الجمهورية الجزائرية الديمقراطية الشعبية

People's Democratic Republic of Algeria

وزارة التعليم العالي و البحث العلمي

Ministry of Higher Education and Scientific Research

Abdelhafid Boussouf-Mila University Center

Institute of Science and Technology Department of Natural and Life Sciences

Thesis prepared for obtaining the Master's Degree

Domain: Science of Life and Nature

Faculty: Biological Sciences **Specialty: Applied Biochemistry**

The Phylogeny of SARS Cov-2

(An Update of The Year 2022)

Presented By:

- **Boulesbàa Wiam**

In front of The Jury:

• **President :**	**Merzoug Amina**	**(M.C.A)**	**Mila University Center**
• **Examiner :**	**Boucherit Hanane**	**(M.C.B)**	**Mila University Center**
• **Supervisor :**	**Boubendir Abdelhafid**	**(M.C.A)**	**Mila University Center**

University Year: 2021/2022

Acknowledgements

First ad foremost, praises and thanks to Allah almighty for the strength he gave me, wisdom and good health in order to finish this thesis successfully.

I would like to express my sincere gratitude to my supervisor **Dr. Boubendir Abdelhafid** for all his guidance and support. I appreciate his continuous encouragement and advice during the writing of this thesis. The completion of this study Could not have been possible without his help.

I extend my respect and thanks to members of the jury who did us the honor of evaluating this work **Mrs Merzoug Amina** president of the jury and **Boucherit Hanane** examiner, who showed interest and availability for the examination of our brief.

I'm extremely grateful for my beloved mother **LEILA** and my dear father **MAHMOUD** for their caring, prayers and sacrifices for educating and preparing me for my future. I also thank my brothers **RAMY** and **ADEM** for always being supportive and helping, wishing them all the best in their life.

I want to give a special thanks to my husband **YASSER** for his belief in me that kept my spirit and motivation high during this process. I'm also thankful for my in-law's family for their support and valuable prayers.

To my friends, thank you all for your unwavering support and for reminding me to take breaks and have fun when I've been stressed out.

A big thanks to all those who have contributed directly or indirectly to the success of this work, whether material or moral.

List of abbreviations

ACE2: Angiotension converting enzyme 2

COVID-19: Coronavirus disease 2019

DNA: Deoxyribonucleic acid

ER: endoplasmic reticulum

GISAID: Global initiative on sharing all influenza data

IBV: infectious bronchitis virus

IRF: Host interferon regulatory factor 3

MEGA: Molecular evolutionary genetics analysis

MERS-CoV: Middle east respiratory syndrome coronavirus

MHV: Mouse Hepatitis coronavirus

mRNA: messenger ribonucleic acid

n-cov: Novel coronavirus

NJ: Neighbor-Joining

ORF: Open reading frame

NSP : Non structural protein

PP : Polyprotein

RNA: Ribonucleic acid

SARS CoV-2: severe acute respiratory syndrome coronavirus 2

TMPRSS2: transmembrane protease Serine 2

TGEV: Transmissible Gastroenteritis Virus

WHO: World Health Organization

List of figures

Figure 1. Transmission of coronaviruses from origin to humans……09

Figure 2. Genomic organization of SARS-CoV-2……10

Figure 3. Structure of SARS-CoV-2……11

Figure 4. Transmission pathways of COVID-19 infection……..12

Figure 5. Pattern Recognition receptors involved in detecting RNA viruses……13

Figure 6. The role of environmental factors in the spread of SARS-CoV-2……16

Figure 7. The genes of SARS-cov-2……18

Figure 8. The structure of Spike Protein……..28

Figure 9. Genomic sequence of SARS-CoV-2 N protein…….....29

Figure 10. Snake diagram of envelope (E) protein……..30

Figure 11. Phylogenetic tree of SARS-CoV-2 samples sequenced in Turkey……33

Figure 12. Phylogenetic tree with Maximum Composite Likelihood distance, representing 222 complete genomes including the four Chilean isolates. The name of the isolates were colored according to the variants as follows: S (blue), G (red), V (purple), unclassified variants (green), and the main clades were highlighted. Chilean strains are marked over the cladogram in the S and G variant clades……..34

Figure 13. Phylogenetic analysis of SARS-CoV-2 emergence in the Netherlands……..35

Figure 14. The main window of MEGA 7……..……..42

Figure 15. Complete genome phylogenetic tree of SARS Cov-2 strains from different area of Algeria. Isolates are represented with country/accession number of origin and the year of isolation……56

Figure 16. Estimates of Evolutionary Divergence between Sequences……..58

Figure 17. Complete genome phylogenetic tree of SARS Cov-2 strains from Algeria and different geographies in the world. Isolates are represented with country/accession number of origin and the year of isolation……..60

Figure 18. Estimates of Evolutionary Divergence between Sequences (GISAID Database)…62

List of tables

Table 1. Taxonomic classification of SARS-CoV-2……08
Table 2. The genes expressed by SARS CoV-2……20
Table 3. Brief description of various non-structural proteins of SARS-CoV-2……25
Table 4. Number of sequences in Algeria (From GISAID)……54
Table 5. Genomic sequences from databases……..…55
Table 6. Genomic sequences from databases with corresponding continent/ country/accession number/date/reference……..58

Table of Contents

Acknowledgements

List of abbreviations

List of figures

List of tables

Table of Contents

Introduction

Part I Bibliographic Research

Chapter 1: Generalities

1.Definition…….06

2. Historical…….06

3. Taxonomy……07

4. Origin……...08

5. Genetic structure and organization……..09

6. Transmission mode…….11

7. Pathophysiology……...12

8. Factors influencing the spread of SARS-CoV-2 around the world……..14

8.1 Geodemographic and environmental factors……..14

8.2 Other factors……...15

Chapter 2: Molecular biology

1. ORFs……18

1.1 Definition……18

1.2. Types of ORFs……...19

2. Non-structural proteins…….21

3. Accessory Proteins……...26

4. Structural proteins……27

4.1. Spike Protein……...27

4.2. Membrane Protein……..28

4.3. Nucleocapsid Protein……28

4.4. Envelope Protein……29

Chapter 3: The phylogeny of SARS Cov-2

1. The phylogeny of whole genome……32

1.1. Phylogenetic analysis of SARS-CoV-2 genomes in Turkey……..32

1.2. Phylogenetic analysis of SARS-CoV-2 genomes in Chile……33
1.3. Phylogenetic analysis of SARS-CoV-2 genomes in Netherlands……..35

Part II Experimental research

1. Material and Methods

1.1. Introduction to databases……...38
1.2. Collection of genomic sequences from databases…….38
1.3. Steps of research on GISAID……38
1.4. The stages of phylogenetic tree construction……41
1.4.1. Introducing MEGA 7……41
1.4.2. Sequence alignment……..42
1.4.3. Estimating Scalable Distances Using Pairwise Distance ……..……...49
1.4.4. Construction of the phylogenetic tree……51

2. Results and discussion

2.1. The study of SARS Cov-2 in different regions of Algeria……54
2.2. The study of SARS Cov-2 in Algeria and different geographies in the world…….58
Conclusion……63
List of references……..65
Abstract
Résumé
ملخص

Introduction

Introduction:

On December 31, 2019 an outbreak caused by a new human pathogen, lately named severe acute respiratory syndrome coronavirus 2 (SARS CoV-2) was notified in Wuhan, China (Gorbalenya et al.,2020).

The virus caused a severe respiratory syndrome, generically defined as coronavirus disease (covid 19) human to human transmission was first reported on January 22, 2020 (Huang et al., 2020)

Coronaviruses are enveloped, single-strand RNA viruses that can infect a wide range of hosts including avian, wild, domestic mammalian species, and humans. Coronaviruses are well known for their ability to mutate rapidly, alter tissue tropism, cross the species barrier, and adapt to different epidemiological situations (Decaro et al., 2010).

SARS CoV-2 is a new human coronavirus which emerged in people's Republic of China at the end of 2019 and is responsible for the global covid 19 pandemic that caused more than 540,000 deaths in six months. Understanding what the origin of this virus is an important issue and it is necessary to determine the mechanisms of its dissemination in order to be able to contain new epidemics. (Sallard et al., 2020)

Several control measures are being instituted by nations around the world to extinguish the SARS-CoV-2 pandemic, including the issuance of travel advisories or even flight bans to and from infected countries, strict quarantine measures and traveler screenings, implementation of mitigation measures by healthcare specialists, application of social distancing measures for schools and popular gatherings, strict personal hygiene such as frequent handwashing, and wearing face masks (Cowling et al., 2020)

The clinical manifestations and lethality of SARS-CoV-2 infection widely vary across individuals, depending on a number of factors (already recognized or just hypothesized), that have been discussed in a dedicated paragraph. Approximately half of the infected subjects remain asymptomatic, and most of the remaining only experience influenza-like symptoms, including fever, cough, sore throat, runny nose, weakness, myalgia, headache and, less frequently, conjunctivitis, hemoptysis, olfactory and taste dysfunction, and diarrhea (Cecilia et al., 2020)

Our study is divided into two essential parts, the first part presents a bibliographical synthesis in which we bring a first chapter which is devoted in the study of severe acute respiratory syndrome coronavirus 2 (SARS Cov-2) a second chapter outlines the molecular biology of the virus and a third chapter of the phylogeny of SARS Cov-2.

The second experimental part is reserved for the phylogenetic study of the full genome of SARS CoV-2 available in databases on the internet, the selected isolates come from different geographies around the world, among them Algerians isolates are targeted. Then comes the construction of the matrix of evolutionary distances and dendrograms. The bioinformatics tool used in the present study is the MEGA 7 software.

.

Part I
Bibliographic Research

Chapter 1: Generalities

Chapter 1: Generalities

1.Definition

Coronavirus disease (covid-19) is an infectious disease caused by the SARS CoV-2 virus. Most people infected with the virus will experience mild to moderate respiratory illness and recover without requiring special treatment. However, some will become seriously ill and require medical attention older people and those with underlying medical conditions like cardiovascular disease, diabetes, chronic respiratory disease, or cancer are more likely to develop serious illness. Anyone can get sick with covid-19 and become seriously ill or die at any age (WHO 2020).

Corona viruses are members of the family coronavidae, the enveloped viruses that possess extraordinarily large single-stranded RNA genomes ranging from 26 to 32 kilobases in length (Shuo et al., 2016)

In late December 2019, a number of local health authorities reported clusters of patients with pneumonia of unknown cause, which were epidemiologically linked to a seafood market in Wuhan, Hubei Province, China. The pathogen, a novel coronavirus (SARS Cov-2), was identified by local hospitals using a surveillance mechanism for "pneumonia of unknown etiology" that was established in the wake of the 2003 SARS outbreak with the aim of allowing timely identification of novel pathogens (Di Wu et al., 2020)

2. Historical

Human coronaviruses were first characterized in the 1960s from respiratory tract infections. The two first isolated viruses were B814 and 229E. Since then, several other coronavirus strains have been isolated from humans using tissue culture (OC16 and OC43). The number of identified coronaviruses has continued to increase significantly to include viruses of several additional animal species such as calves, dogs, cats, bats, sparrows, rabbits, and turkeys (Helmy et al., 2020)

In December 2019, SARS CoV-2 emerged in Wuhan City, China, causing severe respiratory illness and mortality. Early studies reported that it may have evolved from bats, as revealed by phylogenetic analysis and its high identity (96.3%) with the bat coronavirus RaTG13 (Zhou et al., 2020)

It is important to revisit the past pandemics of SARS-CoV and MERS, due to the structural and molecular similarities between these viruses and SARS Cov-2. Historically, SARS originated in

Guangdong, China, then spread to Hong Kong, infected 1755 people with 299 deaths, and caused an economic downturn. Due to the rapid increase of the number of infected cases and infected countries, the WHO declared SARS CoV-2 a pandemic on 11 March 2020 and on 13 March 2020 (WHO 2020)

Interestingly, researchers also reported one amino acid difference in the receptor-binding domain of the S protein of Pangolin-CoV compared to that of SARS CoV-2, suggesting that pangolins might play a role as an intermediate host. Another group of researchers reported that the virus originated from bats based on the genome sequence of SARS-CoV-2, which is 96% identical to bat coronavirus RaTG13. There were speculations that SARS-CoV-2 is a laboratory-engineered CoV and leaked directly from a laboratory in Wuhan where a bat CoV (RaTG13) was recently reported. However, there is no evidence to support this allegation. (Liu et al., 2020).

3. Taxonomy

In terms of taxonomy, SARS CoV-2 is (just) another virus in the species severe acute respiratory syndrome-related coronavirus. In this respect, the discovery of this virus differs considerably from the description of the two other zoonotic coronaviruses, SARS-CoV and MERS-CoV, introduced to humans in the 21st century. Both these viruses were considered novel by this study group based on prototyping two species and two informal subgroups of the Betacoronavirus genus that were recently recognized as subgenera Sarbecovirus and Merbecovirus (Gorbalenya et al., 2020).

Table 1. Taxonomic classification of SARS CoV-2 (Gorbalenya et al., 2020).

Category	Virus
Realm	Riboviria
Order	Nidovirales
Suborder	Cornidovirineae
Family	Coronaviridae
Subfamily	Coronavirinae/ Orthocoronavirinae
Genus	Betacoronavirus
Subgenus	Sarbecovirus
Species	Severe acute respiratory syndrome-related coronavirus
Strain	Severe acute respiratory syndrome coronavirus 2 (SARS CoV-2)

4. Origin

Since the outbreak of severe acute respiratory syndrome (SARS) 18 years ago, a large number of SARS-related coronaviruses (SARSr-CoVs) have been discovered in their natural reservoir host, bats. Previous studies have shown that some bat SARSr-CoVs have the potential to infect humans. Here we report the identification and characterization of a new coronavirus (2019-nCoV), which caused an epidemic of acute respiratory syndrome in humans in Wuhan, China. Full-length genome sequences were obtained from five patients at an early stage of the outbreak. The sequences are almost identical and share 79.6% sequence identity to SARS-CoV. Furthermore, we show that 2019-nCoV is 96% identical at the whole-genome level to a bat coronavirus (Zhou et al., 2020).

According to the ICTV criteria, only the strains found in Rhinolophus bats in European countries, Southeast Asian countries and China are SARSr-CoV variants. Those from Hipposideros bats in Africa are less closely related to SARS-CoV and should be classified as a new coronavirus species (Tong et al., 2009).

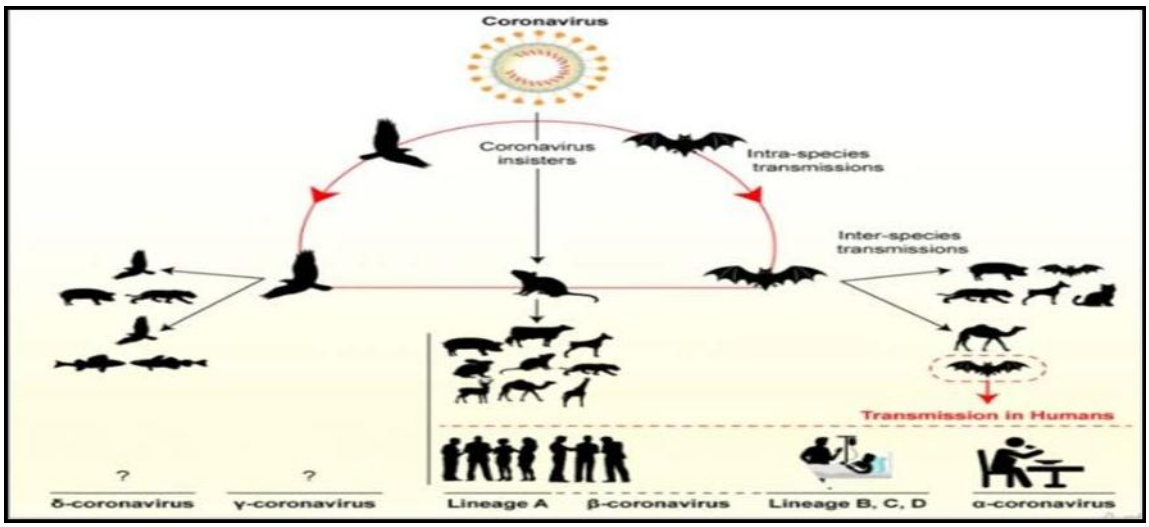


Figure 1. Transmission of coronaviruses from origin to humans (Machhi et al., 2020).

