Mapping Immunogenic Regions in SARS-CoV-2 Variants to Understand PCR Assay Design Using Bioinformatics Tools

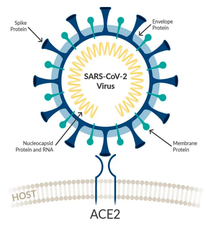
Joanna Sanchez, Christopher Salgado, and Jatniel Morales
Advisor: Vemu Sheela, Ph.D

Abstract

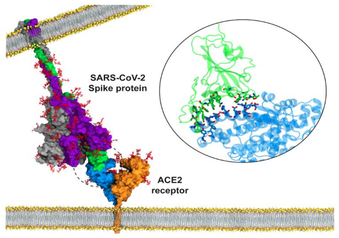
Using the web-based bioinformatics platform Immune Epitope Database (iedb.org) which is funded by NIAID (National Institute of Allergy and Infectious Diseases), we have identified the amino acid residues located on the immunogenic regions of the spike glycoprotein of SARS-CoV-2, the causative agent of COVID-19. The virus can be detected in wastewater since it is excreted in the feces. Waste water monitoring is a cost-effective and non-invasive method for mitigating virus transmission. By identifying the variant of SARS-CoV-2 in each community, it can predict trends and develop early indicators of potential outbreaks. Additionally, by investigating the antigenic determinants of SARS variants in wastewater surveillance using the IEDB, researchers can add surveillance measures to report public health awareness to build community action. The primary objective of this study is to map the immunogenic regions on the spike glycoproteins of the SARS-CoV-2 variants Alpha, Beta, Gamma, Delta, and Omicron using the IEDB which utilizes cutting-edge deep learning algorithms. As a research platform for studying infectious diseases, the IEDB provides published epitope information and prediction tools, including a 3D viewer to analyze immune epitope queries. By detecting SARS-CoV-2 in wastewater and investigating its variants using IEDB epitopes, researchers can predict trends in virus transmission and develop diagnostic kits to monitor the assays in the wastewater.

Introduction

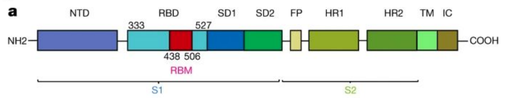
SARS-CoV-2 is a highly infectious virus that causes COVID-19. One of the key components that enable the virus to mutate is the spike glycoprotein, which gives rise to emerging variants. By identifying the variant of SARS-CoV-2 in each community, it is possible to predict trends and develop early indicators of potential outbreaks.Wastewater monitoring is a cost-effective and non-invasive method for detecting the virus, as it can be found in wastewater since it is excreted in feces. Polymerase chain reaction (PCR) assays are a robust way of testing sewage samples that have low concentrations of SARS-CoV-2 genomes. However, as emerging variants continue to mutate and share evolutionary relationships and characteristic mutations, it can be harder to detect specific variants in wastewater. At this moment, PCR faces challenges as it can only target a few mutations and has difficulty detecting samples from a specific variant of interest with high specificity and sensitivity.To address this challenge, the Immune Epitope Database (IEDB), a web-based bioinformatics platform, offers a wide range of epitope analysis and prediction tools. One of the key features of the IEDB platform is its ability to map epitope information and compute immunogenicity for each position in a particular antigen of interest. This information can be used to identify immunogenic hotspots of epitopes, which can provide a framework for researchers t predict the sensitivity and specificity of a PCR assay to detect specific variants of interest. Additionally, by investigating its variants using IEDB epitopes, researchers can predict trends in virus transmission and develop diagnostic kits to monitor the assays in wastewater.



- SARS-CoV-2 virus has Spike-Glycoprotein



- Spike-Glycoprotein binding to host cell receptor ACE2



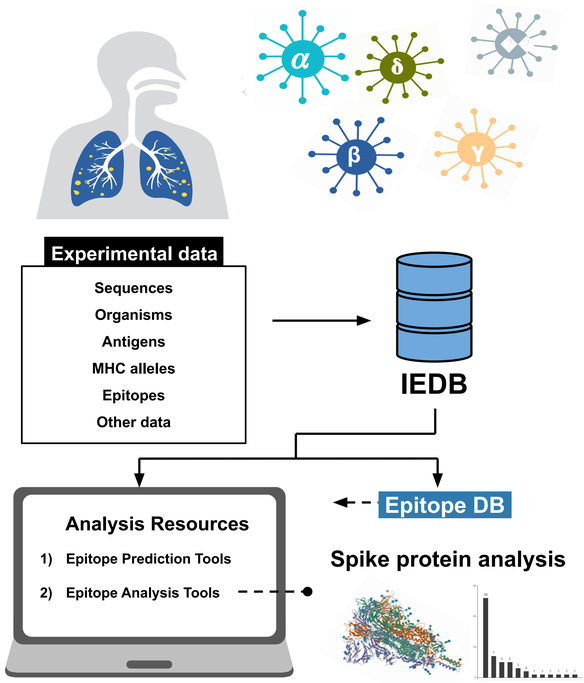
- The linear peptide epitope of SARS-CoV-2 structure
- Receptor binding domain: most immunogenic region

