

SARS-Cov-2 and its variant Omicron Project

Presented by:

Ahmed Mahmoud Hussein (1180364)

Mostafa Neddal Mostafa (1180045)

Mostafa Omar Mostafa (1180342)

Mohamed Fathy Mohamed (1180083)

Presented on: 10-01-2022

Table of Contents

Introduction	3
Methods	3
Ways of comparison & Results	4
Conclusion	5
References.....	8

Introduction

In this project we made a comparative study between SARS-Cov-2(Covid 19), and its variant which is called OMICRON in this study we had a total of 20 sequences we downloaded them from GISAID which is a database that contains a lot of sequences after we used website (COV-Lineages.org) to know which countries have that Variant, then we found out that rank 1 is America (USA) which have a percentage 38% of all infected worldwide followed by India and Turkey by percentage 10% each , we chose Turkey in our project

Methods

Let us go through the details of the study

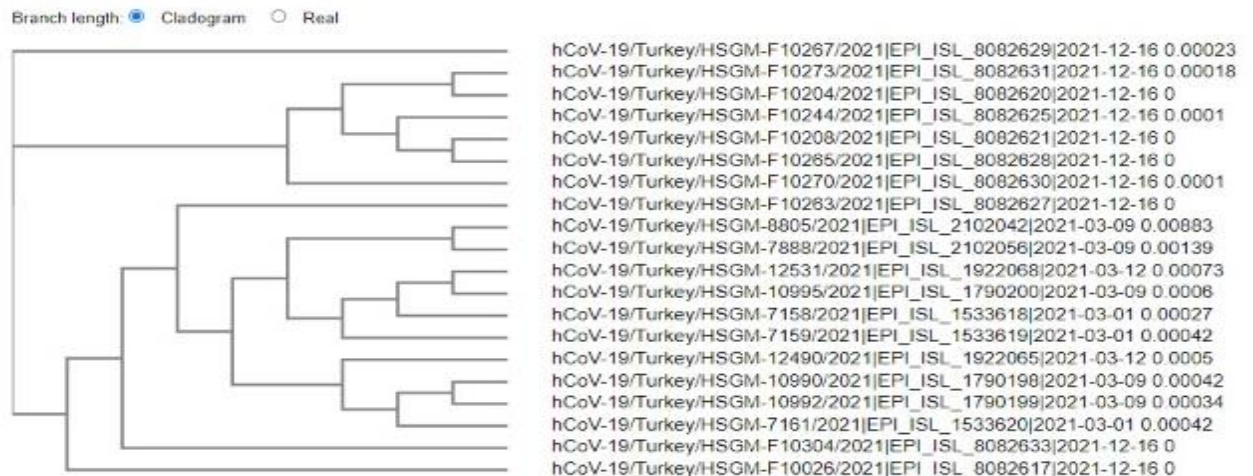
First we used GISAID as we talked before then we used Clustal Omega which helped us in sequence alignment step then it automatically generated the phylogenetic with the alignment result and we will explain it further

Then after getting all the Martials we need we chose python to work with as it very simple to deal with, as we used it to get the CG percentage of the sequences we worked on and also we got from it the Reference sequence that we will normalize everything to it and we checked that it is valid to be used from website (jalview.org)

Then we had to find the dissimilar region so we made multiple sequence alignment between the reference sequence (we got before) and the 10 Sequences of the Omicron variant