All results on the origin, evolution and spread of SARS-CoV-2 come from computer analysis of these sequences, coupled with associated metadata such as the date and place of sequencing, the sequencing technique, etc. In some cases, it has been possible to determine the precise origin of a virus found at a given location by contact tracing (Anna et al., 2021)

5. Genetic structure and organization

Coronaviruses are non-segmented, enveloped and positive sense single-stranded RNA virus genomes. They have a size ranging from 26 to 32 kilobases. The virion consists of a nucleocapsid composed of genomic RNA and phosphorylated nucleocapsid (N) protein, covered by two different types of spike proteins: the hemagglutinin-esterase (HE) (found in some CoVs) and the spike glycoprotein trimmer (S) (found in all CoVs). The S protein in the virus envelop consists of the membrane (M) protein and the envelope (E) protein (Guo et al., 2020)

The genetic makeup of SARS-CoV-2 is composed of 13–15 (12 functional) open reading frames (ORFs) containing ~30,000 nucleotides. The genome contains 38% of the GC content and 11 protein-coding genes, with 12 expressed proteins. The genetic arrangement of ORFs highly resembles the SARS-CoV and MERS-CoV (Rota et al., 2003)

The coronavirus spike protein (S) is a type I glycoprotein that forms the peplomers on coronavirus particles. Some coronaviruses spikes (most from group II and III viruses) are cleaved

into two subunits by a furin-like enzymatic activity during processing in the Golgi. The prototype MHV spike is 180 kDa; for most MHV strains, it is cleaved into two noncovalently associated subunits of about 90 kDa (Stroher et al., 2004)

The M protein is the most abundant virion membrane protein. Aside from its role in viral assembly, the coronavirus M protein is believed to have functions in host interactions. It may be O glycosylated (groups I and III) or N glycosylated (group II). While glycosylation is not essential for viral assembly or infectivity. For MHV, the selection of recombinant viruses with N, O, or no glycosylation demonstrated that while the glycosylation state of M protein does alter the ability to replicate in vitro, it may affect the ability to induce IFN-α in vitro and also to replicate in the liver in vivo (de Haan et al., 2003).

In addition to its role as structural protein, N protein plays a role in transcription and also in pathogenesis (Yount et al., 2003).

While MHV proteins are generally restricted to the cytoplasm, the nucleocapsid proteins of coronaviruses representing groups I, II, and III were shown to localize to the nucleolus as well as to the cytoplasm (Wurm et al 2001)

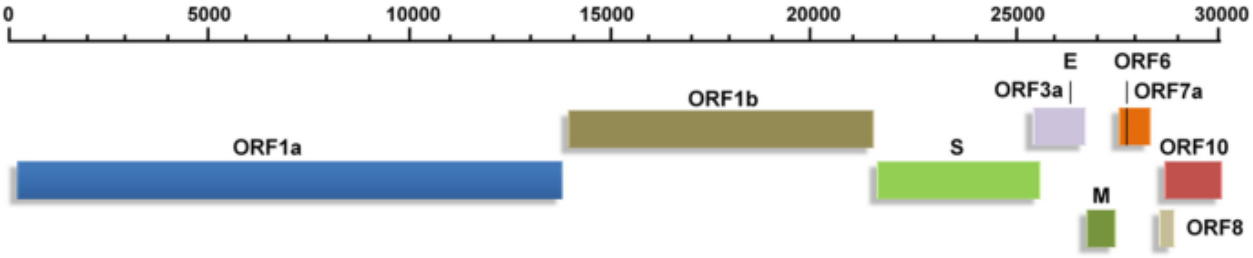


Figure 2. Genomic organization of SARS-CoV-2 (Kumar et al., 2020)

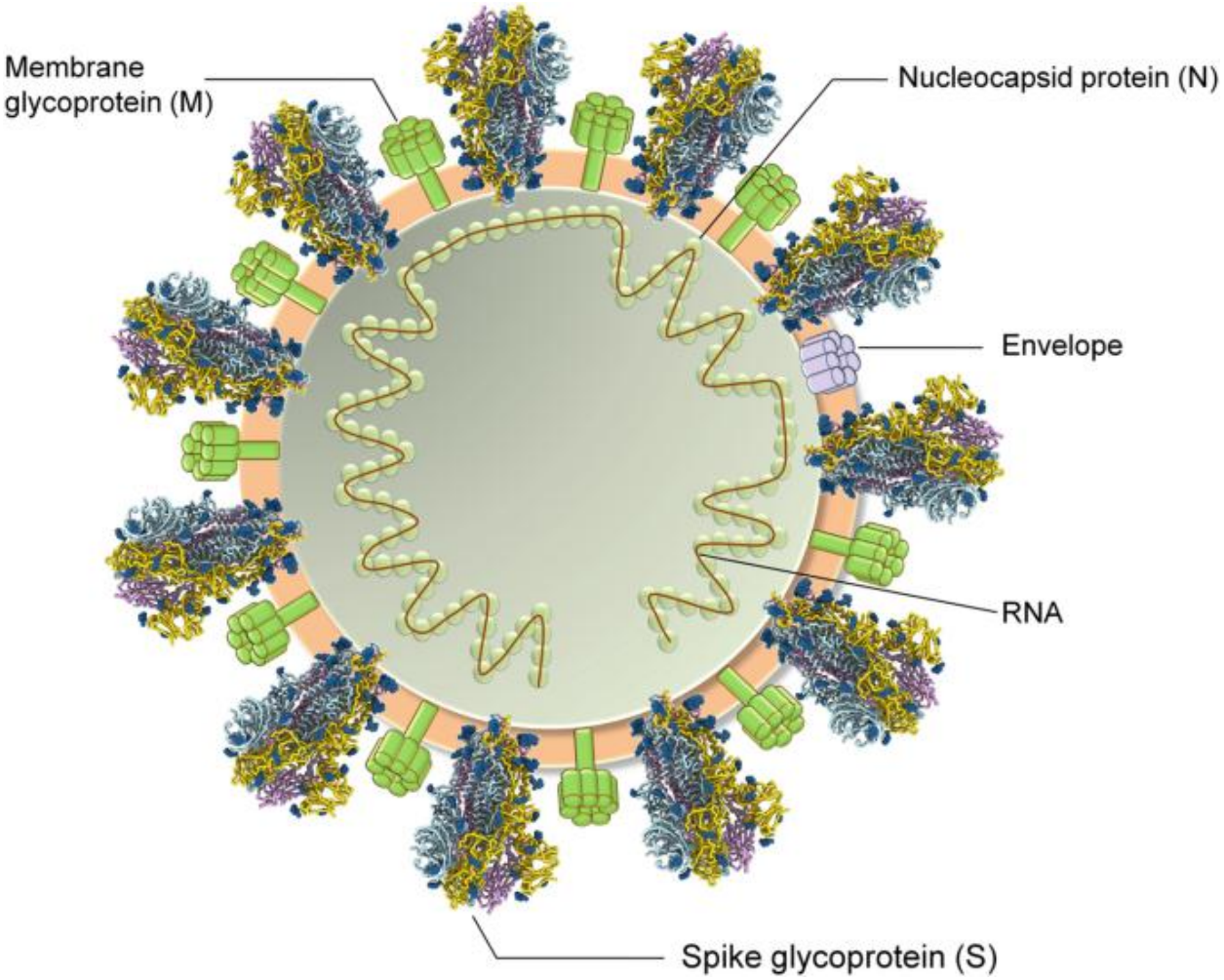


Figure 3. Structure of SARS-CoV-2 (Kumar et al., 2020)

6. Transmission mode

SARS-CoV-2 was indicated as animal-to-human transmission because of the linkage between fish and wild animal market based on the early studied (Li et al., 2020). However, it showed that asymptomatic infected cases with SARS-CoV-2 is highly contagious and transmitted through close contact (Zhao et al., 2020). Therefore, the transmission of SARS-CoV-2 occurs primarily through respiratory droplets and aerosols generated during coughing or sneezing, which may land on the nose, mouth or eyes (Li et al., 2020).

Coronavirus can sustain for a long time on various surfaces which is a major reason for its transmission. This virus can contaminate on different metal surfaces and stay on them from hours

to days, with a maximum span on plastic and stainless and least on the copper surface (Rajiv et al., 2020).

Recently, there is a patient dog infected from its nasal and oral cavity with a low level of COVID-19 virus. It's not shown any signs of disease before. This is the first case of human-to-animal transmission that has been confirmed by the experts from School of Public Health in the Hong Kong University, College of Veterinary Medicine & Life Sciences in the City University of Hong Kong and the World Organization for Animal Health (COVID., 2020).

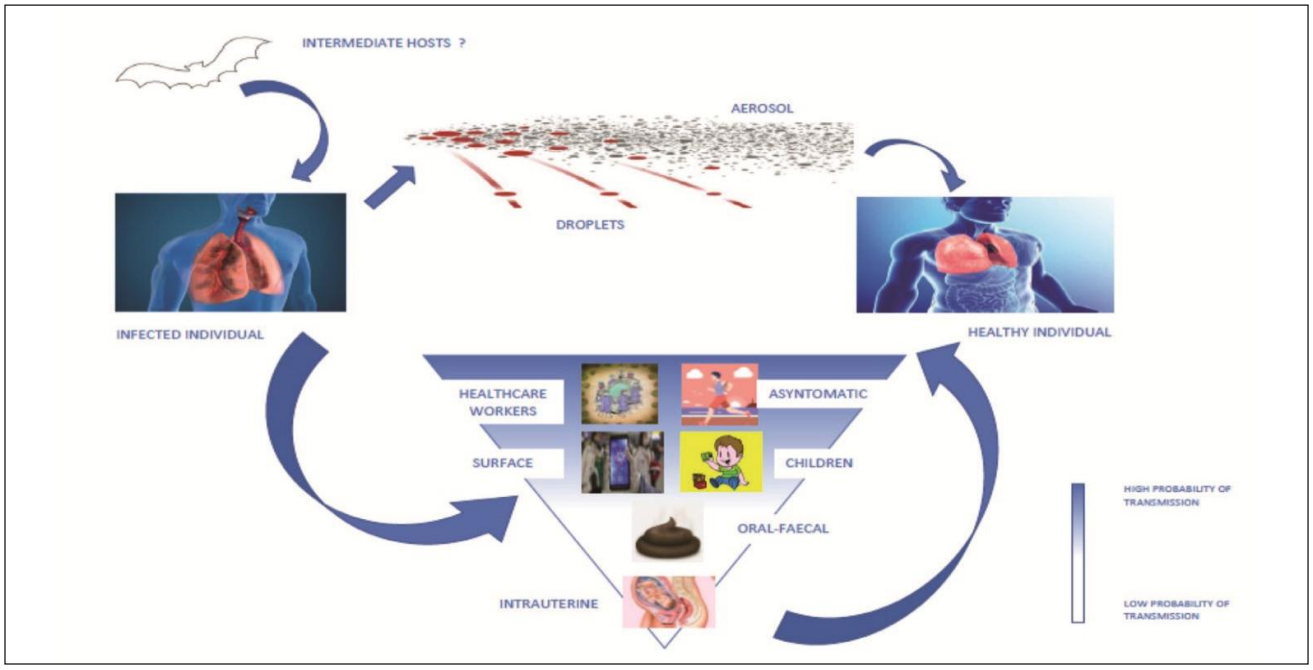


Figure 4. Transmission pathways of COVID-19 infection (Vella et al., 2020)

7. Pathophysiology

Structurally and phylogenetically, SARS-CoV-2 is similar to SARS-CoV and MERS-CoV and is composed of four main structural proteins: spike (S), envelope (E) glycoprotein, nucleocapsid (N), membrane (M) protein, along with 16 nonstructural proteins, and 5-8 accessory proteins (Rajnik et al., 2021).

The Angiotensin Converting Enzyme Receptor (ACE2) is the primary route of entry for the virus. This receptor is expressed for 80% in type II pulmonary alveolar cells, but also in the nasal mucosa, upper respiratory tract, vascular endothelium, cardiac, renal and digestive tissues. The serine protease TMPRSS2, which is highly expressed by endothelial cells in the respiratory and

digestive tracts, also appears to play an important role as a co-receptor for the virus (Couvreur et al., 2021).

Once SARS-CoVs enter the host via the respiratory tract, airway and alveolar epithelial cells, vascular endothelial cells and alveolar macrophages are among their first targets of viral entry. These cells are probably ‘ground-zero’ for early infection and subsequent replication due to their expression of ACE2. Although ACE2 mRNA is detected in human and many mammalian (bat, ferret, cat, dog, etc.) lung biopsies, their expression is rather low compared with extrapulmonary tissues. Thus, the permissiveness of these cells to SARS-CoVs may depend on additional, unappreciated cell-intrinsic factors that aid in efficient infection. (Harrison et al., 2020).

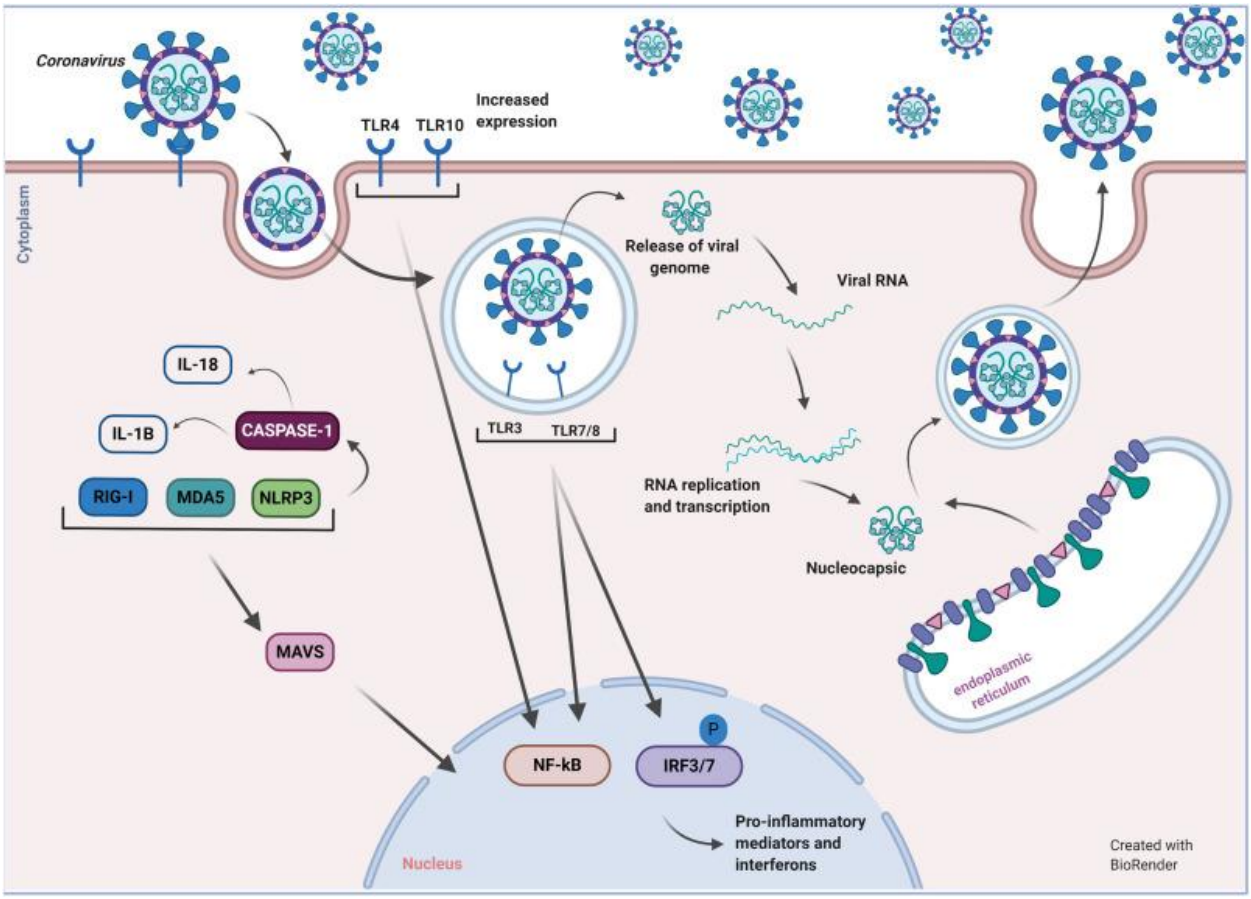


Figure 5. Pattern Recognition receptors involved in detecting RNA viruses (Gerwyn et al., 2020).

