Most frequent in the spike protein of all D.C cases

Mutation	Number of Mutation occurrences	Region
D614G	6961	SD1&SD2
T478K	5690	RBD
G142D	5511	N-Terminal domain
P681H	3469	SD1&SD2
N501Y	3389	RBD

D614G is the SARS-CoV-2 mutational hotspot in Washington D.C. T478K, G142D, P681H and N501Y mutations in NTD, RBD, and SD1 & SD2 regions were the other prominent ones.

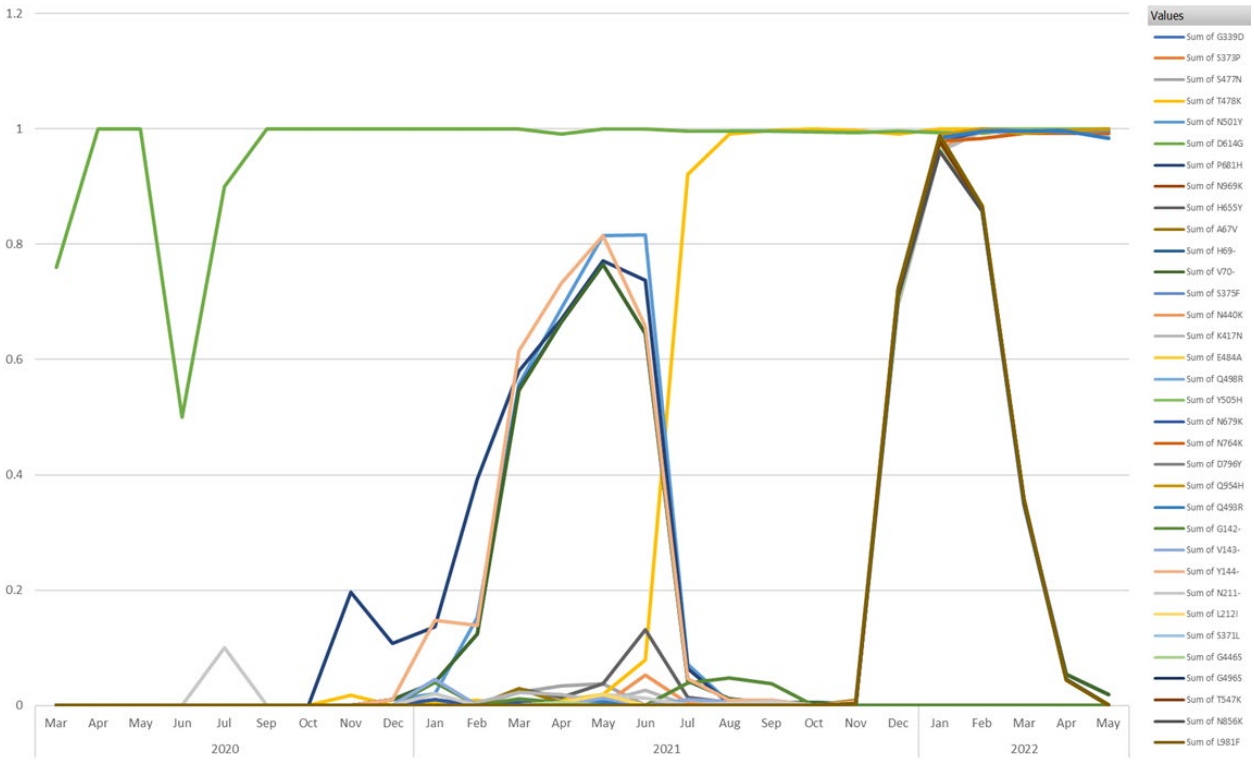

Figure 7. Change in mutation frequency of mutations in omicron variants

From figure 7, it is observed that, frequency of most mutations reached nearing 100% during the omicron period (beyond Dec'21). This indicates omicron was one of the prevalent variants from Jan 2022. The loss of a few mutations such as S371L, G496S, T547K, N856K, and L981F started in Jan 2022 and the mutation frequency of those dropped to near zero in May 2022. This indicates a new lineage of omicron spread in that timeframe.