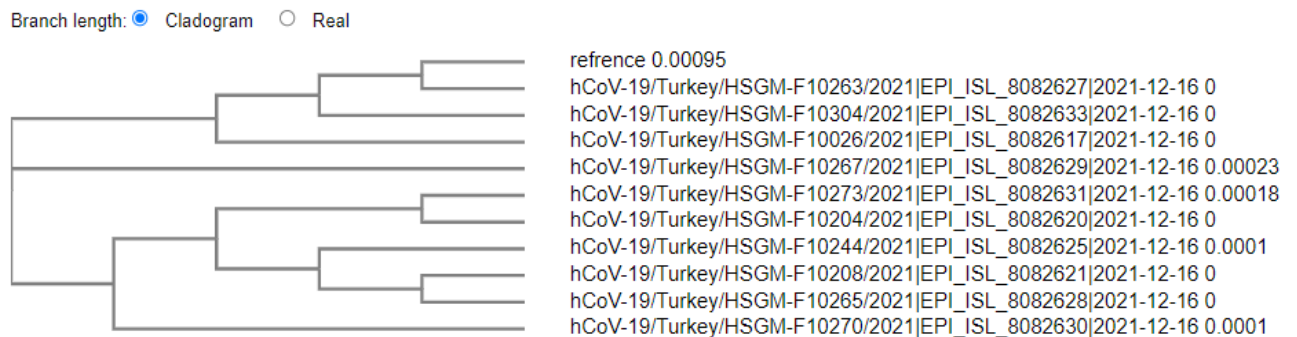
Ways of comparison & Results

1. Phylogenetic tree:



This figure contains the 20 sequences of both Covid19 and its Variant Omicron

For simplicity



We made the reference sequence which contains the sequences of the Covid 19 while the rest is the sequences of the variant Omicron

The phylogenetic Tree is divided into 3 Groups

- 1- Reference sequence and other sequences similar to it
- 2- HSGM-F10267
- 3- the rest of the sequences in tree

That means that the middle branch doesn't has any similarities

Conclusion from Phylogenetic tree

From the phylogenetic Tree we found that the Reference sequence is similar to the sequences with code

- 1- HSGM-F10263
- 2- HSGM-F10304
- 3- HSGM-F10026

The sequences

- 1- HSGM-F10273
- 2- HSGM-F10204
- 3- HSGM-F10244
- 4- HSGM-F10208
- 5- HSGM-F10265

Are all similar to the sequence **HSGM-F10270**

- The reference sequence is similar to **HSGM-F10263** and both are similar to **HSGM-F10304**
- The three sequences mentioned above are similar to **HSGM-F10026**
- **HSGM-F10267** isn't similar to any sequences
- **HSGM-F10273** and **HSGM-F10204** are similar to each other
- **HSGM-F10208** and **HSGM-F10265** are similar to each other and both of them are similar to **HSGM-F10244**

2. Average percentage of the chemical constituents

Reference sequences								
Name	Sequence	Length	A Content	C Content	G Content	T Content	CG Content	
hCoV-19/Turkey/HSGM-7158/2021 EPI_ISL_1533618 2021-03-01	ATTAAAG	29903	29.94348	18.34599	19.60004	32.11049	37.94602548	
hCoV-19/Turkey/HSGM-7159/2021 EPI_ISL_1533619 2021-03-01	ATTAAAG	29903	29.95017	18.34264	19.5967	32.11049	37.93933719	
hCoV-19/Turkey/HSGM-7161/2021 EPI_ISL_1533620 2021-03-01	ATTAAAG	29903	29.9368	18.33595	19.60004	32.12721	37.93599304	
hCoV-19/Turkey/HSGM-10990/2021 EPI_ISL_1790198 2021-03-09	ATTAAAG	29903	29.94348	18.32592	19.5967	32.1339	37.92261646	
hCoV-19/Turkey/HSGM-10992/2021 EPI_ISL_1790199 2021-03-09	ATTAAAG	29903	29.9368	18.34599	19.60673	32.11049	37.95271377	
hCoV-19/Turkey/HSGM-10995/2021 EPI_ISL_1790200 2021-03-09	ATTAAAG	29903	29.95017	18.3393	19.58666	32.12387	37.92596061	
hCoV-19/Turkey/HSGM-12490/2021 EPI_ISL_1922065 2021-03-12	ATTAAAG	29903	29.94014	18.3393	19.60338	32.11718	37.94268134	
hCoV-19/Turkey/HSGM-12531/2021 EPI_ISL_1922068 2021-03-12	ATTAAAG	29903	29.94683	18.32258	19.57663	32.15396	37.89920744	
hCoV-19/Turkey/HSGM-8805/2021 EPI_ISL_2102042 2021-03-09	ATTAAAG	29903	30.10066	18.26907	19.47965	32.15062	37.74872086	
hCoV-19/Turkey/HSGM-7888/2021 EPI_ISL_2102056 2021-03-09	ATTAAAG	29903	30.00702	18.34933	19.57663	32.06702	37.92596061	
Average percentage of the chemical constituents:			29.96556	18.33161	19.58232	32.12052	37.91392168	
Case sequences								
Name	Sequence	Length	A Content	C Content	G Content	T Content	CG Content	
hCoV-19/Turkey/HSGM-F10026/2021 EPI_ISL_8082617 2021-12-16	ACCAACC	29800	29.88926	18.33221	19.62752	32.13423	37.95973154	
hCoV-19/Turkey/HSGM-F10204/2021 EPI_ISL_8082620 2021-12-16	ACCAACC	29800	29.88591	18.32215	19.62752	32.14765	37.94966443	
hCoV-19/Turkey/HSGM-F10208/2021 EPI_ISL_8082621 2021-12-16	ACCAACC	29800	29.88926	18.33221	19.62081	32.14094	37.95302013	
hCoV-19/Turkey/HSGM-F10244/2021 EPI_ISL_8082625 2021-12-16	ACCAACC	29800	29.88591	18.33557	19.62416	32.13758	37.95973154	
hCoV-19/Turkey/HSGM-F10263/2021 EPI_ISL_8082627 2021-12-16	ACCAACC	29800	29.87919	18.3255	19.63087	32.14765	37.95637584	
hCoV-19/Turkey/HSGM-F10265/2021 EPI_ISL_8082628 2021-12-16	ACCAACC	29800	29.88591	18.33893	19.62081	32.13758	37.95973154	
hCoV-19/Turkey/HSGM-F10267/2021 EPI_ISL_8082629 2021-12-16	ACCAACC	29800	29.87919	18.32215	19.63087	32.15101	37.95302013	
hCoV-19/Turkey/HSGM-F10270/2021 EPI_ISL_8082630 2021-12-16	ACCAACC	29800	29.87919	18.32886	19.63087	32.1443	37.95973154	
hCoV-19/Turkey/HSGM-F10273/2021 EPI_ISL_8082631 2021-12-16	ACCAACC	29800	29.87919	18.33221	19.62752	32.1443	37.95973154	
hCoV-19/Turkey/HSGM-F10304/2021 EPI_ISL_8082633 2021-12-16	ACCAACC	29800	29.88926	18.32215	19.62416	32.14765	37.94630872	
Average percentage of the chemical constituents:			29.88423	18.32919	19.62651	32.14329	37.9557047	