8. Factors influencing the spread of SARS-CoV-2 around the world

8.1 Geodemographic and environmental factors

There are several factors involved in transmitting the virus. These conditions can be included in environment and human behavior. The distribution of human population, migration, social interactions, climate change (deforestation, habitat invasion), agricultural growth, and direct contact with domestic and wild animals fall into this category. On the other hand, the effect of environmental factors such as ambient temperature, humidity, etc., relating to the COVID-19 pandemic has not been sufficiently investigated. How the virus enters the body (eyes, ears, mouth, and nose) is not well known by the release of aerosols and droplets containing SARS-CoV-2 in human societies. (Eslami & Mahrokh., 2020)

- **water and waste wasters**

Guaranteeing safe water, collecting sewage, and maintaining effective hygiene during infectious diseases, including COVID-19 pandemic, play a key role in supporting human health (Eslami et al., 2018; WHO 2020). Some studies have reported that coronavirus can remain in water or wastewater sources for days or weeks (Qu et al., 2020).

The SARS-CoV-2 is likely to be significantly inactivated more rapidly than human intestinal viruses without water-borne diseases in contact with oxidants (WHO 2020).

- **insects**

Numerous factors are contributing to the spread of global diseases such as COVID-19, which is spreading worldwide. So far, 7 types of them have been identified. The transmission and evolution of the SARS-CoV-2 from bats to scaly anteaters (pangolins) and then to humans has been reported. Some studies have associated insects such as beetles and domestic insects, which are the main mechanical carriers of pathogens, by contact with contaminated surfaces and patients' secretions involved in transmission. SARS-CoV-2 excretion by stool has been confirmed in some patients. Therefore, feeding domestic insects and beetles from the stool and its mechanical transmission can play a significant role in the transmission of the disease (Eslami & Mahrokh., 2020).

- **Age**

The geographic distribution of countries makes it easy to predict more cases in high-density areas. Moreover, urban areas, especially administrative and/or economic capitals or megacities, are often remarkably densely populated, creating conditions that facilitate the spread of the virus, while rural and Saharan areas stand to benefit greatly. geographical distance and lack of promiscuity, to be at very low risk of contamination. most young individuals would be asymptomatic or have mild forms, and would likely go undetected, with a risk of infecting more people than those who are symptomatic (Hoummadi et al., 2020)

- **Climatic factors**

One study observed the relationship between the numbers of positive daily SARS-CoV-2 cases with three environmental factors: maximum relative humidity, maximum temperature, and maximum wind speed, Although, in most cases, with increasing humidity and wind speed, the prevalence has decreased (Eslami et Mahrokh., 2020).

In Korea, researchers found that the risk of influenza incidence was significantly increased with low daily temperature and low/high relative humidity (RH), but a positively correlated with diurnal temperature range (DTR). Absolute humidity (AH) had significant correlations with influenza viral survival and transmission rates. One important feature of COVID-19 epidemic is that the countries currently suffering most from the disease are most located in the regions with low temperature. Therefore, meteorological factors, such as ambient temperature (AT) and humidity, might play an important role in the spread of the disease. Many factors might influence the COVID-19 epidemic, including social and political factors, geographical factors, climatic factors, etc. When only considering the temperature in single-factor model in the higher-temperature group, every 1 °C increase in the minimum temperature leads to a decrease of the cumulative number of COVID-19 cases by 0.86 etc. reported that weather was related to the spread of COVID-19, but the increase of temperature may not necessarily lead to declines in case counts without the implementation of extensive public health interventions (Liu et al., 2020)

8.2 Other factors

➢ **Tourism**

COVID-19 has emerged in a world closely connected by local and international population on the move, with more people moving for work, education and family reasons, tourism and survival than ever before. Intense population movements, especially of tourists and business workers, have been a key factor in the global spread of the epidemic (Indseth et al., 2021).

➢ **Immigration**

The much higher toll of notified COVID-19 infections among several immigrant groups compared to non-immigrants suggests a need for actions such as enhancing community engagement and health communication strategies to lower the thresholds for being tested. We observed a higher COVID-19 notification rate among immigrants compared to non-immigrants and a much higher rate of hospitalization, with major differences between different immigrant groups (Indseth et al., 2021).

The spread of COVID-19 from China has been clearly linked to those traveling from Wuhan in central China's Hubei province. Therefore, it is important to understand the travel density/volume of passengers carried as well as the routes departing from Wuhan via the major connected regional air transport hubs across China. A model has been developed on the intensity of migration and travel that can explain the outbreak and spread of COVID-19 since its onset in late 2019 (Sirkeci et Mustafa, 2020).

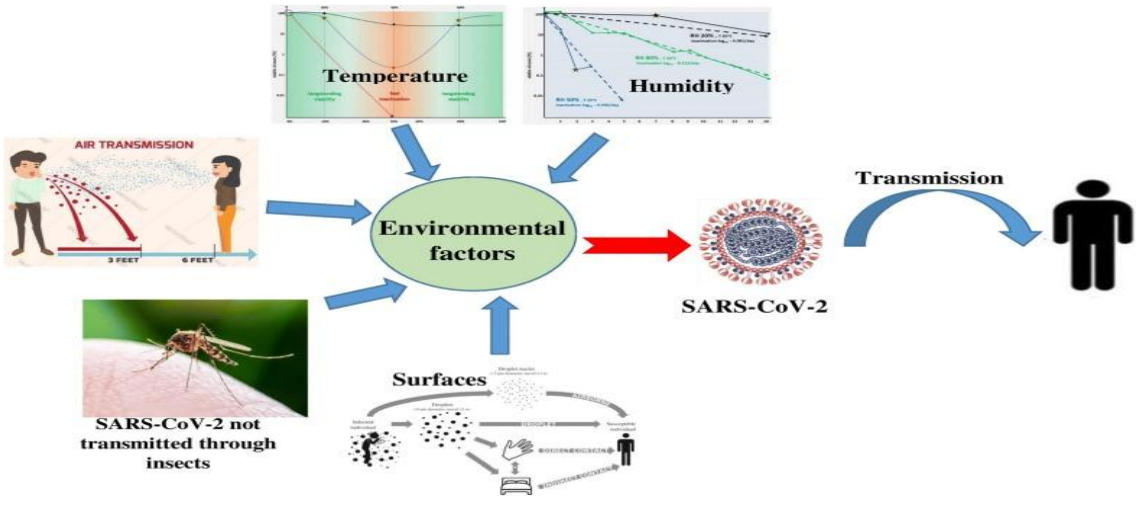


Figure 6. The role of environmental factors in the spread of SARS-CoV-2 (Eslami & Mahrokh., 2020)

Chapter 2:
Molecular biology

Chapter 2: Molecular biology

1. Open Reading Frames (ORFs)

1.1 Definition

The nCoV contains at least six open reading frames (ORFs) and many other accessory genes like other CoVs. The 5′ terminal two-thirds of the genome contains two open reading frames (ORFs), ORF1 and ORF2 which encodes two polyproteins, pp1a and pp1ab, which is further cleaved into 11 and 16 proteins, respectively. These 16 mature proteins are responsible for several important functions in genome maintenance and virus replication. The structural proteins namely spike (S), an envelope protein (E), membrane protein (M) and nucleocapsid (N) are located at the one-third 3′ terminal of the genome. In addition to these genes, there are several accessory proteins which help in virus replication (Yashpal et al., 2020).

In addition, a number of subgroup-specific accessory genes are found interspersed among, or even overlapping, the structural genes. Overlapping genes originate by a mechanism of overprinting, in which nucleotide substitutions in a pre-existing frame induce the expression of a novel protein in an alternative frame. The accessory proteins in coronaviruses vary in number, location and size in the different viral subgroups, and are thought to contain additional functions that are often not required for virus replication, but are involved in pathogenicity in the natural host (Christian et al., 2020)

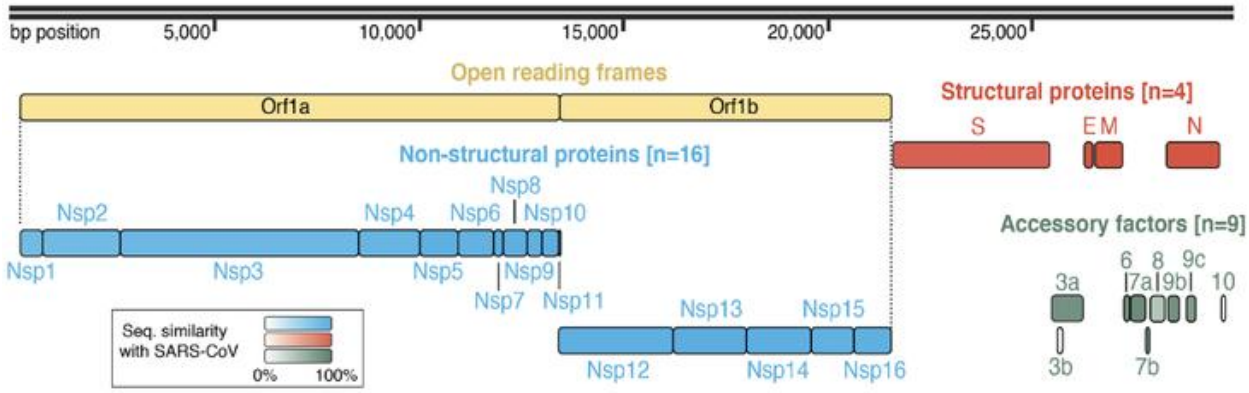


Figure 7. The genes of SARS-cov-2 genome (Gordon et al., 2020)

1.2. Types of ORFs

- **ORF 1**

Upon entry into the cell, the 5′ end of the genome RNA, ORFs 1a and 1b, are translated into pp1a and pp1ab; pp1ab is translated via a frameshift mechanism, which occurs at high frequency (25 to 30%). ORF 1a encodes one or two papain-like proteases (PLpro or PLP) and a picornavirus 3C-like protease (3CLpro), which function to process pp1a and pp1ab into the mature replicase proteins. Also, encoded in the X domain of ORF 1a is a (putative) ADP-ribose 1"-phosphatase activity (Susan et al., 2005)

- **ORF 3**

The ORF3a protein from SARS cov is an ion channel protein related to NLRP3 inflammasome activation. ORF3a interacts with TRAF3, which in turn activates ASC ubiquitination, and as a result, leads to activation of caspase 1 and IL-1β maturation (Siu K-L et al.,2019)

- **ORF 6**

The ORF6 protein from SARS coronavirus is an accessory protein that plays an important role in viral pathogenesis (Zhao J et al.,2009) Using a yeast two-hybrid system, ORF6 was shown to interact with NSP8, the nonstructural protein related to promoting RNA polymerase activity (Kumar P et al.,2007)

- **ORF 7**

ORF7a from SARS coronavirus is an accessory protein that is a type I transmembrane protein and its crystal structure has been determined (Nelson CA et al.,2005)

The ORF7b accessory protein from SARS coronavirus is localized in the Golgi compartment (Schaecher SR et al., 2007)

- **ORF8**

SARS CoV-2 has a single ORF8 protein while SARS CoV has two ORF8 proteins: ORF8a and ORF8b (Le TM et al.,2007) In SARS CoV, the ORF8b protein binds to the IRF association domain (IAD) region of interferon regulatory factor 3 (IRF3), which in turn inactivates interferon signaling (Wong HH et al.,2018) Interestingly, L84S and S62L missense mutations have been reported in various SARS CoV-2 sequences (Khailany RA et al.,2020)

- **ORF10**

ORF10 protein from SARS CoV-2 is comprised of 38-amino acids and its function is unknown. Interestingly, SARS CoV possesses an ORF9b protein, which is not present in SARS CoV-2 (Mu J et al.,2020)

Table 2. The genes expressed by SARS CoV-2(Yoshimoto et al., 2020)

Number(#)	Gene	GeneID	Location	Protein	[LOCUS]
1(7,096)	ORF1ab	43,740,578	266–21,555	ORF1ab polyprotein	[BCB15089.1/BCB97900.1]
1(4,405)	ORF1a	43,740,578	266–13,483	ORF1a polyprotein	[YP_009725295.1]
2(1,273)	ORF2 (S)	43,740,568	21,563–25,384	Spike protein (S protein)	[BCA87361.1]
3(275)	ORF3a	43,740,569	25,393–26,220	ORF3a protein	[BCA87362.1]
4(75)	ORF4 (E)	43,740,570	26,245–26,472	Envelope protein (E protein)	[BCA87363.1]
5(222)	ORF5 (M)	43,740,571	26,523–27,191	Membrane protein (M protein)	[BCA87364.1]
6(61)	ORF6	43,740,572	27,202–27,387	ORF6 protein	[BCA87365.1]
7(121)	ORF7a	43,740,573	27,394–27,759	ORF7a protein	[BCA87366.1]
8(43)	ORF7b	43,740,574	27,756–27,887	ORF7b protein	[BCB15096.1]
9(121)	ORF8	43,740,577	27,894–28,259	ORF8 protein	[BCA87367.1]
10(419)	ORF9 (N)	43,740,575	28,274–29,533	Nucleocapsid phosphoprotein (N protein)	[BCA87368.1]
11(38)	ORF10	43,740,576	29,558–29,674	ORF10 protein	[BCA87369.1]

2. Non-structural proteins

The ORF1ab gene of SARS CoV-2 results in the expression of a polypeptide that is cleaved into 16 nonstructural proteins (Yoshimoto FK 2021).