Table 5. Prominent Mutations in the spike protein that existed across variants

Regions	T478K	N501Y	D614G	P681H	H69-	V70-	Y144-	G142D	T95I
2020	**1**	**1**	**252**	**22**	**1**	**1**	**1**	**0**	**0**
Mar	0	0	60	0	0	0	0	0	0
Apr	0	0	14	0	0	0	0	0	0
May	0	0	4	0	0	0	0	0	0
Jun	0	0	1	0	0	0	0	0	0
Jul	0	0	9	0	0	0	0	0	0
Sep	0	0	1	0	0	0	0	0	0
Oct	0	0	5	0	0	0	0	0	0
Nov	1	0	56	11	0	0	0	0	0
Dec	0	1	102	11	1	1	1	0	0
2021	**3517**	**1221**	**4550**	**1288**	**1207**	**1204**	**1265**	**3341**	**1419**
Jan	1	4	197	27	8	8	29	0	1
Feb	2	32	209	82	26	26	29	0	5
Mar	0	97	174	101	96	95	107	0	13
Apr	2	144	207	140	139	139	153	3	30
May	3	128	157	121	120	120	128	1	15
Jun	6	62	76	56	49	49	50	6	10
Jul	258	20	279	18	12	12	13	227	55
Aug	781	4	785	2	9	9	7	700	140
Sep	764	0	763	0	4	3	7	710	177
Oct	363	0	361	0	2	2	0	349	66
Nov	323	1	322	2	1	1	2	322	101
Dec	1014	729	1020	739	741	740	740	1023	806
2022	**2172**	**2167**	**2159**	**2159**	**677**	**676**	**662**	**2170**	**664**
Jan	328	323	326	322	324	324	324	328	326
Feb	239	239	237	238	205	205	206	238	206
Mar	263	263	263	262	93	92	93	263	93
Apr	855	855	849	853	46	46	39	855	39
May	487	487	484	484	9	9	0	486	0
Grand Total	**5690**	**3389**	**6961**	**3469**	**1885**	**1881**	**1928**	**5511**	**2083**

From table 5 and figure 7, it is observed that D614G mutation exists across variants Alpha, Delta and Omicron. G142D, T478K mutation exists across variants Delta, and Omicron. N501Y, P681H mutation exists across variants Alpha, and Omicron.

Summary of Observations

D614G is the SARS-CoV-2 mutational hotspot in Washington D.C. Various mutations have occurred in the N-terminal domain and the Receptor-Binding Domain of the virus, compared to other areas. The Omicron variant had many mutations and notably, the number of mutations in the RBD region increased. The average number of mutations per sample spiked five-fold during the omicron period. Delta and Omicron variants were prevalent in D.C samples. In Washington DC, the Alpha variant dropped after July while the Delta variant steadily increased spreading from June

until August before dropping. Omicron started increasing from Nov'21. During the omicron period, the frequency and density of mutations in the RBD region increased considerably. G142D, T478K, N501Y, D614G, and P681H mutations in NTD, RBD, and SD1 & SD2 regions were the prominent mutations observed. In addition these are the mutations that existed across the variant.

Results and Discussion

We have discussed below the most frequent mutations that would have influenced the spreadability considering their structural changes. G142D mutation may confer greater immune evasion, replication, and transmission advantages to the virus [4]. Mutations of T478K may result in increased ACE2 binding [5] and antibody escape [6]. N501Y residue increases the affinity for the receptor-binding domain (RBD) to its receptor ACE2 resulting in increased infectivity [7]. The binding affinity of the spike protein to the ACE2 receptor also affects the SARS-CoV-2 replication fitness and disease severity [8]. D614G provides a moderate advantage for infectivity and increases transmissibility because this mutation occurs in the RBD region, which seems to be having mutations that increase the virus' spreading capability.[9][10]. P681H helps SARS-CoV-2 to replicate faster and resist innate immunity [11].

Looking into both Mutation density and mutation frequency in different regions, mutations in the NTD and RBD regions may have an advantage for the virus to spread through a population. Positive pressure and mutations acting on RBD and NTD regions could be important to serve as potential targets for the development of prophylaxes and therapeutics.

Now, let's try to analyze the reason for increased omicron variants' spreadability from our mutation frequency observations. Omicron had more notable mutations than any other SARS-CoV-2 variants. In that, I see a few mutations such as S477N, T478K, N501Y, D614G, P681H, H655Y, A67V, H69-, V70-, N440K, K417N, N679K, Q493R, G142-, V143-, Y144-, N211-, L212I, G446S existed even in older variants.

Infectivity of SARS-CoV-2 is mainly determined by the binding affinity of the ACE2 and RBD complex, although the furin cleavage site plays a crucial role as well. G496S, Y505H, and the triple S371L, S373P, and S375F reduce affinity/binding to ACE2 in a quest to evade the neutralizing antibody response and drive immune evasion.

But, I observed that the frequency of a few mutations such as S371L, G496S, T547K, N856K, and L981F dropped to zero during May, so they may not be a selective advantage to viruses. N764K, D796Y, Q954H, and N969K are in S2, so they may not contribute to the higher immune escape and transmissibility [12].

