**Relation between**

**SARS-Cov-2 and Omicron variant in Botswana**

Baraa Tarek1[⋆](#_bookmark0) Tarek Mohamed2 Mohamed Ayman3 Mohab Hesham4



[iD](https://orcid.org/0000-0000-0000-0000)



[iD](https://orcid.org/0000-0000-0000-0000)



[iD](https://orcid.org/0000-0000-0000-0000)



[iD](https://orcid.org/0000-0000-0000-0000)

1*Section: 1 B.N.:19*

2*Section: 1 B.N.: 45*

3*Section: 2 B.N.:13*

4*Section: 2 B.N.: 34*

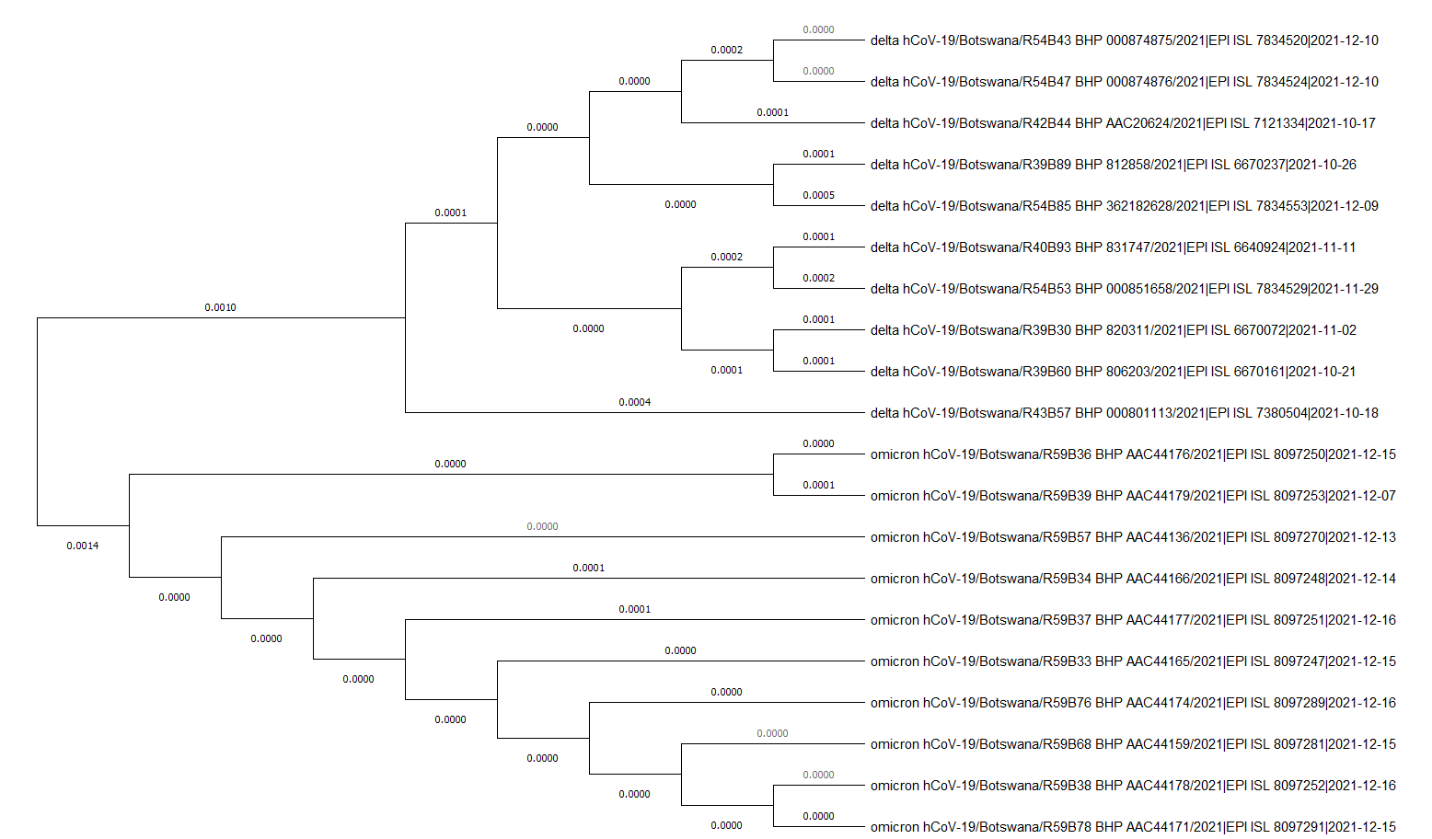
**1 Introduction**

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo.