There are several non-structural proteins, namely NSP1 to NSP 10 and NSP12 to NSP16, encoded by genes located within the 5 0 -region of viral RNA genome. These non-structural proteins with their corresponding functions along with other associated molecular features are tabulated below (Yadav et al., 2021)

- **NSP1**

Nonstructural protein 1 (NSP1) is the first protein of the polyprotein of SARS CoV-2. This protein is also known as the leader protein. This protein is also found in SARS coronavirus and is known to be a potent inhibitor of host gene expression. NSP1 binds to the 40S ribosome of the host cell to inactivate translation and promotes host mRNA degradation selectively, while the viral SARS CoV mRNA remain intact (Huang C et al.,2011)

- **NSP2**

Nonstructural protein 2 (NSP2) is the second protein of the polyprotein of SARS CoV-2. This protein is conserved in SARS CoV, the related beta coronavirus to SARS CoV-2. In SARS CoV, NSP2 was found to bind to two host proteins: prohibitin 1 and prohibitin 2 (PHB1 and PHB2) (Cornillez-Ty CT et al., 2009)

- **NSP3**

NSP3 is the papain-like proteinase protein. This protein is nearly 200 kDa in size and is the largest protein (not including the polyproteins ORF1a and ORF1ab) encoded by the coronaviruses. With such a long sequence, it possesses several conserved domains: ssRNA binding, ADPr binding, G-quadruplex binding, ssRNA binding, protease (papain-like protease), and NSP4 binding), and transmembrane domain. Among the 16 nonstructural proteins, NSP3, NSP4, and NSP6 have transmembrane domains (Sakai Y et al., 2017).

The papain like protease 1 (PL1 protease) of alpha coronavirus (alpha CoV) Transmissible Gastroenteritis Virus (TGEV), which is part of NSP3, was shown to cleave the site between NSP2 and NSP3. Furthermore, this papain like protease domain is responsible for the release of NSP1, NSP2, and NSP3 from the N-terminal region of polyproteins 1a and 1ab from coronaviruses (Lei J et al., 2018).

- **NSP4**

NSP4 interacts with NSP3 and possibly host proteins to confer a role related to membrane rearrangement in SARS CoV. Moreover, the interaction between NSP4 and NSP3 is essential for viral replication (Sakai Y et al., 2017)

- **NSP5**

The NSP5 protein based on the Middle East Respiratory Syndrome (MERS) coronavirus has been characterized. NSP5 cleaves at 11 distinct sites to yield mature and intermediate nonstructural proteins (NSPs) (Tomar S et al.,2015)

- **NSP6**

The NSP6 protein of the avian coronavirus (infectious bronchitis virus, IBV) was shown to generate autophagosomes from the endoplasmic reticulum (ER) Autophagosomes facilitate assembly of replicase proteins. Furthermore, NSP6 limited autophagosome/lysosome expansion, which in turn prevents autophagosomes from delivering viral components for degradation in lysosomes (Cottam EM et al.,2014) With SARS CoV, NSP6 was shown to induce membrane vesicles (Angelini MM et al.,2013)

- **NSP7**

NSP7 is required to form a complex with NSP8 and NSP12 to yield the RNA polymerase activity of NSP8 (te Velthuis AJ et al.,2012)

- **NSP8**

NSP8 is a peptide cofactor that makes a heterodimer with NSP7 (the other peptide cofactor), and this NSP7-NSP8 heterodimer complexes with NSP12. In addition to the NSP7-NSP8 heterodimer, an NSP8 monomer unit also complexes with NSP12, which ultimately forms the

RNA polymerase complex. The cryo-EM structure of this complex has been solved (Gao Y et al., 2020)

- **NSP9**

NSP9 from the porcine reproductive and respiratory syndrome virus (PRRSV) has been found to interact with the DEAD-box RNA helicase 5 (DDX5) cellular protein (Zhao S et al.,2015)

- **NSP10**

NSP10 has been shown to interact with NSP14 in SARS coronavirus, and this interaction stimulates activity of NSP14. NSP 14 is known to function as an S-adenosylmethionine (SAM)-dependent (guanine-N7) methyl transferase (N7-MTase) (Ma Y et al.,2015). Furthermore, NSP10 has also been shown to stimulate the activity of NSP16, which is a 2′-O-methyltransferase (Wang Y et al.,2015).

- **NSP11**

The function of NSP11 seems to be unknown. NSP11 is made of thirteen amino acids and the first nine amino acids are identical to the first nine in NSP12 (Subissi L et al.,2014).

- **NSP12**

NSP12 is the RNA-dependent RNA polymerase that copies viral RNA. As mentioned, NSP12 makes a complex with an NSP7-NSP8 heterodimer and an NSP8 monomer to confer processivity of NSP12. NSP12 exhibits poor processivity in RNA synthesis—that is the presence of NSP7 and NSP8 lowers the dissociation rate of NSP12 from RNA (Subissi L et al., 2014).

- **NSP13**

SARS CoV was used to characterize the helicase enzyme, NSP13, which unwinds duplex RNA (Jang K-J et al.,2020).

The crystal structure of NSP13 of SARS Cov has been reported (Jia Z et al., 2019). Furthermore, it has been shown that binding of NSP12 with NSP13 can enhance the helicase activity of NSP13. In addition to its helicase activity, NSP13 of SARS CoV is also known to possess 5′-triphosphatase activity, which is responsible for introducing the 5′-terminal cap of the viral mRNA. Both

eukaryotic and most viral mRNA have a 5′-terminal cap structure: m7G(5)ppp(5)N-. This 5′-terminal cap is the site of recognition for translation and plays a role in splicing, nuclear export, translation, and stability of mRNA (Ivanov KA et al., 2004).

- **NSP14**

Nsp14, a bifunctional replicase subunit with exoribonuclease activity (proofreading) that acts in a 3´ to 5´ direction on both ssRNA and dsRNA and a guanine-N7 methyltransferase activity at its C-terminal (Gorla et al., 2020).

The guanine-N7-methyltransferase activity is part of the process for introducing the 5′-cap of the virus, which involves multiple steps: the gamma-phosphate of the 5′end of nascent mRNA is removed by the RNA triphosphatase (NSP13) (Ivanov KA et al., 2004)

- **NSP15**

NSP15 of SARS coronavirus has been biochemically characterized as an endoribonuclease that cleaves RNA at uridylates at the 3′-position to form a 2′-3′ cyclic phosphodiester product (Bhardwaj K et al.,2006)

The NSP15 protein specifically targets and degrades the viral polyuridine sequences to prevent the host immune sensing system from detecting the virus (Hackbart M et al.,2020). NSP15 uses manganese as a cofactor to promote endoribonuclease activity (Bhardwaj K et al.,2004) It has been suggested that NSP15 degrades viral dsRNA to prevent host recognition (Deng X et al.,2018).

- **NSP16**

NSP16 for coronavirus has been biochemically (feline coronavirus, FCoV) and structurally (complex of NSP10-NSP16 for SARS CoV) characterized. The viral RNA has a 5′-cap, which protects it from mRNA degradation by 5′-exoribonucleases, promotes mRNA translation, and prevents the viral RNA from being recognized by innate immunity mechanisms (Decroly E et al., 2011).

The RNA cap is an N7-methylated guanine nucleotide connected through a 5′-5′ triphosphate bridge to the first transcribed nucleotide (adenine). NSP16 methylates the 2′-hydroxy group of adenine using S-adenosylmethionine as the methyl source (Decroly E et al.,2011).

Table 03 summarizes the list of SARS-CoV-2 non-structural proteins (NSPs) and their molecular functions

Table 03. Brief description of various non-structural proteins of SARS-CoV-2 (Yadav et al., 2021)

Name	Protein (Full Name)	Length (aa)	Range	Role	Accession No.
NSP1	N-terminal product of the viral replicase	180	1–180	Leader protein which acts as host translation inhibitor and also degrade host mRNAs	YP_009725297
NSP2	N-terminal product	638	181–818	Binds to prohibitin 1 and prohibitin 2 (PHB1 and PHB2)	YP_009725298
NSP3	Papain-like proteinase	1945	819–2763	Responsible for release of NSP1, NSP2, and NSP3 from the N-terminal region of pp1a and 1ab	YP_009725299
NSP4	Membrane-spanning protein containing transmembrane domain 2	500	2764–3263	Viral replication transcription complex and it helps modify ER Membranes	YP_009725300
NSP5	Proteinase and main proteinase	306	3264–3569	Cleaves at multiple distinct sites to yield mature and intermediate nonstructural proteins	YP_009725301
NSP6	Putative transmembrane domain	290	3570–3859	Induces formation of ER-derived autophagosomes As well as induces double-membrane vesicles	YP_009725302
NSP7	RNA-dependent RNA polymerase	83	3860–3942	Forms complex with NSP8 and NSP12 to yield the RNA polymerase activity of NSP8	YP_009725303
NSP8	Multimeric RNA polymerase; replicase	198	3943–4140	Makes heterodimer with NSP8 and 12	YP_009725304
NSP9	single-stranded RNA-binding viral protein	198	4141–4253	May bind to helicase	YP_009725305

NSP10	Growth-factor-like protein possessing two zinc binding motifs	139	4254–4392	Yet to be deciphered	YP_009725306
NSP11	Consists of 13 amino acids (sadaqsflngfav) andidentical to the first segment of Nsp12	13	4393–4405	Unknown	YP_009725312
NSP12	RNA-dependent RNA polymerase	932	4393–5324	Replication and methylation	YP_009725307
NSP13	RNA-dependent RNA polymerase (Pol/RdRp)	932	5325–5925	A helicase core domain that binds ATP. Zinc-binding domain is involved in replication and transcription	YP_009725307
NSP14	Proofreading Exoribonuclease domain (ExoN/nsp14)	527	5926–6452	Exoribonuclease activity acting in a 3^0-5^0 direction and N^7-guanine methyltransferase activity	YP_009725309
NSP15	EndoRNAse; nsp15-A1 and nsp15B-NendoU	346	6453–6798	Mn(2 +)-dependent endoribonuclease activity	YP_009725310.1
NSP16	2^0-O-ribose methyltransferase	298	6799–7096	Methyltransferase that mediates mRNA cap 2^0-O-ribose methylation to the 5^0-cap structure of viral mRNAs	YP_009725311

3. Accessory Proteins

The majority of proteins of SARS CoV have been characterized in detail. The proteins of SARS CoV consist of two large polyproteins: ORF1a and ORF1ab (that proteolytically cleave to form 16 nonstructural proteins), four structural proteins: spike (S), envelope (E), membrane (M), and nucleocapsid (N), and eight accessory proteins: ORF3a, ORF3b (NP_828853.1, not present in SARS CoV-2), ORF6, ORF7a, ORF7b, ORF8a, ORF8b, and ORF9b (NP_828859.1, not present

in SARS CoV-2). Although accessory proteins have been viewed as dispensable for viral replication in vitro, some have been shown to play an important role in virus-host interactions in vivo (Liu DX et al., 2014)

Evolutionary analysis indicated that some of these accessory proteins were dominating the early evolutionary trends of SARS-CoV-2 (Velazquez et al., 2020).

4. Structural proteins

4.1 Spike Protein

The envelope of corona-virion contains protruding projections from its surface called the large surface glycoproteins or spike proteins responsible for recognizing the host's receptor followed by its binding to it and fusing with its membrane. Due to the crown-shaped appearance of these projections it has been named coronavirus {corona-a crown (Latin)} (Satarker, S., & Nampoothiri, M., 2020).

SARS CoV spike mouse polyclonal antibodies potently inhibited SARS CoV-2 spike protein mediated entry into cells (Walls AC et al., 2020).

The spike protein consists of an ectodomain element, transmembrane moiety and a short intracellular C fragment. The viral ectodomain inherits two subunits, namely S1 that facilitates receptor binding and S2 that facilitates membrane fusion. It is structured like a clove built from three S1 subunits and S2 stem formed of a trimer as shown in Figure 8 (Chen et al., 2020)

Therefore, the S protein plays a crucial role in viral entry into the host cells and the structural abilities in this newly discovered SARS-CoV-2 boost its intended actions. The fact that these protruding spikes are the first point of contact with host receptors, therapeutic strategies can be applied to prevent its binding to target receptors and prevent viral entry into host cells (Satarker, S., & Nampoothiri, M., 2020).

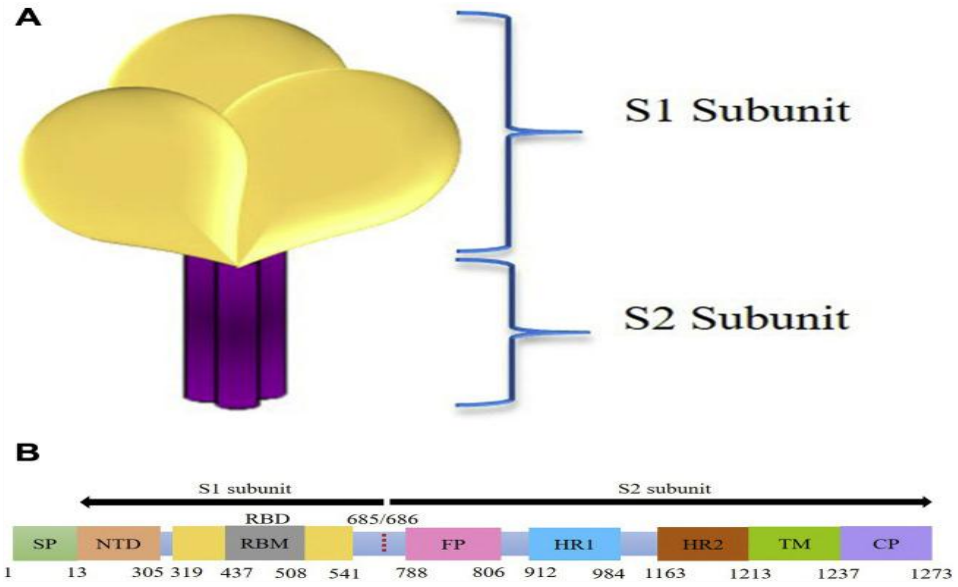


Figure 8. The structure of Spike Protein (Xia et al., 2020)

4.2 Membrane Protein

The M protein is the most abundant virion membrane protein. Aside from its role in viral assembly, the coronavirus M protein is believed to have functions in host interactions. It may be O glycosylated (groups I and III) or N glycosylated (group II). While glycosylation is not essential for viral assembly or infectivity, the glycosylation state of M protein is likely to play a role in virus-host interaction. For TGEV, the M protein has been shown to have interferogenic activity, and mutations in the M protein ectodomain that impair N glycosylation decrease this activity. For MHV, the selection of recombinant viruses with N, O, or no glycosylation demonstrated that while the glycosylation state of M protein does alter the ability to replicate in vitro, it may affect the ability to induce IFN-α in vitro and also to replicate in the liver in vivo (Weiss et al., 2005)

4.3 Nucleocapsid Protein

The nucleocapsid (N) protein of coronaviruses is a structural protein that binds directly to viral RNA and providing stability (Grunewald Me et al., 2018). Furthermore, the N protein of SARS

CoV-2 has been found to antagonize antiviral RNAi (Mu J et al., 2020). In another study, the nucleocapsid protein of SARS CoV was found to inhibit the activity of cyclin-cyclin-dependent kinase (cyclin-CDK) complex. Inactivation of the cyclin-CDK complex results in hypophosphorylation of the retinoblastoma protein and in turn inhibits S phase (genome replication) progression in the cell cycle (Surjit M et al., 2006).

The N protein genome consists of a serine (SR) rich linker region sandwiched between an N terminal domain (NTD) and a C terminal domain (CTD) (Satarker, S., & Nampoothiri, M., 2020).