After ruling out the possibility of influence by many notable mutations, the remaining mutations are S375F, E484A, Q498R, and Y505H. These 4 mutations happened in the positions closer to the ACE-2 receptor and are also present at the interface between the spike protein and ACE-2 receptor.

In addition, S375F mutation is responsible for low S1/S2 cleavage efficacy and extensive Omicron transmission [13]. E484A increases neutralizing antibody resistance and has the stronger vaccine-breakthrough capability. Y505H also weakens many known antibodies and RBD complexes. [14]. Q498R is likely to increase ACE2 affinity by many-fold [15], which potentially enhances the viral infectivity to the host cells.

Hence, we can conclude that these mutations could be the main reason for high spreadability of Omicron. They will also potentially affect vaccine performance and drugs designed at the interface of protein-protein interactions.

Conclusion

Mutation density (MD) and Mutation frequency (MF) can indicate the virus' spreading capability. But the selective advantage of the virus can only be identified based on the region where a mutation has occurred. Many notable mutations have occurred in the N-terminal domain, RBD, and SD2 regions. D614G is identified as the most prevalent

mutation. Data indicated that, NTD, RBD and SD2 may have a selective advantage, but the variant comparison (Omicron, Delta & Alpha) showed us mutations in RBD/RBM regions are the main cause of the virus' transmission capability. D614G evidently provides a moderate advantage for infectivity and increases transmissibility.

Variant agnostically, mutations G142D, N501Y, and P681H supported to provide a selective advantage for SARS-COV-2 by increasing the ACE2 binding and antibody escape. S375F, E484A, Q498R, and Y505H are the mutations that contributed to high spreadability of omicron and should be an important consideration for antibody, vaccine, and drug development against omicron. Mutations G142D, T478K, N501Y, D614G, and P681H existed across the variants until Omicron, hence they are highly predicted to be in the future SARS-COV2 variants. These mutations would be targets for precautionary dose developments.

In summary, though we have found out that mutation density and mutation frequency could be the cause of the virus' spreading capability, we can only be sure by identifying which region a mutation has occurred. From the data we can see that many mutations have occurred in the N-terminal region, RBD region, and SD2 region. This may lead us to think that these regions can have a selective advantage on the virus but in the end we found out that the RBD region is the main cause for the virus' transmission capability.

Further Studies

I would like to understand how to identify mutations providing a selective disadvantage (hindering a virus's ability to spread) or become neutral (no effect on a virus's ability to spread). R-value could bring some quantification in identifying the spreadability of variants. Hence, I should also analyze how different variants and their R-value correlate to spreadability. In addition, would love to understand how the imdevimab/regen cov (spike protein antibody) reduces the ACE2 bind.

Acknowledgements

I acknowledge my parents, science teacher and mentors, who supported me with clarifications and doing reviews during my project execution. In addition, I would like to thank, my brother Raghuagaan and my sister Nikhitha for keeping me entertained while I was drafting this paper.

References

[1] Starr, T. N., Greaney, A. J., Hilton, S. K., Ellis, D., Crawford, K. H. D., Dingens, A. S., Navarro, M. J., Bowen, J. E., Tortorici, M. A., Walls, A. C., King, N. P., Veesler, D., & Bloom, J. D. (2020). Deep mutational scanning of SARS-COV-2 receptor binding domain reveals constraints on folding and Ace2 binding. Cell, 182(5). https://doi.org/10.1016/j.cell.2020.08.012

[2] Zharko Daniloski, Tristan X Jordan, Juliana K Ilmain, Xinyi Guo, Gira Bhabha, Benjamin R tenOever, Neville E Sanjana (2021) The Spike D614G mutation increases SARS-CoV-2 infection of multiple human cell types eLife 10:e65365. https://doi.org/10.7554/eLife.65365

[3] Guruprasad, L. (2021). Human sars COV -2 spike protein mutations. Proteins: Structure, Function, and Bioinformatics, 89(5), 569–576. https://doi.org/10.1002/prot.26042

[4] Shen, L., Triche, T. J., Bard, J. D., Biegel, J. A., Judkins, A. R., & Gai, X. (2021). Spike protein NTD mutation G142D in SARS-COV-2 delta VOC lineages is associated with frequent back mutations, increased viral loads, and immune evasion. https://doi.org/10.1101/2021.09.12.21263475

[5] Cherian, S.; Potdar, V.; Jadhav, S.; Yadav, P.; Gupta, N.; Das, M.; Rakshit, P.; Singh, S.; Abraham, P.; Panda, S.; et al. SARS-CoV-2 Spike Mutations, L452R, T478K, E484Q and P681R, in the Second Wave of COVID-19 in Maharashtra, India. Microorganisms 2021, 9, 1542. https://doi.org/10.3390/microorganisms9071542

