A Comparative Study of SARS-CoV-2 Delta and Omicron Variants

Team 02

Department of Systems and Biomedical Engineering

Cairo university

1. Introduction

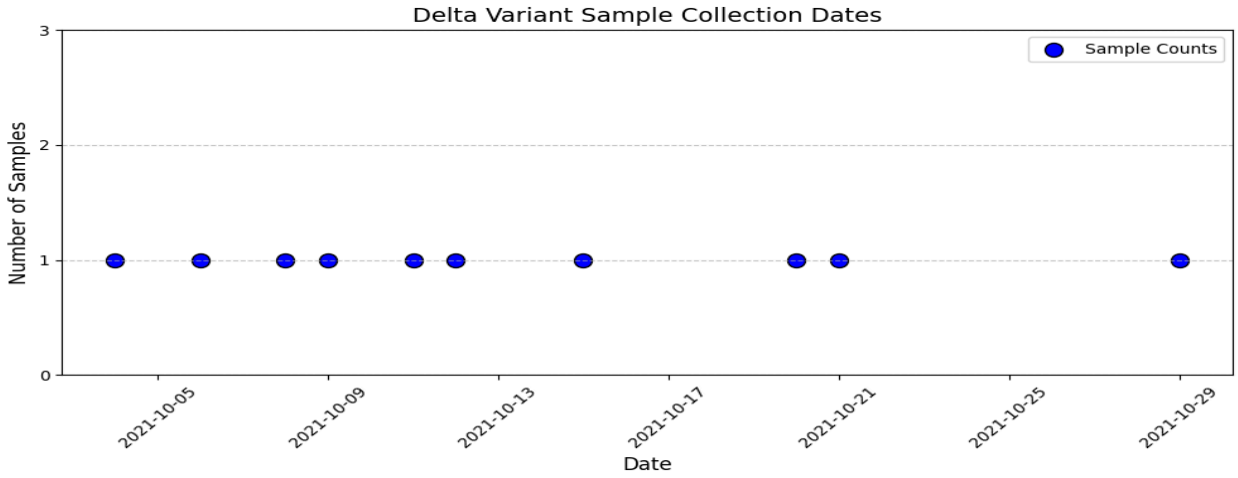
SARS-CoV-2, a single-stranded RNA virus first detected in December 2019 in Wuhan, China, is responsible for the COVID-19 pandemic. By February 2023, it caused over 754 million infections and 6 million deaths globally. In Vietnam, the virus triggered four outbreaks, significantly impacting daily life. The virus exhibits a broad spectrum of clinical manifestations, ranging from asymptomatic cases to severe outcomes like respiratory failure and multiorgan dysfunction. Its high mutation rate has led to diverse variants, influencing transmissibility, disease severity, and vaccine efficacy. Despite prior genome sequencing studies in Vietnam, there is limited research on the correlation between SARS-CoV-2 variants and clinical outcomes. In this study, we focus on investigating the genetic mutations between the Delta variant and the Omicron variant.

2. Methods

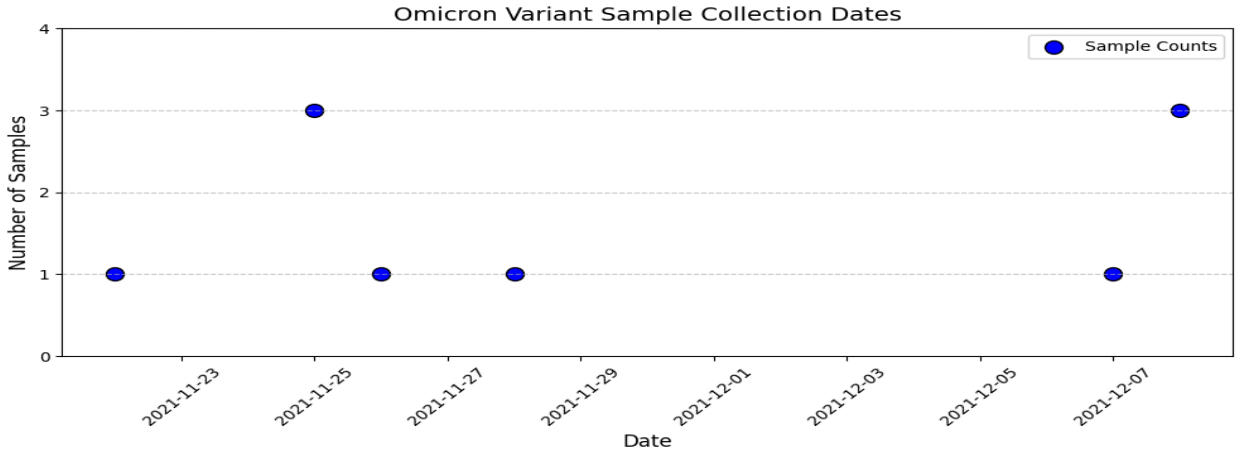
2.1. Data Description

We used a dataset consisting of 30 sequences obtained from 30 different samples from Ghana's population, 15 of which are from the Delta variant of SARS-CoV-2 and the remaining 15 from the Omicron variant. We randomly selected 10 sequences from each category for investigation in our study. The average length of both the Delta and Omicron sequences is 29,601 residues. In our comparison, the Delta variant sequences serve as the reference sequences, and the Omicron sequences serve as the case sequences.

