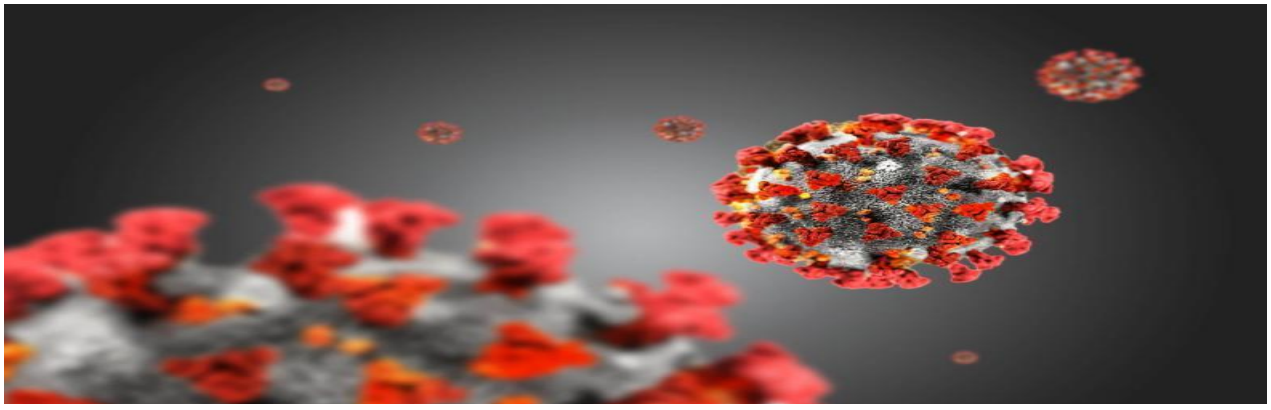
BMD301 Project's Report

Nile University

[COVID-19 WAVES]



Team Members:

Shrouk Hesham	*19106271*
Hussin Fekry	*19105777*
Omnia Salah	*19106208*
Manar Mohamed	*19100552*
Seifeldin Mohamed	*19105145*
Abdelrahaman Mahmoud	*19104609*

Under Supervision:

Dr. Mohamed Elhadidi & Eng. Manar Rashed & Eng. Eman Mohamed

Abstract

Coronavirus disease (COVID-19) is a newly found coronavirus that causes an infectious disease. When there is a sudden rise in cases, it is called an epidemic. Recently, a number of next-generation sequencing (NGS)-based methods have been effective in tracing the origins and fully understanding the evolution of infectious agents, looking into the spread and transmission chains of outbreaks, facilitating the development of efficient and quick molecular diagnostic tests, and advancing the search for treatments and vaccines. The invention of effective diagnostic molecular tests as well as the formation of successful measures and strategies to mitigate the spread of the pandemic were dependent upon the development of effective and quick sequencing methods to reconstruct the genomic sequence of SARS-CoV-2, the pathogenetic agent of COVID-19. In our project, significant differences between pandemic phases or waves are discussed by downloading COVID-19 data from NCBI database. Performing multiple sequence alignment for our COVID-19 samples. In addition, generating a phylogenetic tree and visualizing it. Furthermore, Using BLAST to align COVID-19 spike protein with other viral spike proteins. Finally, by choosing the most homologous protein and annotating it using NCBI and Uniprot. Therefore, we can generate COVID-19 sequences at different time points, which can help us later in understanding the evolution of infectious diseases.

Table of Contents

1. Introduction

In December 2019 in China Wuhan city, COVID-19 appeared, then the disease spread until it became a pandemic, and many countries went through different waves of COVID-19, and each wave is slightly different in its symptoms a cluster of patients with pneumonia of unknown cause was linked to a seafood wholesale market in Wuhan, China. Globally, there have been more than 177 million cases, and the number likely to decline is below 500,000 people per day. The absolute number of passings surpasses 3.8 million individuals; notwithstanding, everyday passings have diminished to under 10,000 individuals each day. The vast majority tainted with the Coronavirus infection had encountered gentle to-direct respiratory ailment and recuperated without extraordinary treatment.