[6] Liu Z, VanBlargan LA, Bloyet L-M, Rothlauf PW, Chen RE, Stumpf S, Zhao H, Errico JM, Theel ES, Liebeskind MJ et al (2021b) Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. Cell Host Microbe 29: 477–488. https://doi.org/10.1016/j.chom.2021.01.014

[7] Starr, T. N. et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. Cell 182, 1295–1310. https://doi.org/10.1016/j.cell.2020.08.012 (2020)

[8] V'kovski, P., Kratzel, A., Steiner, S. et al. Coronavirus biology and replication: implications for SARS-CoV-2. Nat Rev Microbiol 19, 155–170 (2021). https://doi.org/10.1038/s41579-020-00468-6

[9] Harvey, W.T., Carabelli, A.M., Jackson, B. et al. SARS-CoV-2 variants, spike mutations, and immune escape. Nat Rev Microbiol 19, 409–424 (2021). https://doi.org/10.1038/s41579-021-00573-0

[10] Zhang, L., Jackson, C. B., Mou, et al. (2020). SARS-COV-2 spike-protein D614G mutation increases virion spike density and infectivity. Nature Communications, 11(1). https://doi.org/10.1038/s41467-020-19808-4

[11] Lista, M. J., Winstone, H., Wilson, H. D., Dyer, A., Pickering, S., Galao, R. P., De Lorenzo, G., Cowton, V. M., Furnon, W., Suarez, N., Orton, R., Palmarini, M., Patel, A. H., Snell, L., Nebbia, G., Swanson, C., & Neil, S. J. (2021). The P681H mutation in the spike glycoprotein confers type I interferon resistance in the SARS-COV-2 alpha (B.1.1.7) variant. https://doi.org/10.1101/2021.11.09.467693

[12] Ou, J., Lan, W., Wu, X. et al. Tracking SARS-CoV-2 Omicron diverse spike gene mutations identify multiple inter-variant recombination events. Sig Transduct Target Ther 7, 138 (2022). https://doi.org/10.1038/s41392-022-00992-2

[13] SARS-CoV-2 spike S375F mutation characterizes the Omicron BA.1 variant. Izumi Kimura et al, bioRxiv preprint 2022. https://doi.org/10.1101/2022.04.03.486864

[14] Chen, J., Wang, R., Gilby, N. B., & Wei, G. W. (2022). Omicron Variant (B.1.1.529): Infectivity, Vaccine Breakthrough, and Antibody Resistance. Journal of chemical information and modeling, 62(2), 412–422. https://doi.org/10.1021/acs.jcim.1c01451

[15] Wanwisa D et al, SARS-CoV-2 Omicron-B.1.1.529 leads to widespread escape from neutralizing antibody responses. VOLUME 185, ISSUE 3, P467-484. https://doi.org/10.1016/j.cell.2021.12.046

Other Sources

[AA] Biology of SARS-COV-2. HHMI BioInteractive. (2021, November 5). Retrieved November 15, 2021, from https://www.biointeractive.org/classroom-resources/biology-sars-cov-2

[BB] COVID Data Tracker https://covid.cdc.gov/covid-data-tracker/#datatracker-home

[CC] Centers for Disease Control and Prevention. (n.d.). SARS-COV-2 variant classifications and definitions. Centers for Disease Control and Prevention. Retrieved December 5, 2021, from https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html

[DD] UniPro UGENE tool http://ugene.net/

[EE] MFFTT, https://mafft.cbrc.jp/alignment/software/closelyrelatedviralgenomes.html

[FF] Briefings in Bioinformatics, Volume 22, Issue 2, March 2021, Pages 642–663, https://doi.org/10.1093/bib/bbaa232

[GG] SARS-COV-2 mutations complicate COVID-19 pandemic response. Charles River Laboratories. (n.d.). Retrieved November 15, 2021, from https://www.criver.com/eureka/sars-cov-2-mutations-complicate-covid-19-pandemic-response

[HH] SIB Swiss Institute of Bioinformatics. Retrieved November 15, 2021, from https://viralzone.expasy.org/9077

[II] U.S. National Library of Medicine. (n.d.). SARS-COV-2 resources - NCBI. National Center for Biotechnology Information. Retrieved November 15, 2021, from https://www.ncbi.nlm.nih.gov/sars-cov-2/

[JJ] Wikimedia Foundation. (2021, December 4). SARS-COV-2 delta variant. Wikipedia. Retrieved December 5, 2021, from https://en.wikipedia.org/wiki/SARS-CoV-2_Delta_variant