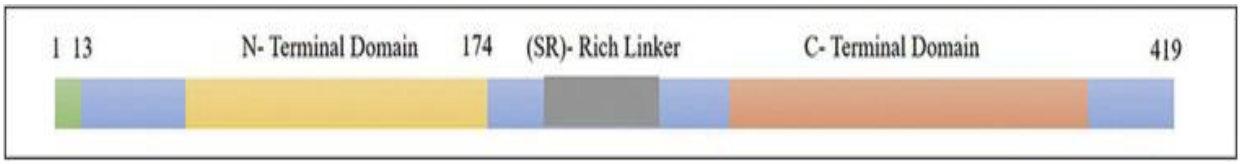


Figure 9. Genomic sequence of SARS-CoV-2 N protein (Yan et al., 2006)

4.4 Envelope Protein

The envelope protein is a small integral membrane protein in coronaviruses, which can oligomerize and create an ion channel (Verdia-Baguena C et al.,2012)

E protein, when expressed alone or when expressed together with M, forms virus-like particles. Surprisingly, it was possible to select a recombinant MHV with a deletion of the E gene. Such a recombinant MHV has low infectivity and replicates poorly, indicating that while it is nonessential for MHV, E plays an important role in production of infectious virus. The E protein of TGEV, however, is essential; disruption of the E gene within TGEV proteins is lethal. Apart from its role in virus assembly, E protein has additional functions during infection. It has recently been demonstrated that the E protein of SARS-CoV has cation-selective ion channel activity (Weiss et al., 2005)

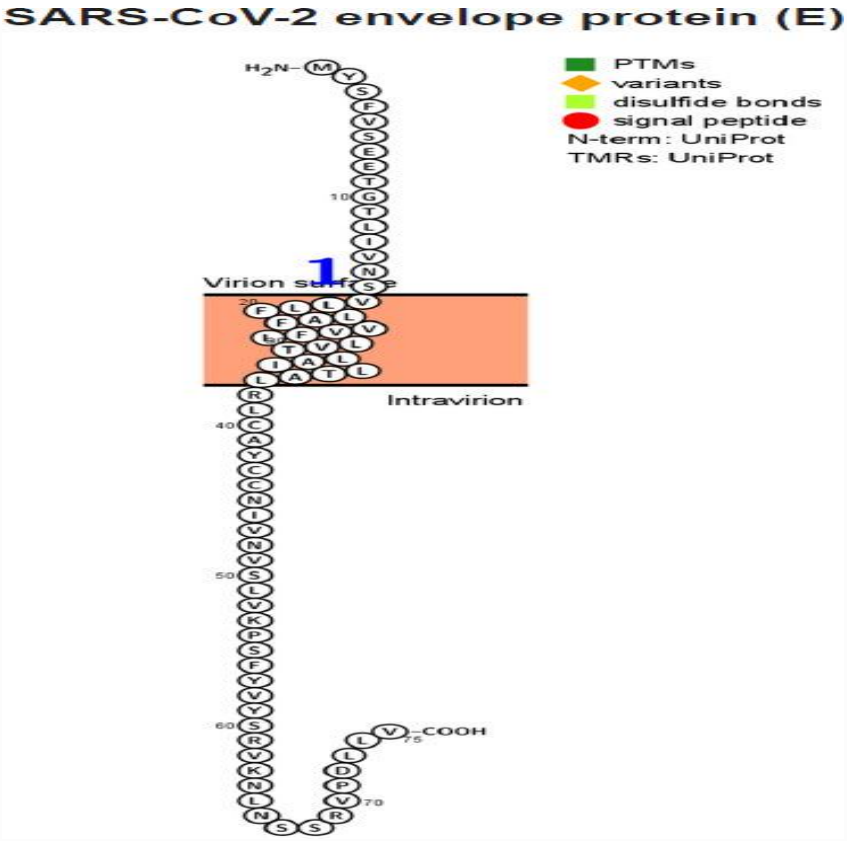


Figure 10. Snake diagram of envelope (E) protein (Thomas., 2020)

Chapter 3:

The phylogeny of SARS Cov-2

1. The phylogeny of whole genome

In a phylogenetic network analysis of 160 complete human severe acute respiratory syndrome coronavirus 2 (SARS Cov-2) genomes, we find three central variants distinguished by amino acid changes, which we have named A, B, and C, with A being the ancestral type according to the bat outgroup coronavirus. The A and C types are found in significant proportions outside East Asia, that is, in Europeans and Americans. In contrast, the B type is the most common type in East Asia, and its ancestral genome appears not to have spread outside East Asia without first mutating into derived B types, pointing to founder effects or immunological or environmental resistance against this type outside Asia. The network faithfully traces routes of infections for documented coronavirus disease 2019 (COVID-19) cases, indicating that phylogenetic networks can likewise be successfully used to help trace undocumented COVID-19 infection sources, which can then be quarantined to prevent recurrent spread of the disease worldwide (Forster et al., 2020)

1.1 Phylogenetic analysis of SARS-CoV-2 genomes in Turkey

COVID-19 has effectively spread worldwide. As of May 2020, Turkey is among the top ten countries with the most cases. A comprehensive genomic characterization of the virus isolates in Turkey is yet to be carried out. Here, we built a phylogenetic tree with globally obtained 15,277 severe acute respiratory syndrome coronavirus 2 (SARS CoV-2) genomes. They identified the subtypes based on the phylogenetic clustering in comparison with the previously annotated classifications. They performed a phylogenetic analysis of the first 30 SARS CoV-2 genomes isolated and sequenced in Turkey. They suggest that the first introduction of the virus to the country is earlier than the first reported case of infection. Virus genomes isolated from Turkey are dispersed among most types in the phylogenetic tree. They find 2 of the seventeen subclusters enriched with the isolates of Turkey, which likely have spread expansively in the country. Finally, they traced virus genomes based on their phylogenetic placements. This analysis suggested multiple independent international introductions of the virus and revealed a hub for the inland transmission. They released a web application to track the global and interprovincial virus spread of the isolates from Turkey in comparison to thousands of genomes worldwide (Adebali et al., 2020).

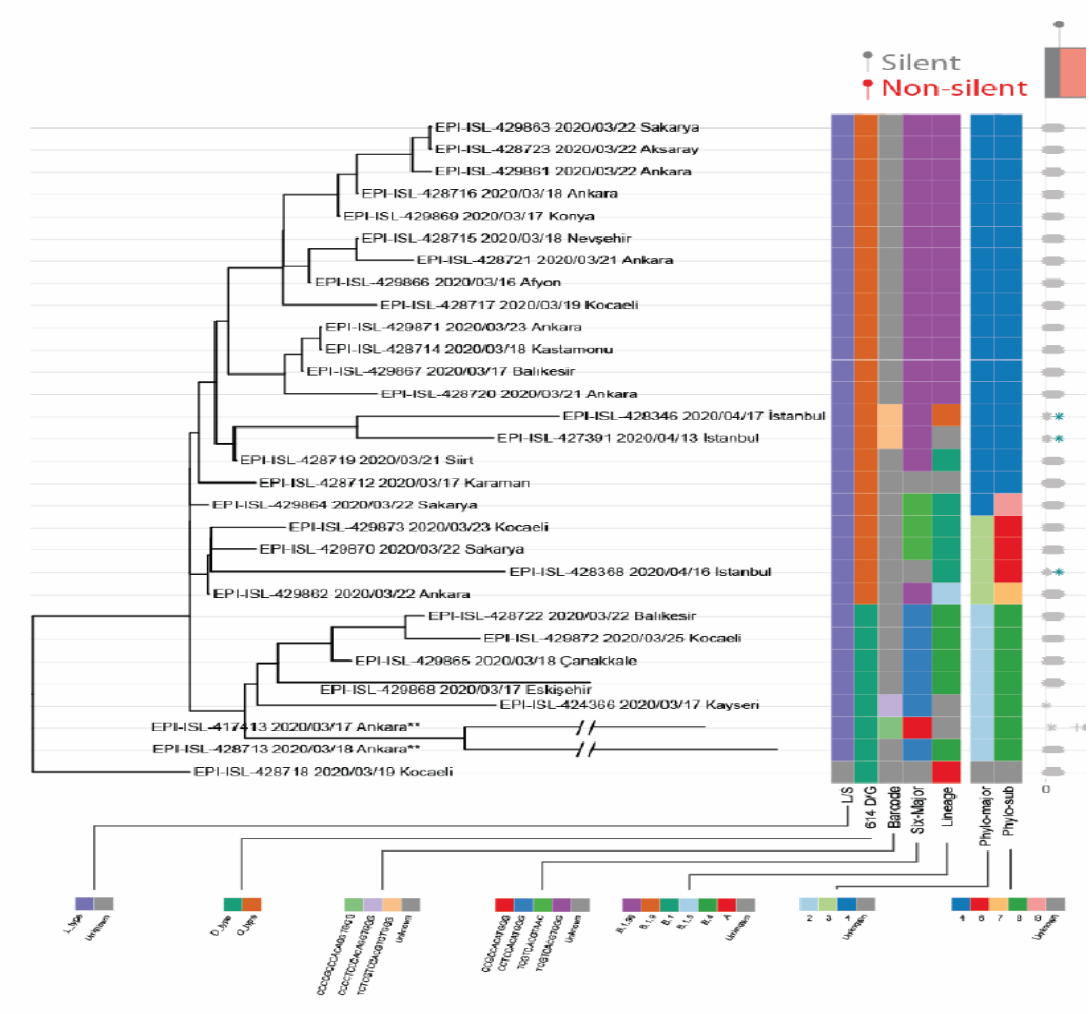


Figure 11. Phylogenetic tree of SARS-CoV-2 samples sequenced in Turkey (Adebali et al., 2020)

1.2. Phylogenetic analysis of SARS-CoV-2 genomes in Chile

The current pandemic caused by the new coronavirus is a worldwide public health concern. To aboard this emergency, and like never before, scientific groups around the world have been working in a fast and coordinated way to get the maximum of information about this virus when it has been almost 3 months since the first cases were detected in Wuhan province in China. The complete genome sequences of around 450 isolates are available, and studies about similarities and differences among them and with the close related viruses that caused similar epidemics in this century. In this work, they studied the complete genome of the first four cases of the new

coronavirus disease in Chile, from patients who traveled to Europe and Southeast Asia. Their findings reveal at least two different viral variants entries to Chilean territory, coming from Europe and Asia. They also sub- classified the isolates into variants according to punctual mutations in the genome. Their work contributes to global information about transmission dynamics and the importance to take control measures to stop the spread of the infection (Andres et al., 2020).

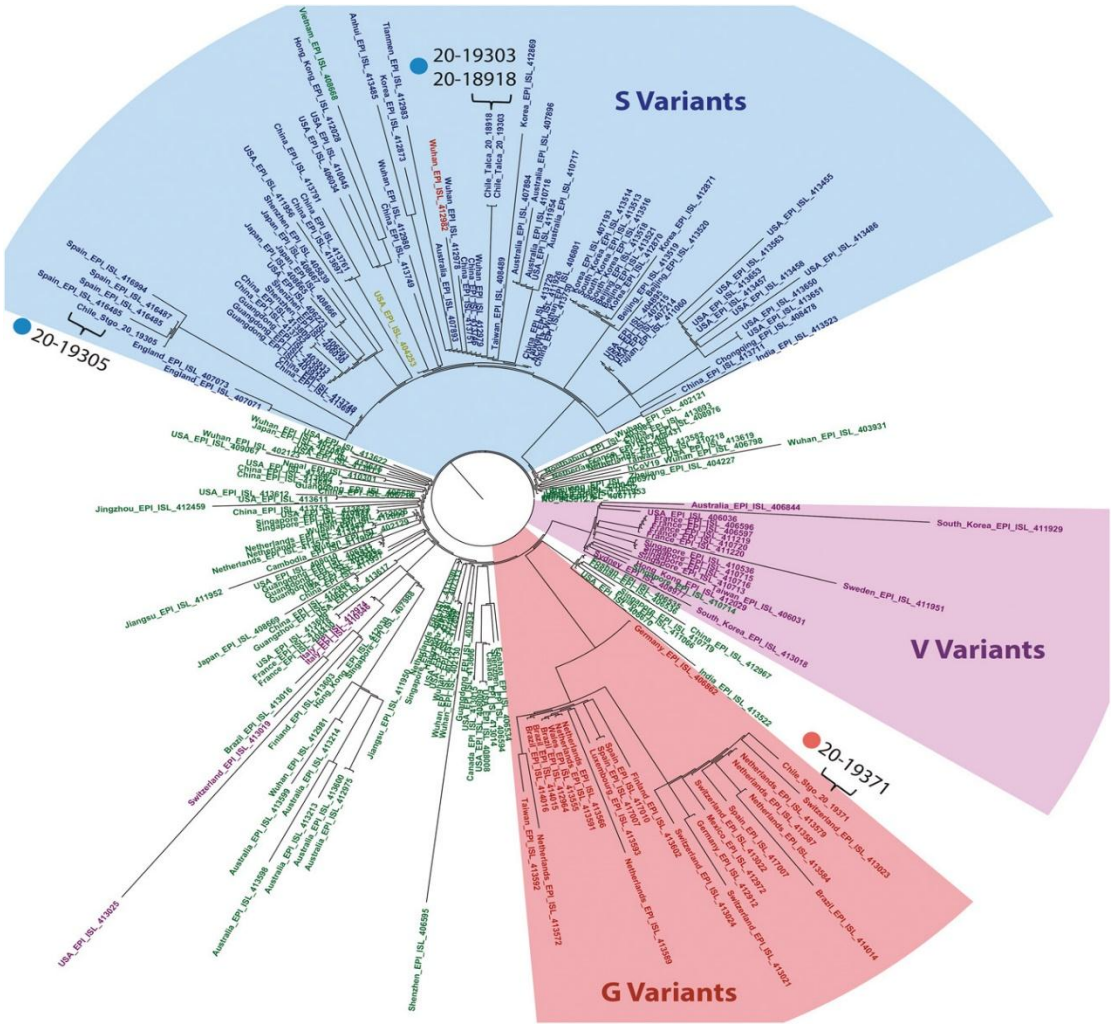


Figure 12. Phylogenetic tree with Maximum Composite Likelihood distance, representing 222 complete genomes including the four Chilean isolates. The name of the isolates was colored according to the variants as follows: S (blue), G (red), V (purple), unclassified variants (green), and the main clades were highlighted. Chilean strains are marked over the cladogram in the S and G variant clades (Andres et al., 2020)

1.3. Phylogenetic analysis of SARS-CoV-2 genomes in Netherlands

In late December 2019, a cluster of cases of pneumonia of unknown etiology were reported linked to a market in Wuhan, China. The causative agent was identified as the species severe acute respiratory syndrome-related coronavirus and was named SARS-CoV-2. By 16 April the virus had spread to 185 different countries, infected over 2,000,000 people and resulted in over 130,000 deaths. In the Netherlands, the first case of SARS-CoV-2 was notified on 27 February. The outbreak started with several different introductory events from Italy, Austria, Germany and France followed by local amplification in, and later also outside, the south of the Netherlands. The combination of near to real-time whole-genome sequence analysis and epidemiology resulted in reliable assessments of the extent of SARS-CoV-2 transmission in the community, facilitating early decision-making to control local transmission of SARS-CoV-2 in the Netherlands. They demonstrate how these data were generated and analyzed, and how SARS-CoV-2 whole-genome sequencing, in combination with epidemiological data, was used to inform public health decision-making in the Netherlands (Oude et al., 2020)