Conclusion from Chemical Constituents

From the results, we see that all numbers of chemical constituents are close to each other however, after applying more analysis we concluded that the variance of CG content in covid-19 reference sequences (0.003596553) is higher than the variance of CG content in Omicron case sequences (0.0000245234) and in addition, the average of CG content in the covid-19 sequences is less than the average of CG content in the Omicron case sequences, Therefore the covid 19 sequences is less stable than the Omicron sequences as a result the chance of mutation in covid 19 reference sequences is higher than the Omicron case sequences.

3. Dissimilar Regions/Columns between the alignment of the case sequences and the consensus sequence

[[0, 29], [202], [2831], [3036], [5385], [6512, 6514], [8392], [10028], [10448], [11282, 11290], [11536], [12640], [13194], [15239], [18162], [21007], [21761], [21764, 21769], [21845], [21986, 21994], [22577], [22881], [22897], [22991], [22994], [23012], [23039], [23047], [23062], [23074], [23201], [23524], [23598], [23603], [23853], [23947], [24129], [24423], [24468], [24502], [24871], [24999], [25155], [25583], [26269], [26529], [26576], [26708], [27258], [27293], [27806], [28270], [28310], [28361, 28369], [28880, 28882], [29400], [29415, 29416], [29418], [29420], [29866, 29902]]

ATTAAAGGTTTATACCTTCCCAGGTAACA	Severe acute respiratory syndrome coronavirus 2
C	
A	
T	
T	
GT	
G	
C	
C	
GTTTGTCT	
A	
C	
T	
C	
A	
C	
C	
TACAT	
C	
GTGTTTAT	
G	
T	
G	
G	
C	

A	
A	
G	
A	
T	
C	
C	
T	
C	
C	
G	
C	
A	
T	
C	
G	
C	
C	
C	
C	
A	
C	
G	
A	
C	
C	
A	
C	
GAGAACGC	
NN	
T	
C	
A	
C	
TGACAAAAAAAAAAAAAAAAAAAAAAAAAAAA	

Conclusion from Dissimilar Regions

By analyzing the dissimilar regions through blast, we found that the severe acute respiratory syndrome coronavirus 2 is found in the consensus sequence but not in the case sequences.

References

- ✓ Biopython website <https://biopython.org/>
- ✓ <https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Get&RID=XK0X47MT016>
- ✓ GISAID
- ✓ Clustal Omega
- ✓ jalview.org
- ✓ ebi.ac.uk
- ✓ COV-Lineages.org