Acknowledgments

We would like to thank the faculty advisor, Sheela Vemu, Ph.D., Associate Professor of Biology at Waubonsee Community College. We would like to express our appreciation for the input from the scientists working on IEDB – Immune Epitope Database, Dr. Mahita Jarjapu, Dr. Marcus Mendes, and Senior Project Manager Nina Blazeska, La Jolla Institute for Allergy and Immunology, La Jolla, CA.

Materials and Methods

The IEDB at www.iedb.org is a powerful research platform for investigating infectious diseases. It provides published epitope information and prediction tools, as well as a robust bioinformatics tool that allows for the visualization of epitopes through an enhanced 3D viewer, which uses NCBI's iCn3D viewer and response frequency graphs. The tool maps and visualizes peptidic epitopes along the length of the protein sequence of a target or reference protein, enabling users to investigate the frequency of immune assays that include positive and/or negative immune responses in different protein regions. Our epitope search began with IEDB's powerful selection tool that allowed us to filter for **B cell assays** only, and we specifically chose **SARS-CoV-2** as our epitope source, with **human** as the host.



Research Goals

By detecting SARS-CoV-2 in wastewater and investigating its variants using IEDB epitopes, researchers can predict trends in virus transmission and develop diagnostic kits to monitor the assays in the wastewater. Our current research goal is **to compare immune evasion of five different variants of concern by analysing IEDB data.**

- Identify conserved epitopes in these different variants; Alpha, Beta, Gamma, Delta, and Omicron.
- This can identify immunogenic hotspots of epitope recognition in a protein as compared to other areas that are not recognized.
- Identify specific mutation for each variant to provide a framework for future PCR assay design by using prediction tools such as IEDB.

References

[1] Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters B. The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res. 2018 Oct 24. doi: 10.1093/nar/gky1006. PMID: 30357391; PMCID: PMC6324067

[2] Motozono, C., Toyoda, M., Zahradnik, J. et al. (2021). SARS-CoV-2 spike L452R variant evades cellular immunity and increases infectivity. Cell Host & Microbe, 29(7), 1124-1136.e11. https://doi.org/10.1016/j.chom.2021.06.006

[3] Taka, E., Yilmaz, S. Z., Golcuk, M. et al. (2021). Critical Interactions Between the SARS-CoV-2 Spike Glycoprotein and the Human ACE2 Receptor. The Journal of Physical Chemistry B 2021 125 (21), 5537-5548
DOI: 10.1021/acs.jpcb.1c02048

[4] Lan, J., Ge, J., Yu, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581, 215–220 (2020). https://doi.org/10.1038/s41586-020-2180-5

[5] "Tools to Study SARS-CoV-2-Host Interactions", accessed Apr 6. 2023, https://www.caymanchem.com/news/tools-to-study-sars-cov-2-host-interactions

Results

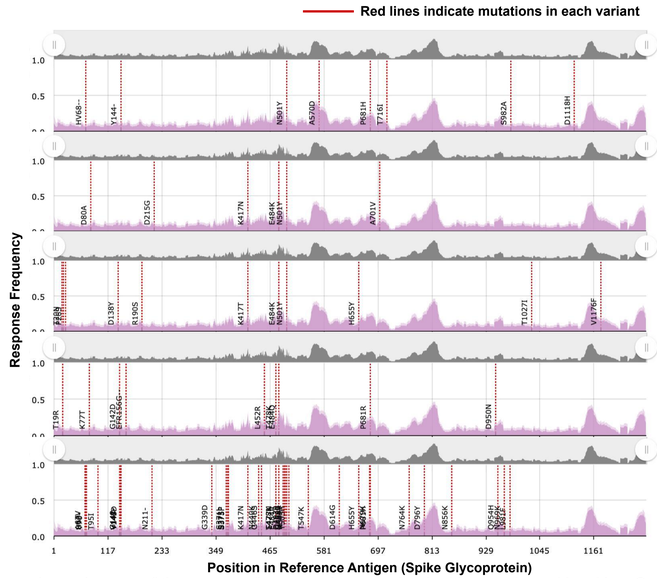
Alpha (B.1.1.7)
Primitive determination: **September 2020**
First origin: **United Kingdom**