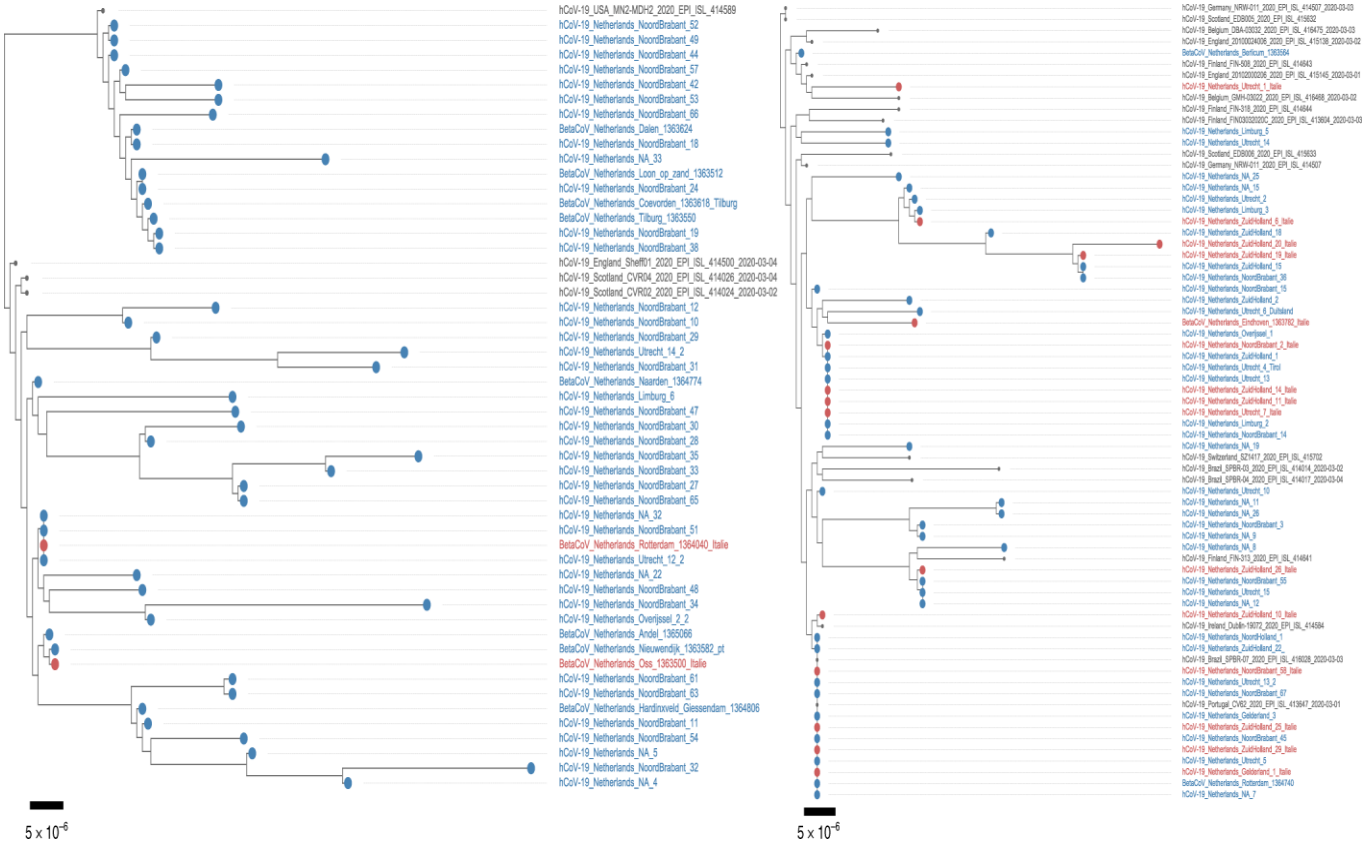


Figure 13. Phylogenetic analysis of SARS CoV-2 emergence in the Netherlands (Oude et al., 2020)

Part II

Experimental research

Material and Methods

1. Material and Methods

1.1 Introduction to databases

- **Presentation of GISAID :**

GISAID is a global science initiative and primary source established in 2008 that provides open access to genomic data of influenza viruses and the coronavirus responsible for the COVID-19 pandemic. On January 10, 2020, the first whole-genome sequences of SARS-cov2 were made available on GISAID, which enabled global responses to the pandemic, including the development of the first vaccines and diagnostic tests to detect SARS-CoV-2. As of Mai 12 2022 deadline of our phylogenetic analysis, the GISAID database (https://www.gisaid.org/) had compiled 11,134,425 coronavirus genomes from different geographies.

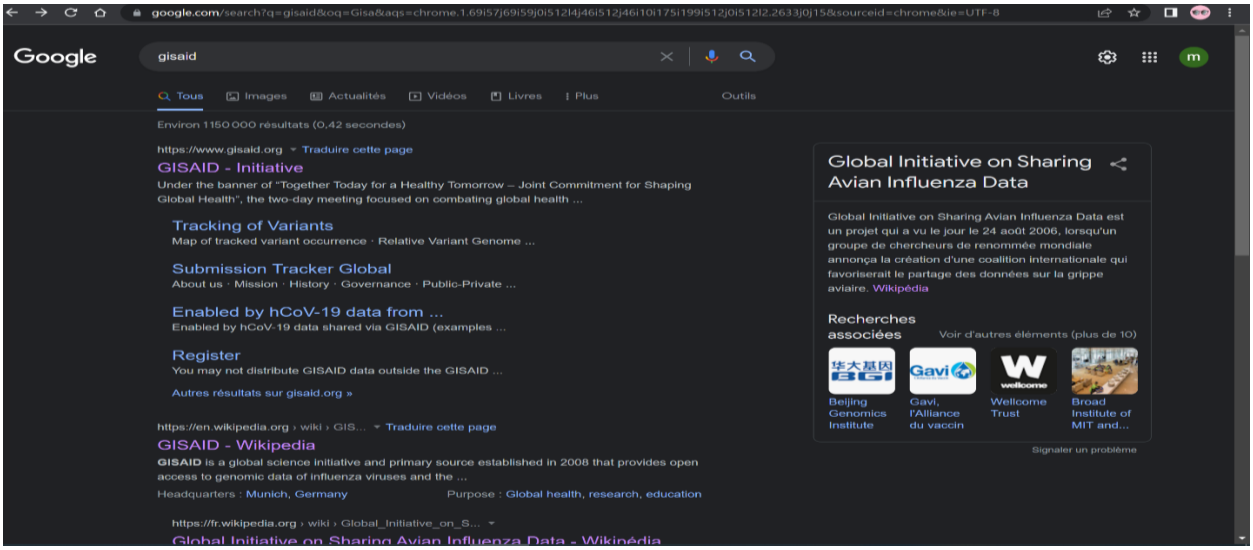
1.2. Collection of genomic sequences from databases

SARS-CoV-2 genome sequences were assembled from GISAID (https://www.gisaid.org/) in FASTA format.

A total of 264 sequences of SARS-CoV-2from Algeria have been found.

1.3 Steps of research on GISAID

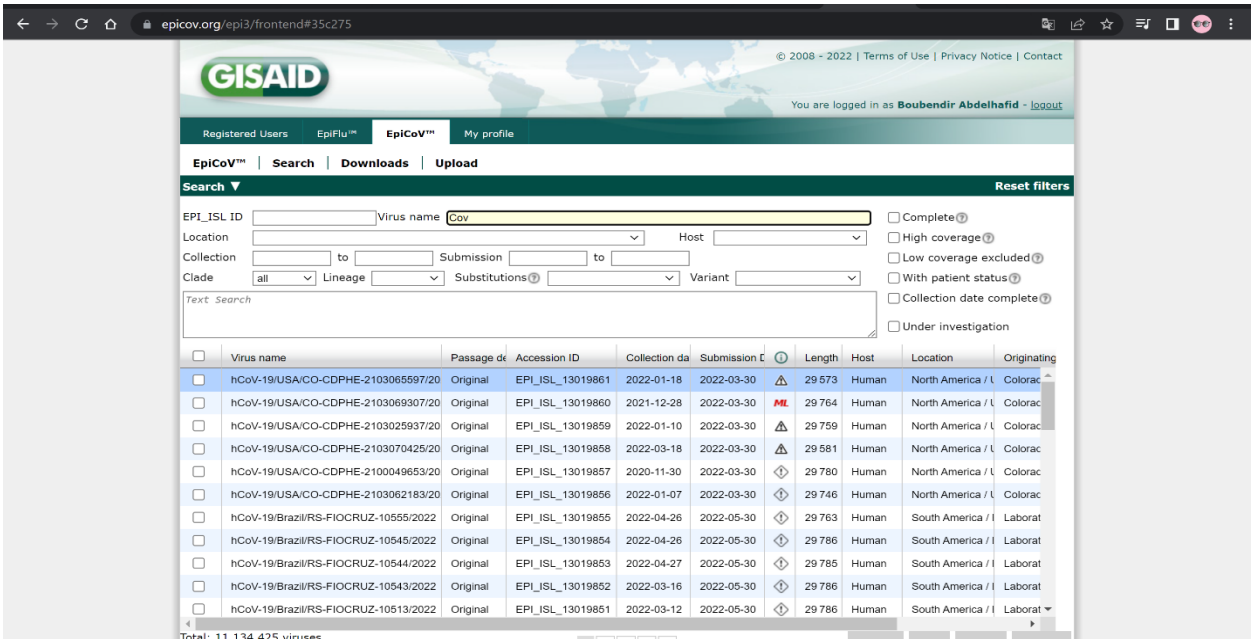
The following illustrations summarize the steps in the search for genomic sequences on GISAID:



Step 1. Open Google browser page (https://www.gisaid.org/)

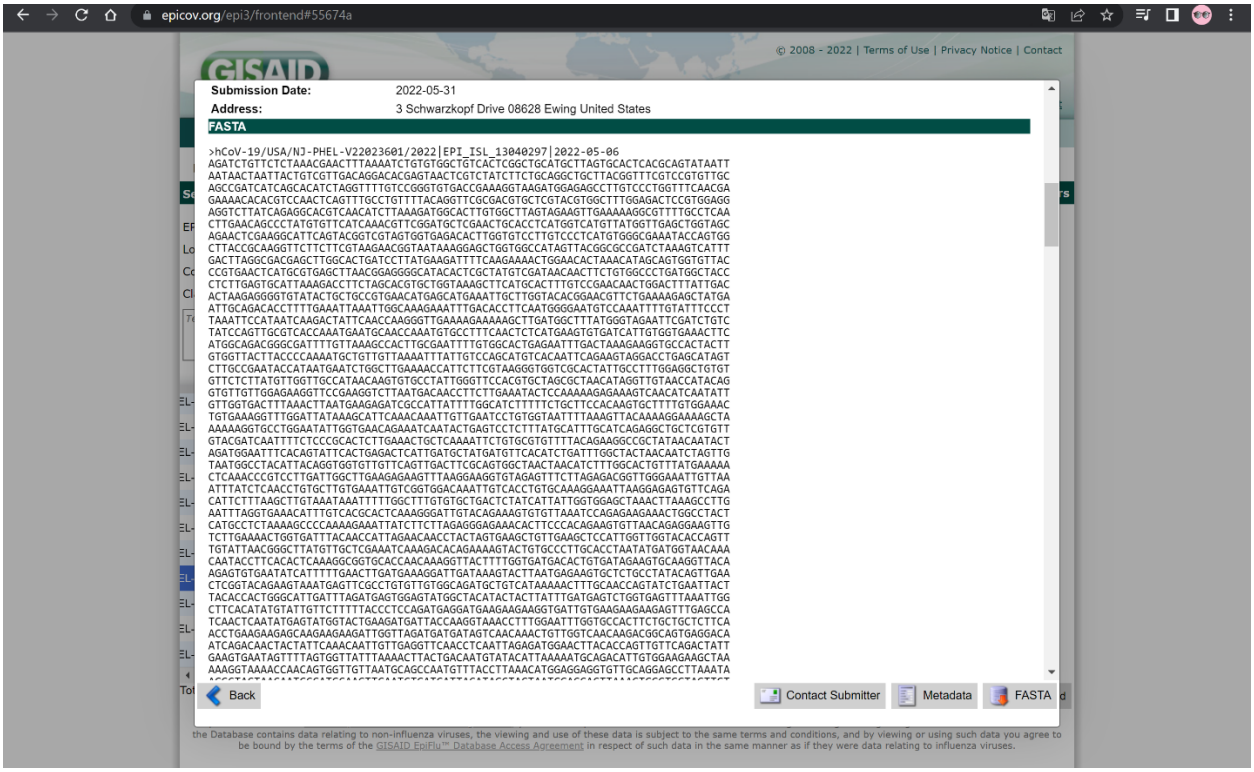
Step 2. Home page of GISAID

- After Login and entering the keywords (which is cov), all the genomic sequences of SARS-CoV-2 are loaded into each page containing 50 sequences.



Step 3. SARS-CoV-2 genome search results on GISAID

Step 4. The choice of whole genome sequences.



Step 5. Download SARS-CoV-2 Full Genome Sequences in Format FASTA by clicking on "FASTA"

1.4 The stages of phylogenetic tree construction

1.4.1 Introducing MEGA 7

We present the latest version of the Molecular Evolutionary Genetics Analysis (MEGA) software, which contains many sophisticated methods and tools for phylogenomics and phylomedicine. In this major upgrade, M EGA has been optimized for use on 64-bit computing systems for analyzing larger datasets. Researchers can now explore and analyze tens of thousands of sequences in MEGA. The new version also provides an advanced wizard for building timetrees and includes a new functionality to automatically predict gene duplication events in gene family trees. The 64-bit M EGA is made available in two interfaces: graphical and command line. The graphical user interface (GUI) is a native Microsoft Windows application that can also be used on Mac OS X. The command line M EGA is available as native applications for Windows, Linux, and Mac OS X. They are intended for use in high-throughput and scripted analysis. Both versions are available from www.megasoftware.net free of charge.

Molecular Evolutionary Genetics Analysis (MEGA) software is now being applied to increasingly bigger datasets. This necessitated technological advancement of the computation core and the user interface of MEGA. Researchers also need to conduct high-throughput and scripted analyses on their operating system of choice, which requires that MEGA be available in native cross-platform implementation. We have advanced the MEGA software suite to address these needs of researchers performing comparative analyses of DNA and protein sequences of increasing larger datasets.