## Methods

## Phylogenetic Tree

## A phylogenetic tree is a diagram that represents evolutionary relationships among organisms or genes from common ancestor, it helps us to know more about biological diversity and structuring classification , In trees, two genes are more related if they have a more recent common ancestor and less related if they have a less recent common ancestor, neighbor-joining is an algorithm used to reconstruct phylogenetic trees from distance data , it’s used to find neighbors based on distance at each stage of clustering, points with least distances are neighbors , we used the phylogenetic tree in our project to analyze the relation between delta and omicron by the help of an external software “ MEGA “ .



## Consensus Sequence

A consensus sequence is a sequence of [DNA, RNA](https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/dna-rna-hybridization), or protein that represents aligned, related sequences. The consensus sequence of the related sequences can be defined in different ways, but is normally defined by the most common nucleotide(s) or amino acid residue(s) at each position.

It serves as a simplified representation of the viral population. It represents the results of multiple sequence alignments in which related sequences are compared to each other.

In DNA molecules, consensus sequences are useful for representing short elements that are binding sites for other molecules. Such elements may be recognized by, for example, when considering sequence-dependent enzymes such as RNA polymerase. (**RNA polymerase**, is an enzyme that synthesizes RNA from a DNA template. Using the enzyme helicase, RNAP locally opens the double-stranded DNA so that one strand of the exposed nucleotides can be used as a template for the synthesis of RNA, a process called transcription.)

In proteins, consensus sequences may represent entire protein molecules or short fragments of them that correspond to conserved regions of importance for structure and function.

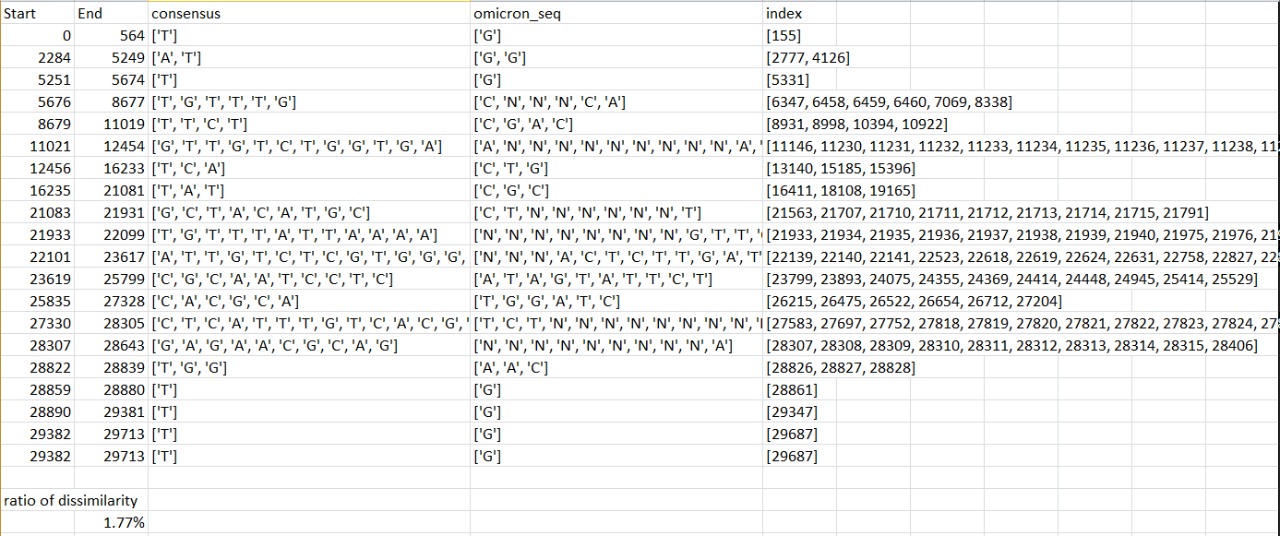
A protein binding site, represented by a consensus sequence, may be a short sequence of nucleotides which is found several times in the genome and is thought to play the same role in its different locations. For example, many transcription factors recognize particular patterns in the promoters of the genes they regulate. In the same way, restriction enzymes usually have palindormic consensus sequences, usually corresponding to the site where they cut the DNA.

Thus consensus sequence is defined as the idealized sequence that represents the predominant base at each position. All the actual examples shouldn't differ from the consensus by more than a few substitutions.

Any mutation allowing a mutated nucleotide in the core promoter sequence to look more like the consensus sequence is known as an **up mutation**. This kind of mutation will generally make the promoter stronger, and thus the RNA polymerase forms a tighter bind to the DNA it wishes to transcribe and transcription is up-regulated. On the contrary, mutations that destroy conserved nucleotides in the consensus sequence are known as **down mutations**. These types of mutations down-regulate transcription since RNA polymerase can no longer bind as tightly to the core promoter sequence.