More seasoned individuals with hidden clinical issues like cardiovascular infection, diabetes, ongoing respiratory sickness, and disease are bound to foster difficult ailments. The Coronavirus infection spreads fundamentally through drops of spit or release from the nose when a tainted individual hacks or wheezes, so it means a lot to rehearse respiratory behavior. Numerous nations have encountered different floods of Covid episodes. During the 2020 pandemic, observational information show that attributes changed between waves. Propels in Preventive Medication Volume 2021, In correlation with the subsequent wave, the extent of neighborhood bunches (24.8% versus 45.7%) was lower in the third wave, and individual contact transmission (38.5% versus 25.9%) and obscure courses of transmission (23.5% versus 20.8%) were higher. Thus, numerous states and wellbeing specialists, including the WHO, have been effectively teaching individuals to go to preventive lengths to lessen the spread of the infection, including lockdown measures. Waves in Thailand have been followed to super-spreading occasions at diversion foundations, bars, bars, Karaoke lounges, and a few sorts of betting scenes in various nation locales. These occasions prompted a development of the spread of Coronavirus to numerous areas since the gamble areas were locales that pulled in swarming, broadened associations, and high turnover. An episode in Walk 2020 was related with participants at an enclosing arena Bangkok who spread the infection to different regions as they voyaged home or on business. One of the elements behind the flare-up was the thick day to day

environments of the transients in the encompassing local area and the absence of individual safety measures to forestall its spread. This study thought about the qualities of various waves during the Coronavirus pandemic in Thailand, where more than 150,000 Coronavirus cases have been affirmed. Albeit the day-to-day number of contaminations locally has settled at around 2,000 every day, privately communicated cases have been affirmed in many territories. The spread of Coronavirus was a bunch occasion conveyed over various areas. This occasion dispersed Coronavirus to countless territories. In this task we will apply various groupings between the various rushes of Coronavirus, what's more, we will apply pairwise alignment between Coronavirus spike protein with other viral spike proteins.

Problem Definition:

Spread of COVID-19 in China Wuhan city turned into a pandemic and millions of individuals got contaminated by it, and numerous nations went through various rushes of Coronavirus, and each wave is somewhat disparate in its side effects. The illness is an exceptionally infectious respiratory sickness brought about by the SARS-CoV-2 infection. SARS-CoV-2 is remembered to spread from one individual to another through drops delivered when a tainted individual hacks, sniffles, or talks. It might likewise be spread by contacting a surface with the infection on it and afterward contacting one's mouth, nose, or eyes. The most widely recognized signs and side effects of Coronavirus are fever, hack, and inconvenience relaxing. Exhaustion, muscle torment, chills, cerebral pain, sore throat, runny nose, sickness or regurgitating, looseness of the bowels, and a deficiency of taste or smell may likewise happen. The signs and side effects might be gentle or extreme and typically seem 2 to 14 days after openness to the SARS-CoV-2 infection. Certain individuals might not have any side effects however are yet ready to spread the infection. The vast majority with Coronavirus recuperates without requiring extraordinary treatment. However, others are at higher gamble of difficult ailment. Our Exploration is finished to treat Coronavirus various waves and forestall contamination with SARS-CoV-2.

Purpose:

The main purpose of working on such a project is that the analysis of these waves can be useful to monitor data; for example, to identify the rise in future waves in the country and expect what will happen within these waves and how it could be effective and its size, through the search, we have defined some steps to reach the aimed results. Firstly, we have to use the NCBI database to download COVID-19 sequences (at least twenty samples) at different time points. Perform multiple sequence alignment for our samples. Generate a phylogenetic tree and visualize it. Use BLAST to align COVID-19 spike protein with other viral spike proteins. Choose the most homologous protein and annotate it using NCBI or any preferred database. All analysis steps must be delivered in bash script. That will help the ministry of health in each country to know the right or the correct way to protect their societies.

2. Methodology

In our methods, we used sequencing technology in order to make it easy to generate COVID-19 sequences and compare them at different time, so we made our methods in 6 steps:

2.1. Study Design: In the USA, we directed a review partner investigation of all Coronavirus test screens and cases. Between January 2020 and May 2021, information was gathered (17 months) by involving NCBI data set for Coronavirus to download Coronavirus successions at various time focuses. All through this study we've attempted to sort out to look at the succession of Coronavirus infections inside various time point, to investigate these waves.

2.2. Data Gathering: COVID-19 data collected in USA, between January 2020 and May 2021, including number of test screens, number of confirmed cases and deaths, and sociodemographic variables such as gender, age, nationality, and risk factor sources population. We got to use NCBI Covid-19 database within viral samples instead of using GISAID database because of the late permission, so we have extracted our reported data from NCBI with different Lineage and made sure that all of them their cross- references is Bio Samples and their location in the same country

but with different time in the timeline. We have downloaded +20 samples till now with their fasta format then converted it into notepad to ease the usage.