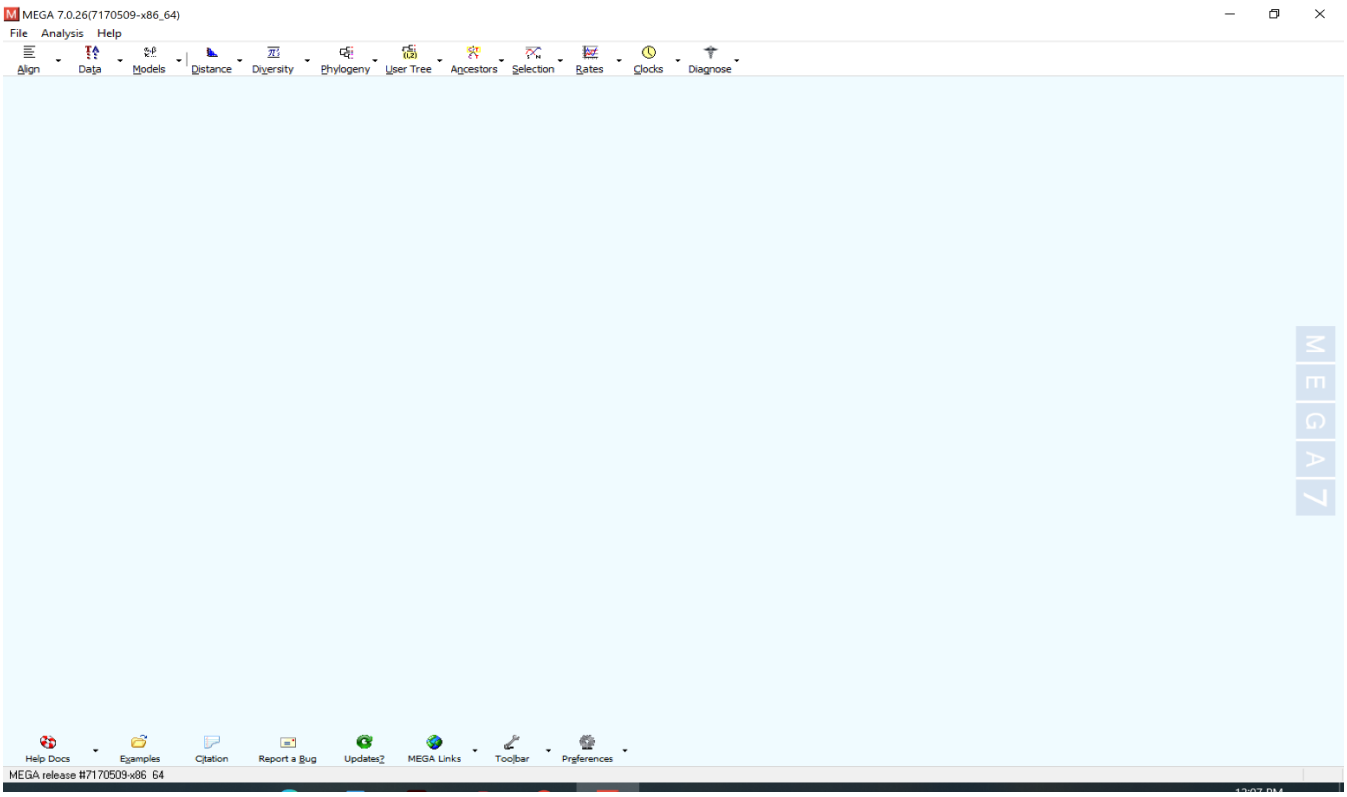
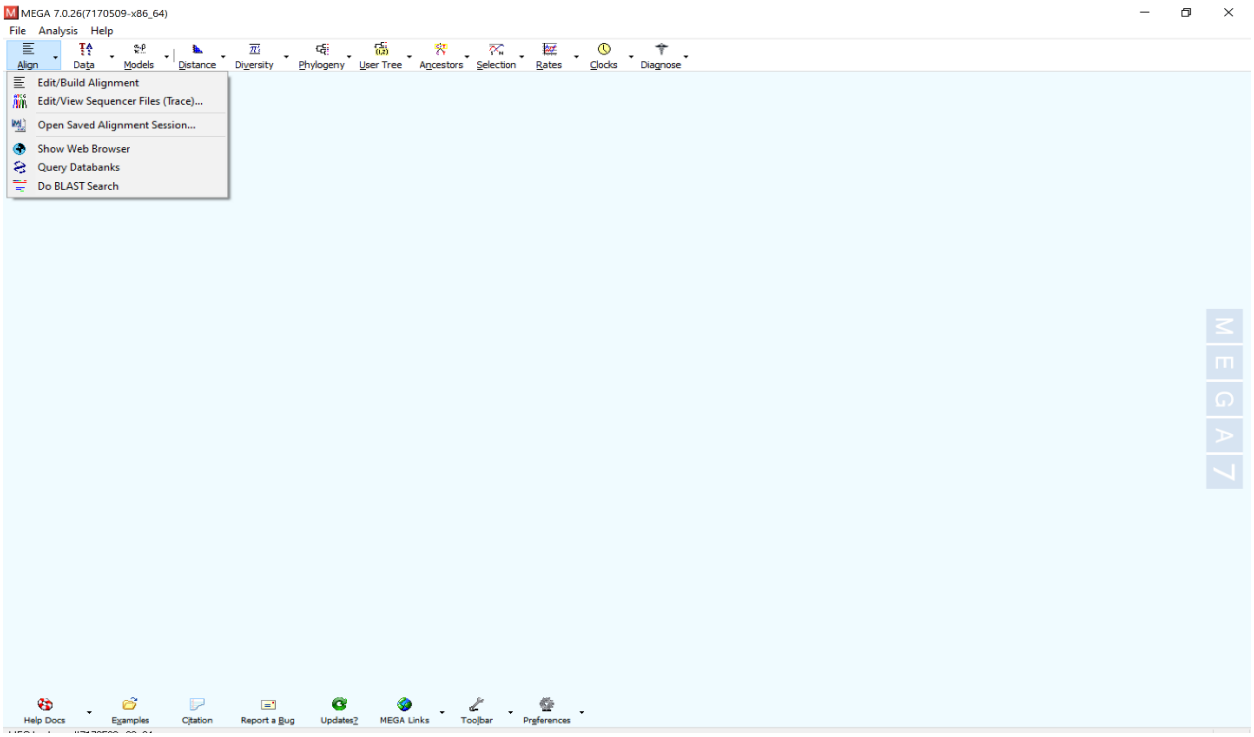


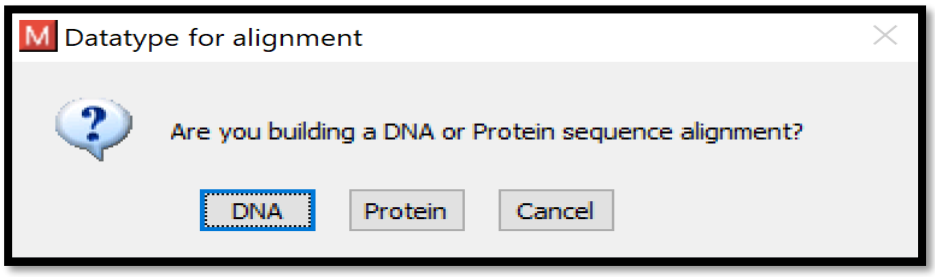
Figure 14. The main window of MEGA 7

1.4.2. Sequence alignment

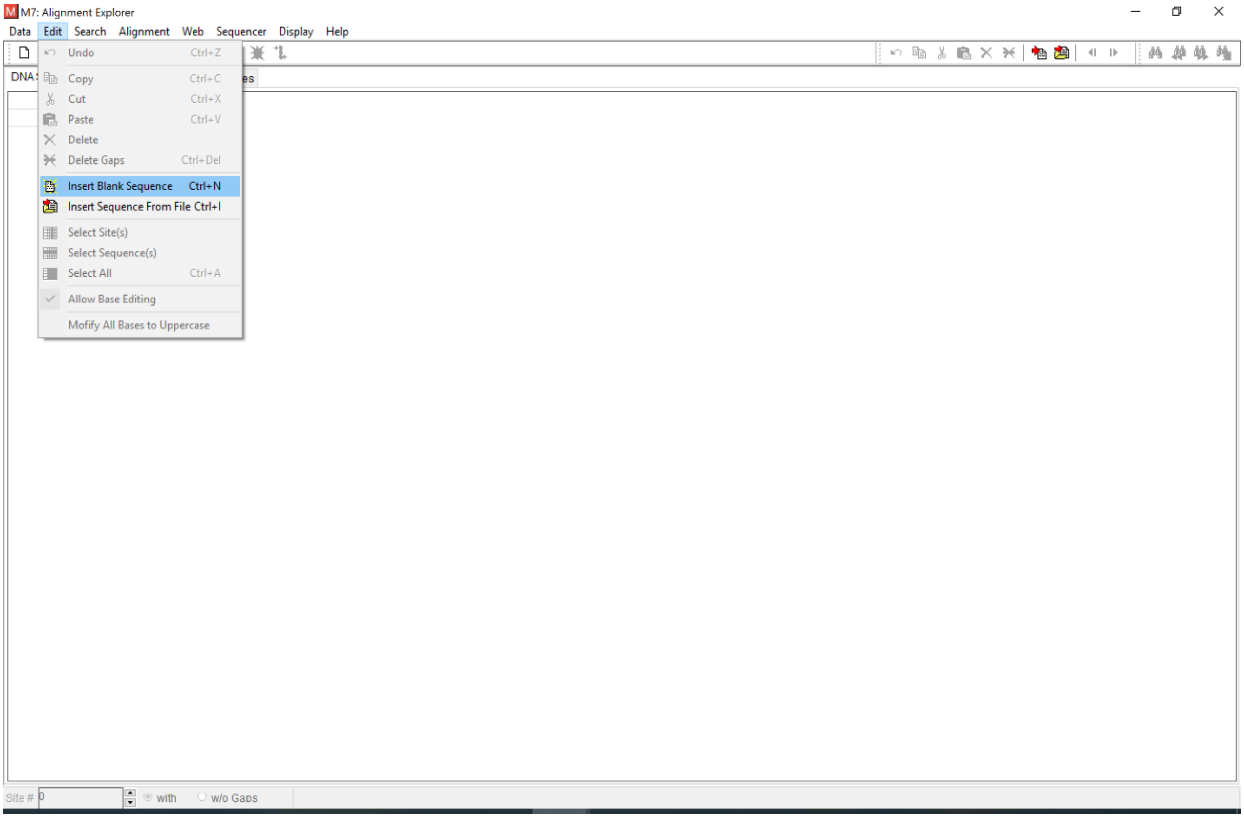


Step 6. Open the MEGA 7 program, click on Align and select Edit /Build Alignment.

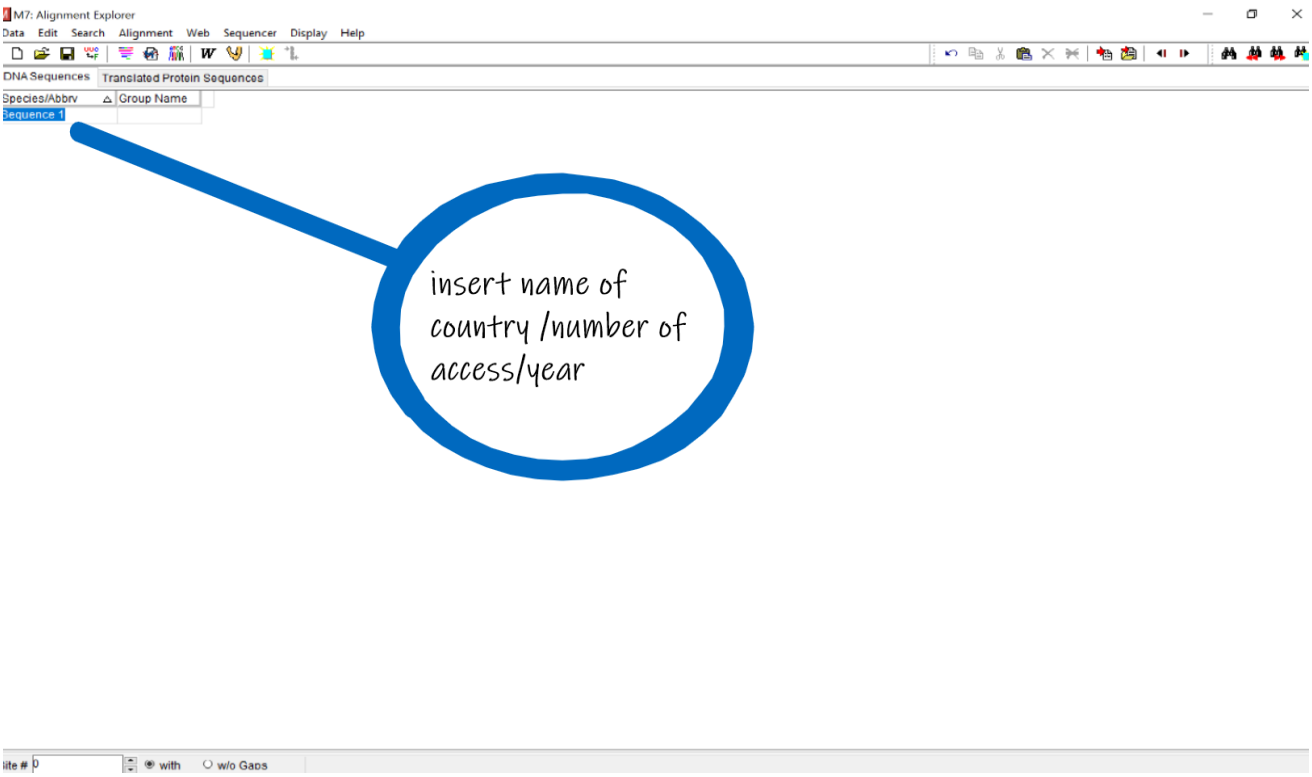
Step 7. Select create a new alignment and click on OK**.**



Step 8. Click on DNA.

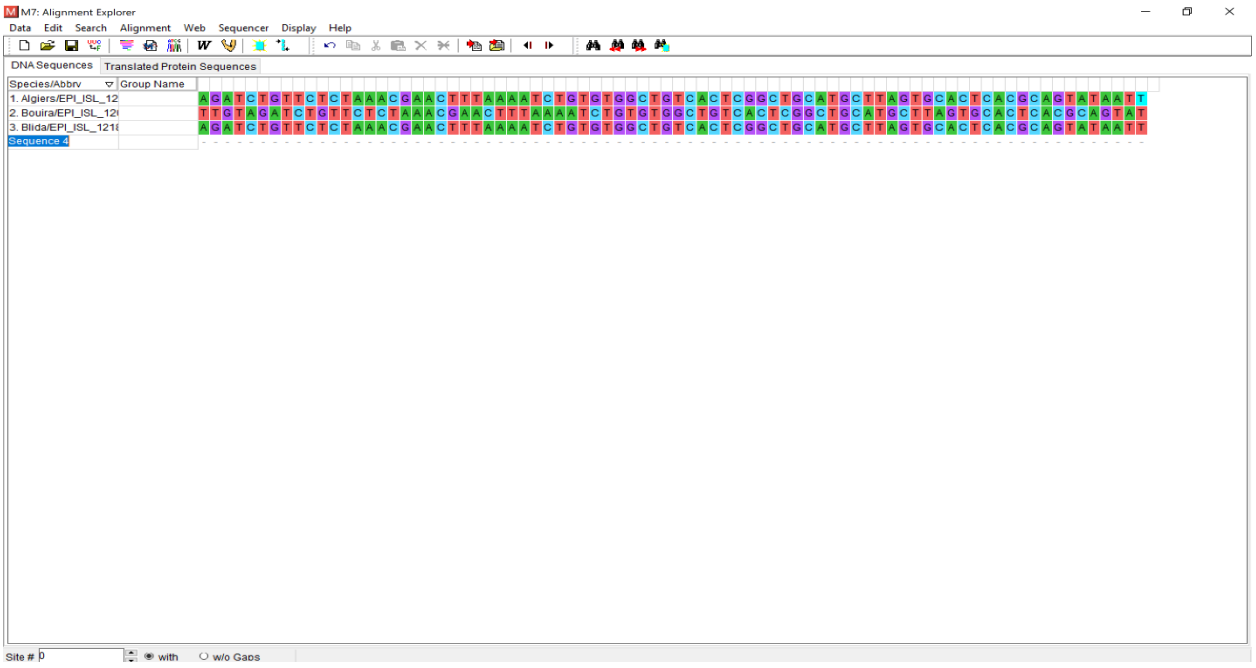


Step 9. Click on Edit then “insert blank sequence.