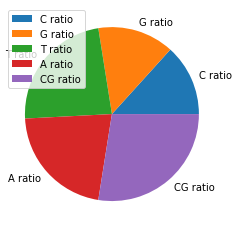
Samples from Code output:



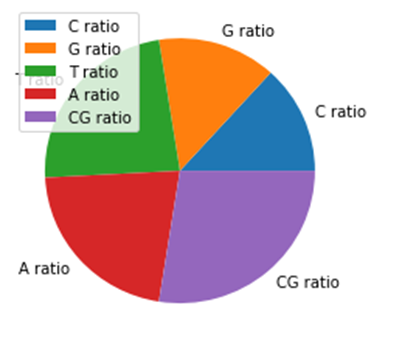


## 2.3 Constituent Percentages

Upon calculating the chemical constituents of variants we used a code that iterates over each sequence of the Omicron and calculates the sum of the calculated constituent in it. Then we calculated its percentage from the whole sequence. After doing so for all of the 10 sequences of the variant we calculated the average percentage of each constituent in Omicron SARS-Cov-2. The Percentages were calculated for



(Delta of hCoV-19)

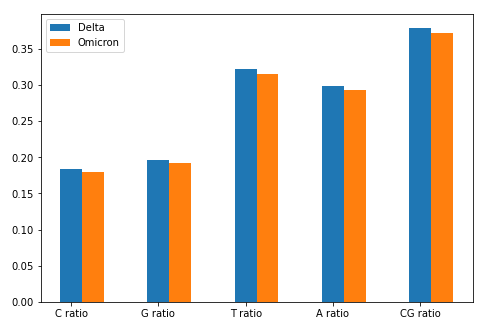


(Omicron of hCoV-19

delta SARS-Cov-2 using its consensus sequence only not iterating over the 10 sequences. First we used the code that returns the consensus then calculated the percentages in it. The previous steps were done using libraries in python such as bioPython, re and numpy. Then we used some data visualization tools such as pie and bar charts in matplotlib to make the comparison process a little bit easier.

The Outputs of this method yielded that the average ratio of:

* C in delta hCoV-19 is 0.1832 while in its omicron variant its 0.1793
* G in delta hCoV-19 is 0.1961 while in its omicron variant its 0.1924
* T in delta hCoV-19 is 0.3216 while in its omicron variant its 0.3154
* A in delta hCoV-19 is 0.2989 while in its omicron variant its 0.2932
* CG in delta hCoV-19 is 0.3794 while in its omicron variant its 0.3718



(Bar Chart Comparison)

1. **Results and Conclusions**

As we can see , delta sequence samples are very similar in their constructions and the distances between the neighbors is very small , but in omicron samples , they are approximately typical , as we see many branches are of length of zero which means that the neighbors are typical in their structure

We can also see that the mutations between delta and omicron isn’t big and they are close to each other’s in structure because the branches length is relatively small, which is a good indication for us as we can understand the structure and behavior of omicron based on our delta studies and also our vaccines and drugs developed for delta may deal with omicron also.

1. Ratio of dissimilarity is around 1.77% which means that the omicron Sequences are very similar to consensus sequence.
2. There are a lot of undefined nucleotides in the omicron sequences so there are some errors in our statistics.
3. The most mutated nucleotide is T and most of the time T nucleotide is mutated to G or undefined nucleotide in the omicron sequences.
4. From index 0 to 6000 and the last 1000 nucleotides the sequences are mostly typical and any mutated nucleotide in consensus is mutated to G.
5. Lot of mutations are to undefined nucleotides in omicrons, So we can’t make sure if it’s dissimilar or not.
6. There are lots of dissimilarities between each sequence of the 10 sequences of omicron.
7. The CG ratios fell with percentage of 0.76% in the Omicron variant resulting a dna with a slightly lower stability than the delta hCoV-19.
8. Most of the Chemical constituent’s percentages in the Omicron variant decreased by a very small amount which may be due to the presence of lots of NaN values in the Omicron sequences
9. Mathematically there’s a slight change in the percentages but biologically we can deduce that the 2 variants are very chemically close to each other as the change is insignificant.
10. **References**
11. <https://en.wikipedia.org/wiki/Consensus_sequence>
12. <https://en.wikipedia.org/wiki/RNA_polymerase>
13. <https://www.sciencedirect.com/topics/medicine-and-dentistry/consensus-sequence>