2.3. Study Population: All recorded ages, genders, nationalities, and at-risk source populations were obtained from the COVID-19 NCBI database. Cases in which such information was missing were excluded. Firstly, we've prepared all fasta sequence in one file, wrote bash script which it takes from user file name and run them on the muscle alignment, and at the end we constructed the phylogenies tree. Without any coding from the user.

2.4. Static analysis Bash scripting tool: By using bash scripting tool, we performed multiple sequence alignment for our retrieved samples. Then, we generated a phylogenetic tree and visualized it. Prepare the sequence in one file to apply the multiple sequence alignment to compare all the RNAs in covid-19 virus through the different waves and generate the phylogenetic tree to identify the development and the division of the virus with its new versions within the waves and detect the time that took for these changes.

2.5. Use BLAST aligning tool: To align COVID-19 spike protein with other viral spike proteins. We worked on the spike protein that we have with the database to find the most homologs spike and get its information. We created a for loop to make a pairwise alignment along the database we use through using **Blast**.

2.6. Evaluation phase: Finally choosing the most homologous protein and annotating it using Uniprot and NCBI or any preferred database. After we divided each blast file separately against the database, we were able to identify the gene that is most likely to result in recombination by performing the significant homologous search step. This step involves the alignment of two homologous sequences and the identification of the homologous site. So, we have chosen our most homologous gene based on the criteria of E-value, bit score, and the best coverage of each file.

3. Results& Discussion

Results so far, through working on this project, we have used the NGS sequencing made it easy to generate COVID-19 sequences and compare them at different time points. Using what we have learn in the course to be able to analyze COVID-19 data at different waves. After we extracted the data from **"NCBI Covid-19"** We have chosen to work on the timeline of waves for the same country **"The United States"** at different times using "20 samples" with different Lineage with **"fasta format"**, then performing multiple sequence alignment for our 20 samples using **"Muscle Alignment"**. We have get the results:

Fig (1 &2) shows: The Multiple sequence alignments for our NCBI 20 samples

Furthermore, generating a **"Phylogenetic Tree"** and visualizing.