Delta variant samples were collected between 04/10/2021 and 29/10/2021, with no two samples collected on the same day.



Omicron variant samples were collected between 22/11/2021 and 08/12/2021.



2.2 Consensus Sequence Construction

Firstly, we constructed a consensus sequence to serve as a single reference sequence representing the 10 reference sequences of the Delta variant. To construct the consensus sequence, we first aligned the 10 sequences using the "ClustalW" library. Then we constructed the consensus sequence such that its residues were the most common residues at each column in the 10 aligned sequences.

2.3 Phylogenetic Tree Construction

To compare the case sequences with the reference sequences, we first aligned all of the 20 sequences. Then, we constructed a phylogenetic tree using the neighbour joining technique. We used the "Phylo" module from the "Bio Python" library for constructing the tree and "matplotlib" for improved visualization.

Phylogenetic Tree Construction steps

1. **Distance Matrix Calculation**: The pairwise evolutionary distances between sequences were calculated using the identity method.
2. **Tree Construction Method**: The Neighbor-Joining (NJ) algorithm was employed to generate the phylogenetic tree. This method clusters sequences based on their evolutionary distances.

2.4 Percentage of the Chemical Constituents in the two variants

We calculated the percentage of each nucleotide base in the reference sequences as well as in the case sequences. We also calculated their CG content to determine which of the two variants is more resistant to further mutations.

2.5 Identification of regions of mismatch between the two variants

To compare the genetic differences between the Delta variant and the Omicron variant, we identified the regions of consecutive mismatches and individual mismatch residues.
For each column in the alignment, the nucleotide residues from all sequences were extracted and analyzed. A mismatch was identified when the column contained more than one unique nucleotide

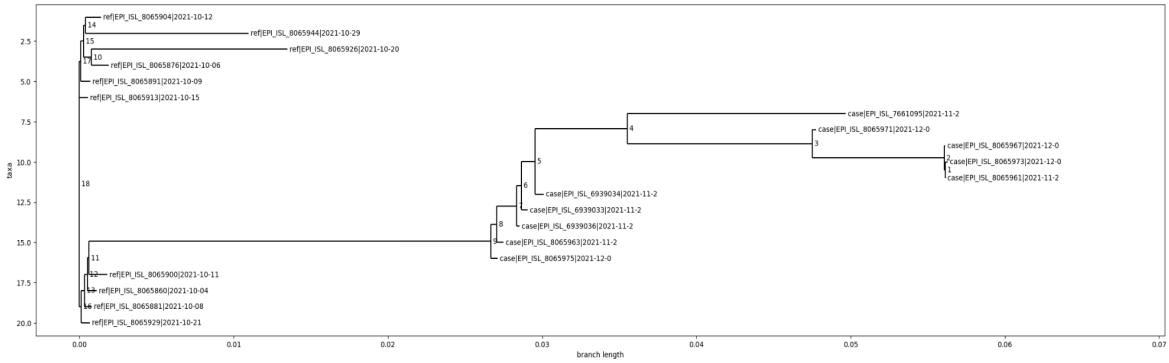
residue (excluding gaps "-"). The locations of individual mismatch residues as well as regions of consecutive mismatches were recorded

We also classified the identified mutations into variant-related mutations and sample-related mutations. Variant-related mutations occur when at least 9 out of the 10 case sequences have similar residues which differ from that of the consensus sequence, indicating that this mutation is most likely variant-dependent. In contrast, sample-related mutations occur when variability exists in multiple case sequences, or when one sequence uniquely differs from the rest.