Beta (B.1.351)
Primitive determination: **October 2020**
First origin: **South Africa**

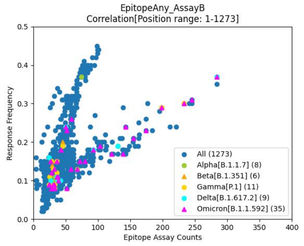
Gamma (P.1)
Primitive determination: **January 2021**
First origin: **Brazil/ Japan**

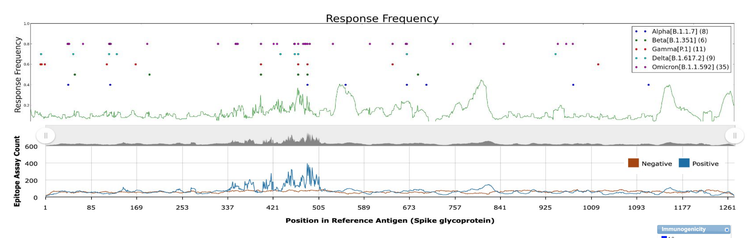
Delta (B.1.617.2)
Primitive determination: **October 2020**
First origin: **India**

Omicron (B.1.1.529)
Primitive determination: **November 2021**
First origin: **South Africa**



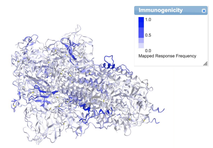
The ImmunomeBrowser graph described above shows the map of the response frequency data associated with an amino acid residue. Our results suggest that certain mutations are conserved among these five variants. The immunogenicity hotspots were found in the residue range of 300 - 550, corresponds to the receptor binding domain. The positive epitope assay counts showed activity, while the response frequency exhibited no change when tested against our reference antigen. Omicron showed the most mutations within the overall range of residues particularly within the hotspot range.





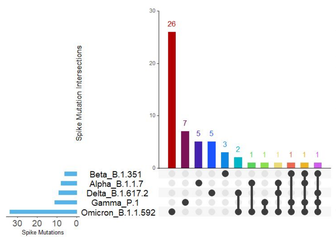
We see two trends:

- Smaller counts with respect to response more frequently responded.
- Larger count with respect to response less frequently responded.
- Among, all five variants omicron has the most immunogenic regions and follows the larger count trend within respect to higher response frequency.



Next Steps: Application in Wastewater-based Epidemiology

Difficulty of PCR assay use in wastewater

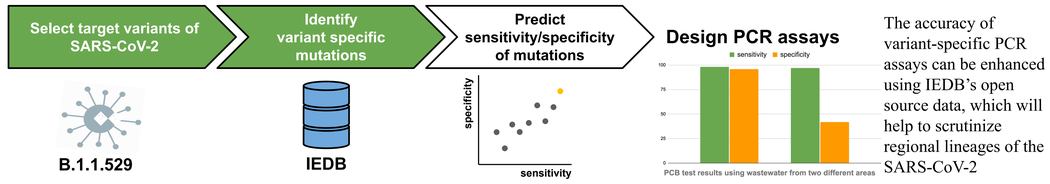


It is well known that PCR is more robust for sewage samples that have low concentrations of SARS-CoV-2 genomes and contain different types of impurities. However, PCR assays typically targets only a 1~3 mutations.

Left figure shows the shared mutations (amino acid residues) on the spike glycoprotein among the various SARS-CoV-2 variants: Beta, Alpha, Delta, Gamma and Omicron. Variants with shared mutation residues are depicted by a linking line between each variant (represented as dots).

- Omicron variant possesses 26 unique mutations on the Spike glycoprotein, and over 30 total Spike mutations
- Amino acid Residues 484 (magenta), 501 (yellow), 417 (orange) and 681 (gold) share a high degree of overlap.

The overlap makes PCR assays have difficulty detecting samples from a specific variant of interest with high specificity and sensitivity as it only target a few mutations.



The accuracy of variant-specific PCR assays can be enhanced using IEDB's open source data, which will help to scrutinize regional lineages of the SARS-CoV-2