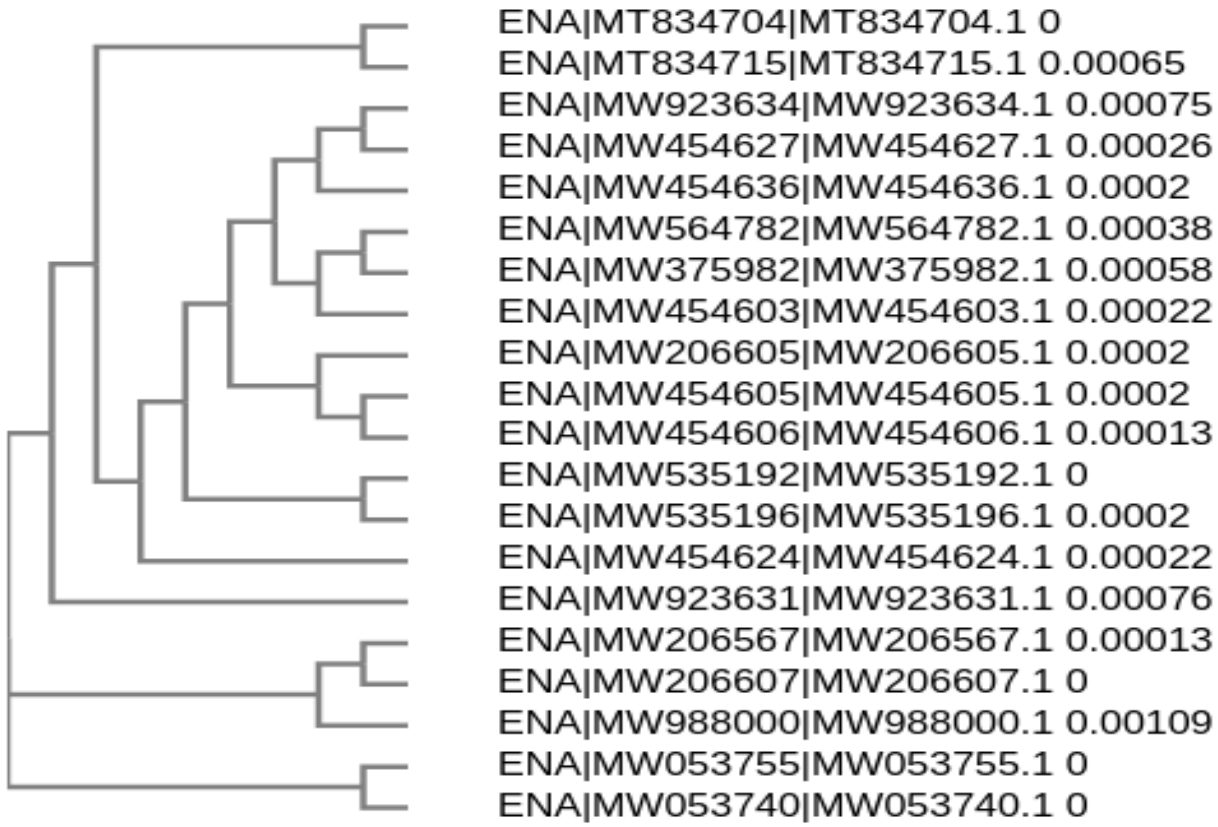


Fig (3 &4) shows: Our **Phylogenetic Tree** visualized for our NCBI 20 samples

Moreover, in order to be able to compare BLAST results done by using different matrices and gap penalties, we used Bit Score (S'). As Bit Score (S') considered a more reliable principle for assuming homology. An average bit score of 50 indicates that a protein is almost always significant.

In addition, we used E-value for sequence matches to the query that are related to chance rather than homology are more likely to be found in larger databases. For homology matches, blast hits with an E-value less than 0.01 can still considered good hits. E-values below 10 will include hits that cannot be considered relevant but may provide a suggestion to possible relationships.

After we divided each blast file separately against the database, we were able to identify the gene that is most likely to result in recombination by performing the significant homologous search step. This step involves the alignment of two homologous sequences and the identification of the homologous site. So, we have chosen our most homologous gene based on the criteria of E-value, bit score, and the best coverage of each file, **we have found the following:**

Best Files Results	**Sequences producing significant alignment**	Bit Score(S')	E-Value
File 3	Four KT182953.1 Middle East respiratory syndrome coronavirus isol	***45.4***	***1e-05***
	two OK642747.1 Porcine epidemic diarrhea virus isolate XJ1904-34	***45.4***	***1e-05***
File 7	four KT182953.1 Middle East respiratory syndrome coronavirus isol	***63.9***	***4e-12***
	two OK642747.1 Porcine epidemic diarrhea virus isolate XJ1904-34	***39.9***	***6e-05***
File 11	four KT182953.1 Middle East respiratory syndrome coronavirus isol	***63.9***	***4e-12***
	two OK642747.1 Porcine epidemic diarrhea virus isolate XJ1904-34	***45.4***	***1e-06***
File 13	four KT182953.1 Middle East respiratory syndrome coronavirus isol	***63.9***	***4e-12***
	two OK642747.1 Porcine epidemic diarrhea virus isolate XJ1904-34	***41.7***	***2e-05***
File 19	two OK642747.1 Porcine epidemic diarrhea virus isolate XJ1904-34	***45.4***	***1e-07***

We have found that files (3, 7, 11 , 13 , 19) gives us best Bit score and E-value

After we have chosen the most homologous gene based on the criteria upon E-value, bit score, coverage. Next, we searched for the gene in other databases and mentioned its cellular and molecular function and how it affects the disease.

So, we have chosen our best sequences which is:

- **KT182953.1 Middle East respiratory syndrome coronavirus**

After we used Uniport to mention in detail the function of the protein, gene ontology, cellular and molecular function and how it affects the disease, we have found that:

✓ **Molecular Function**:

- lipid binding
- cyclase activity
- oxidoreductase activity
- transferase activity
- hydrolase activity
- lipid binding
- cyclase activity
- oxidoreductase activity
- transferase activity
- hydrolase activity
- Others…

✓ **Cellular Function**

- cell wall
- Nucleus
- nuclear envelope
- Mitochondrion

✓ **Cellular Component**

- cell wall
- Nucleus
- nuclear envelope
- Mitochondrion

- **How a Middle East respiratory syndrome coronavirus affects the Coronavirus Disease?**

Middle East respiratory syndrome (MERS) is a viral respiratory contamination brought about by Middle East respiratory syndrome disorder related Covid (MERS-CoV). Side effects might go from none, to gentle, to severe. Typical side effects incorporate fever, hack, looseness of the bowels, and windedness. The illness is normally more serious in those with other medical conditions.

Over 2,500 cases have been accounted for as of January 2021, remembering 45 cases for the year 2020. Around 35% of the people who are determined to have the sickness die from it.