3. Results and Discussion

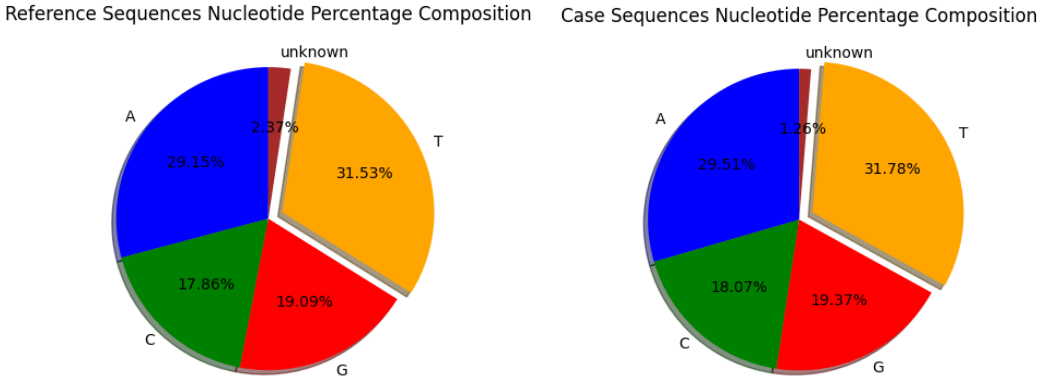
3.1 Phylogenetic tree

The resulting phylogenetic tree showed that the case sequences formed a single cluster, while the reference sequences formed two distinct clusters. The minimum distance detected was 0.000101, observed between the following two case sequences: case|EPI_ISL_8065961|2021-11-2 and case|EPI_ISL_8065973|2021-12-0.



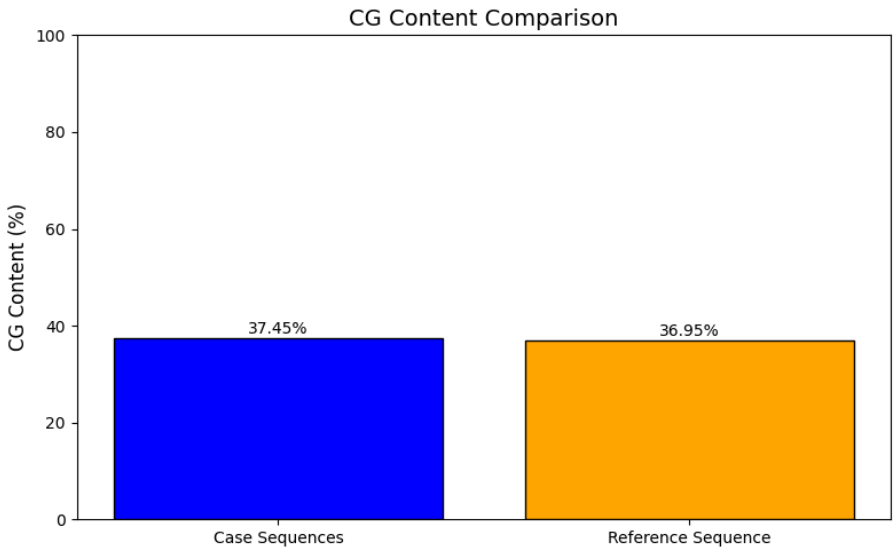
3.2 Percentage of the Chemical Constituents in the two variants

The percentage of each nucleotide in both variants are almost equal.



The CG content of each of the two variants is around 37%, which indicates that the number of triple hydrogen bonds is lower than that of dihydro0

gen bonds.



3.3 Mismatched Regions

The number of columns experiencing mismatches is 1,892, representing 6.35% of the columns. 36.15% of the identified mismatches were classified as variant-related mutations, while the rest were sample-related. The number of individual mismatch residues is 75 and the number of mismatch regions is 22.

4. Conclusion

The comparative analysis of the Delta and Omicron variants of SARS-CoV-2 highlights key genetic differences that contribute to our understanding of their evolution. Our findings provide valuable insights into the extent of mutations that differentiate the two variants and underscore the importance of ongoing genomic investigation to manage emerging SARS-CoV-2 variants effectively.

5. References

- Y. Wang *et al.*, “Human SARS-CoV-2 has evolved to reduce CG dinucleotide in its open reading frames,” *Scientific Reports*, vol. 10, no. 1, Jul. 2020, doi: https://doi.org/10.1038/s41598-020-69342-y.
- L. Van Nam, T. C. Dien, L. V. N. Bang, P. N. Thach, and L. Van Duyet, “Genetic features of SARS-CoV-2 Alpha, Delta, and Omicron variants and their association with the clinical severity of COVID-19 in Vietnam,” *IJID Regions*, vol. 11, p. 100348, Mar. 2024, doi: https://doi.org/10.1016/j.ijregi.2024.03.003.

6. Members contributions

- Hadeer sherif and Mariam Hatem: Multiple sequence alignment, Consensus sequence construction, Phylogenetic tree construction, Percentage of the Chemical Constituents in the two variants, documentation
- Nada Amr and Salema Abdeltawab: Multiple sequence alignment, Data visualization, Mismatch region identification and classification, documentation