MERS-CoV is a Covid accepted to be initially from bats. Nonetheless, people are normally contaminated by camels, either during direct contact or in a roundabout way. Spread between people regularly requires close contact with a contaminated individual. Its spread is unprecedented beyond hospitals. Thus, its gamble to the worldwide populace is presently considered to be genuinely low. Analysis is by rRT-PCR testing of blood and respiratory examples.

Past disease with MERS can present cross-responsive insusceptibility to SARS-CoV-2 and give halfway security against Coronavirus In any case, co-disease with SARS-CoV-2 and MERS is conceivable and could prompt a recombination occasion.

The Middle East respiratory syndrome (MERS) CoV has impacted more than 1,000 patients with more than 35% casualty since its development in 2012. All essential instances of MERS are epidemiologically connected to the Center East. A portion of these patients had reached camels which shed infection as well as had positive serology. Most optional cases relate to medical services-related groups. The illness is particularly serious in older men with comorbidities. Clinical seriousness might relate to MERS-CoV's capacity to contaminate an expansive scope of cells with DPP4 articulation, dodge the host intrinsic invulnerable reaction, and prompt cytokine dysregulation. Switch recording PCR on respiratory, as well as extrapulmonary examples, quickly lays out determination. Steady treatment with extracorporeal layer oxygenation and dialysis. Therefore, all of this will help as expected in patients with organ disappointment.

4. Conclusion

In our project, significant differences between pandemic phases or waves were concluded by downloading COVID-19 database from NCBI database. Performing multiple sequence alignment for our COVID-19 samples. In addition, generating a phylogenetic tree and visualizing it. Furthermore, Using BLAST to align COVID-19 spike protein with other viral spike proteins. Finally, by choosing the most homologous which is **KT182953.1 Middle East respiratory syndrome coronavirus** and annotating it using NCBI and Uniprot. Therefore, we can generate COVID-19 sequences at different time points, which can help us later in understanding the evolution of infectious diseases.

5. Suggestions & Future Work

In our future work, significant differences between pandemic phases or waves will be more discussed by downloading COVID-19 database from GISAID or NCBI database. Furthermore, working on the development of efficient and rapid sequencing methods to reconstruct the genomic sequence of SARS-CoV-2, the pathogenetic agent of COVID-19, as an essential part in the development of diagnostic molecular tests as well as the development of effective measures and strategies to mitigate the pandemic's spread. In addition, there should be a COVID-19 vaccination plan for individuals who are at risk of experiencing severe symptoms and the general population in outbreak areas to increase immunity in communities. Therefore, generating COVID-19 sequences at different time points can help us later in the evolution of infectious medical field.

6. References

I. I. Khan, A. Haleem, and M. Javaid, "Analysing COVID-19 pandemic through cases, deaths, and recoveries," Journal of Oral Biology and Craniofacial Research, vol. 10, no. 4, pp. 450–469, 2020

II. S. A. Shams, A. Haleem, and M. Javaid, "Analyzing COVID19 pandemic for unequal distribution of tests, identified cases, deaths, and fatality rates in the top 18 countries," Diabetes & Metabolic Syndrome: Clinical Research & Reviews, vol. 14, no. 5, pp. 953–961, 2020

III. S. Samadizadeh, M. Masoudi, M. Rastegar, V. Salimi, M. B. Shahbaz, and A. Tahamtan, "COVID-19: why does disease severity vary among individuals?," Respiratory Medicine, vol. 180, Article ID 106356, 2021.

IV.] R. Forman, S. Shah, P. Jeurissen, M. Jit, and E. Mossialos, "COVID-19 vaccine challenges: what have we learned so far and what remains to be done?," Health Policy, vol. 125, no. 5, pp. 553–567, 2021.

V. V. Parcha, K. S. Booker, R. Kalra et al., "A retrospective cohort study of 12,306 pediatric COVID-19 patients in the United States," Scientific Reports, vol. 11, no. 1, p. 10231, 2021.

VI. S. Suthar, S. Das, A. Nagpure et al., "Epidemiology and diagnosis, environmental resources quality and socio-economic perspectives for COVID-19 pandemic," Journal of Environmental Management, vol. 280, Article ID 111700, 2021

VII. V. Soriano, P. Ganado-Pinilla, M. Sanchez-Santos et al., "Main differences between the first and second waves of COVID-19 in Madrid, Spain," International Journal of Infectious Diseases, vol. 105, pp. 374–376, 2021.