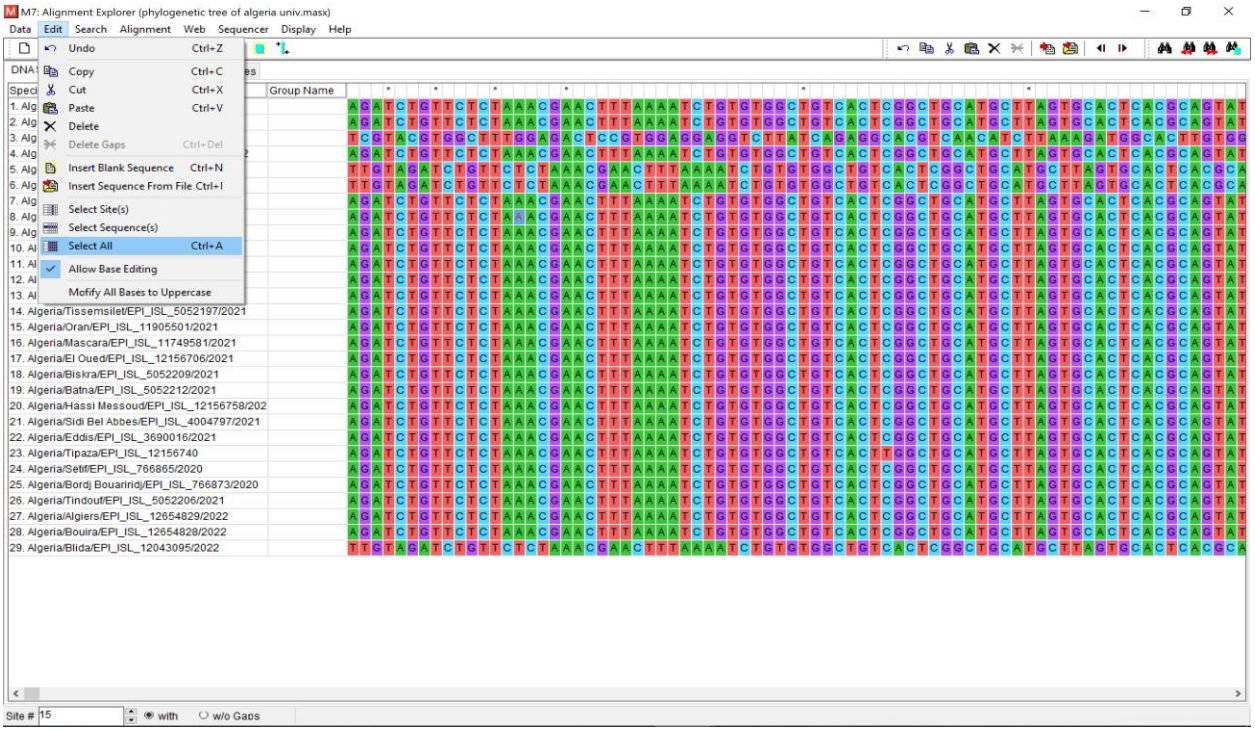
Step 10. Click on sequence and write country name/access number/year.

- Then copy paste the sequence .

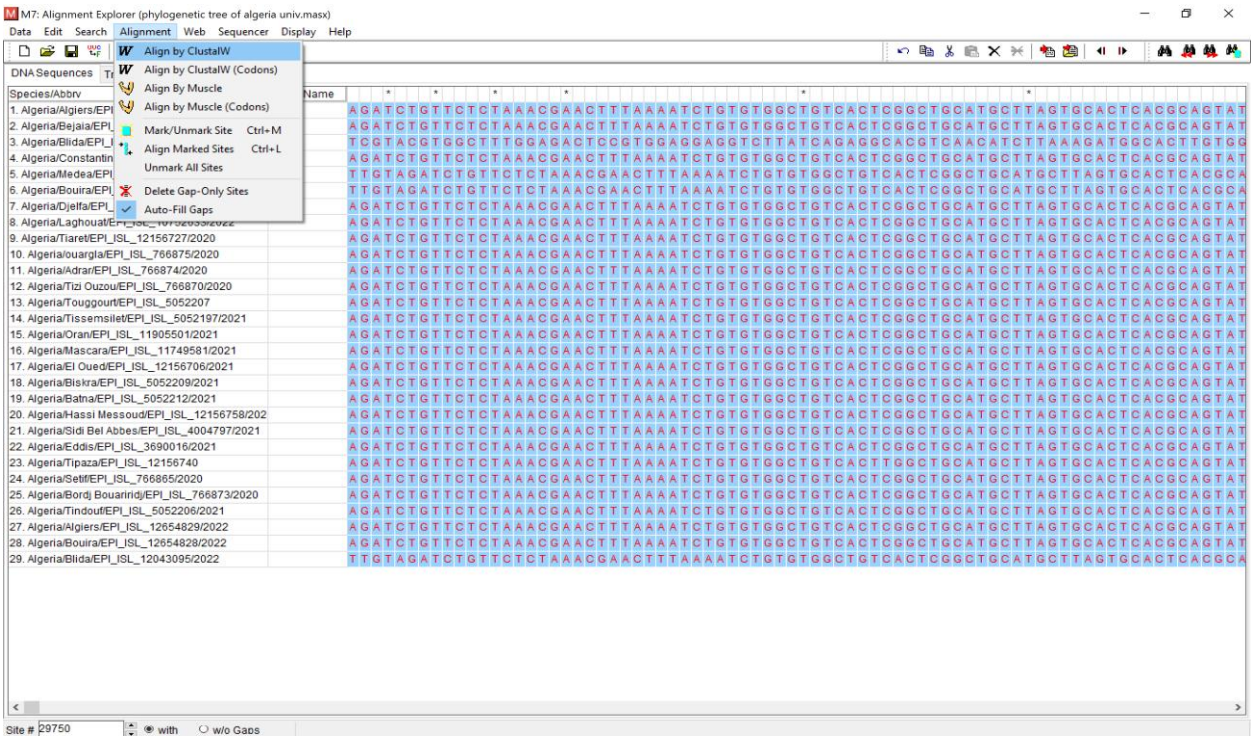


Step 11. Insertion of 29 SARS-CoV-2 DNA sequences.

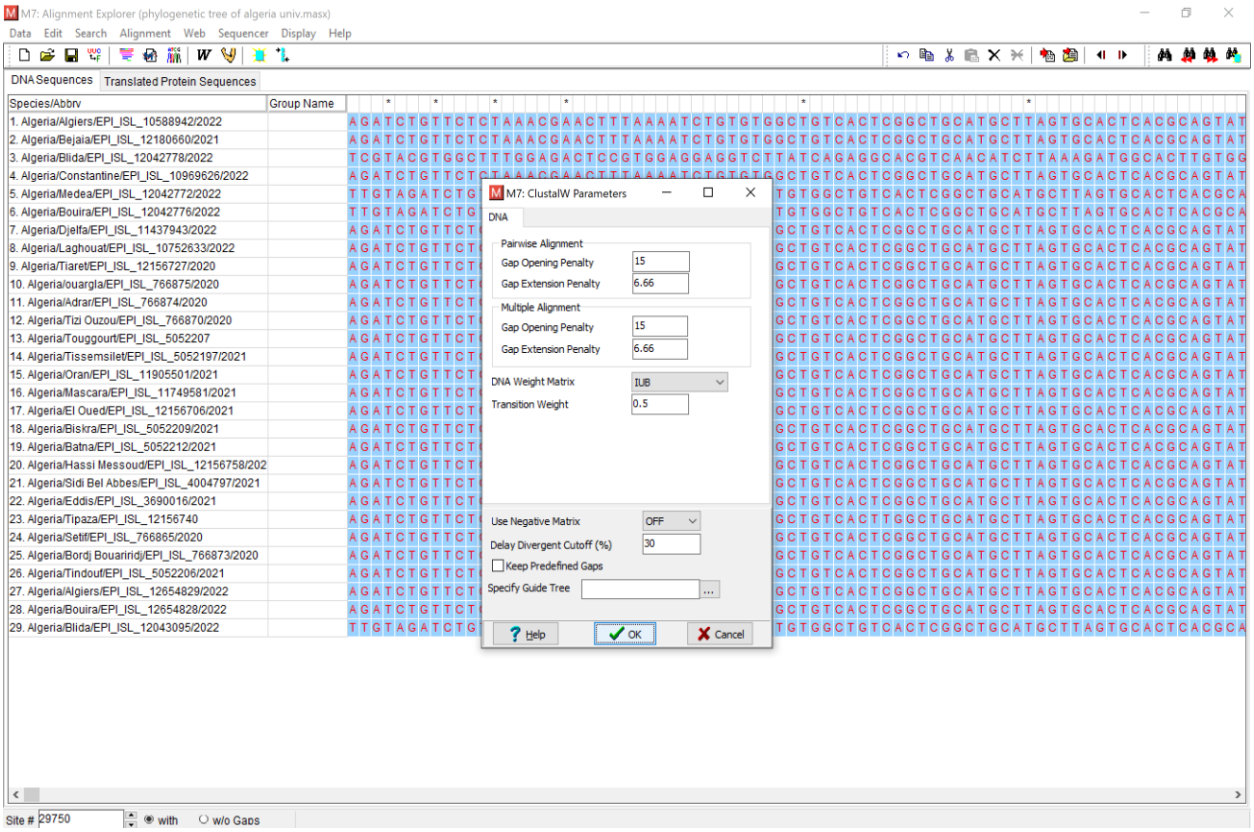
Step 12. Insertion result of the 29 Virus DNA samples obtained from GISAID.



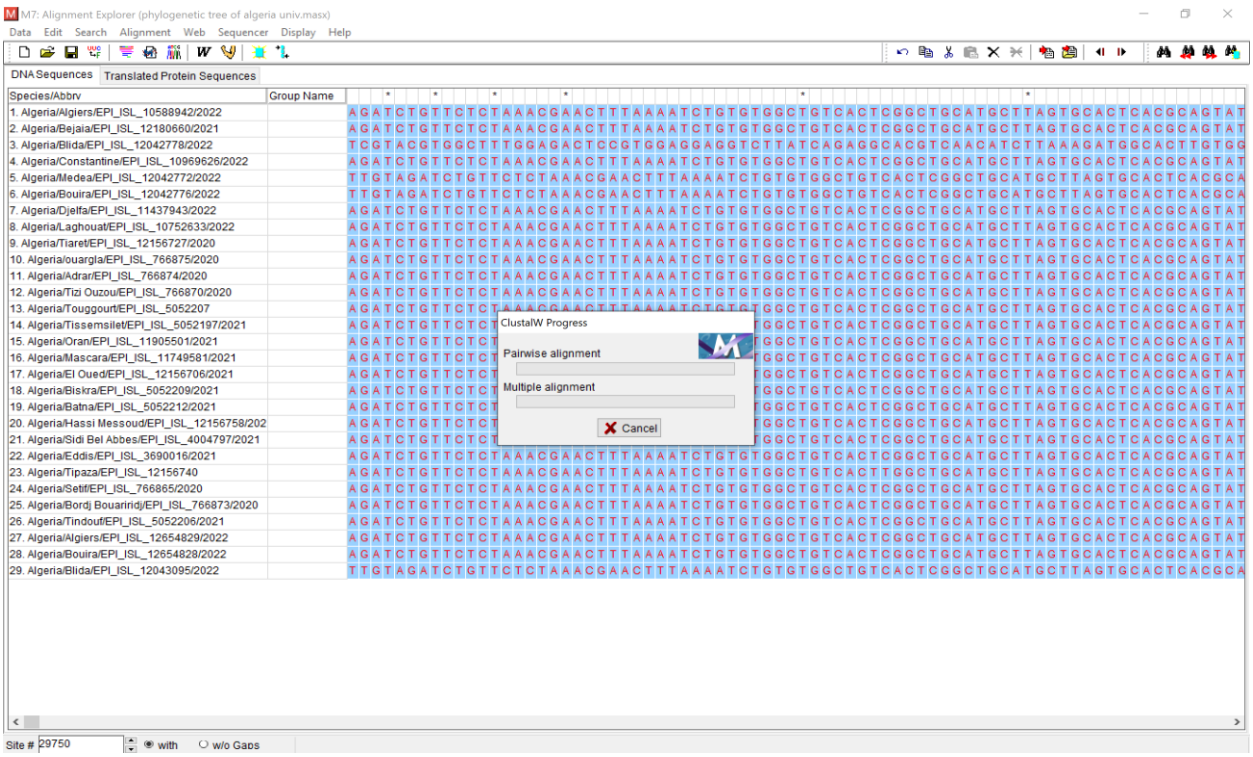
Step 13. Select all sequences: Edit / select all.



Step 14. Activate Alignment: Alignment/Align by Clustaw.



Step15. Confirmation the activation of Alignment: OK.

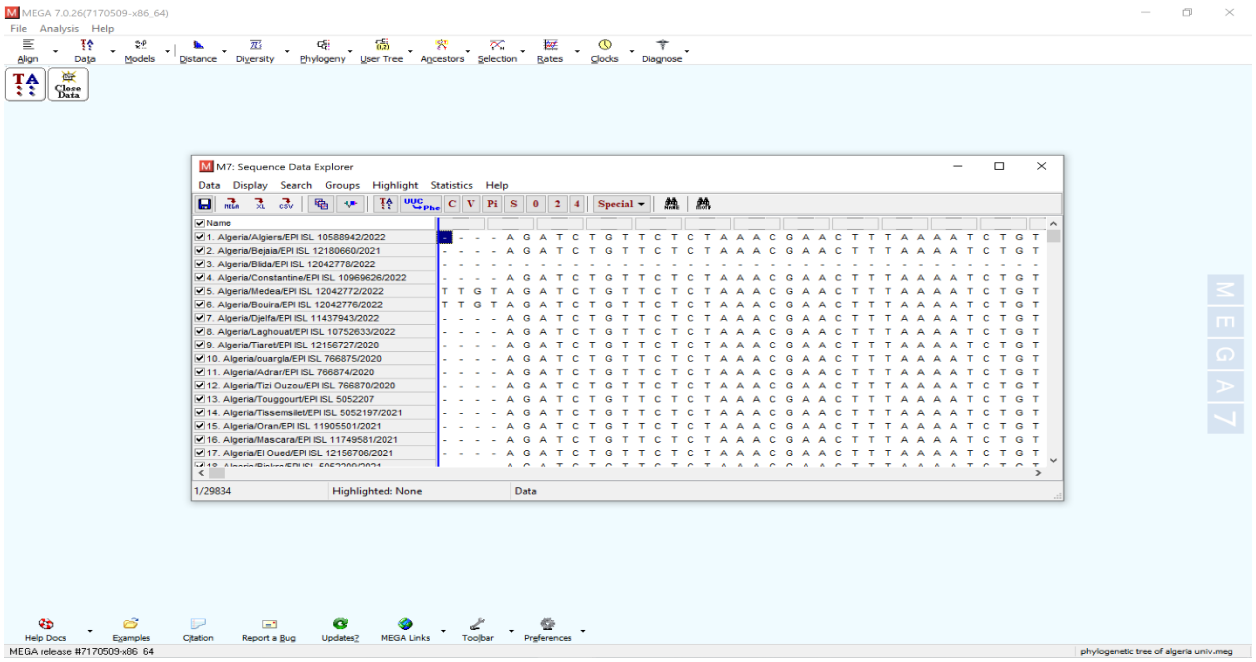


Step 16. You will have the result of the alignment after more than 7 hours.

Step 17. Result of alignment.

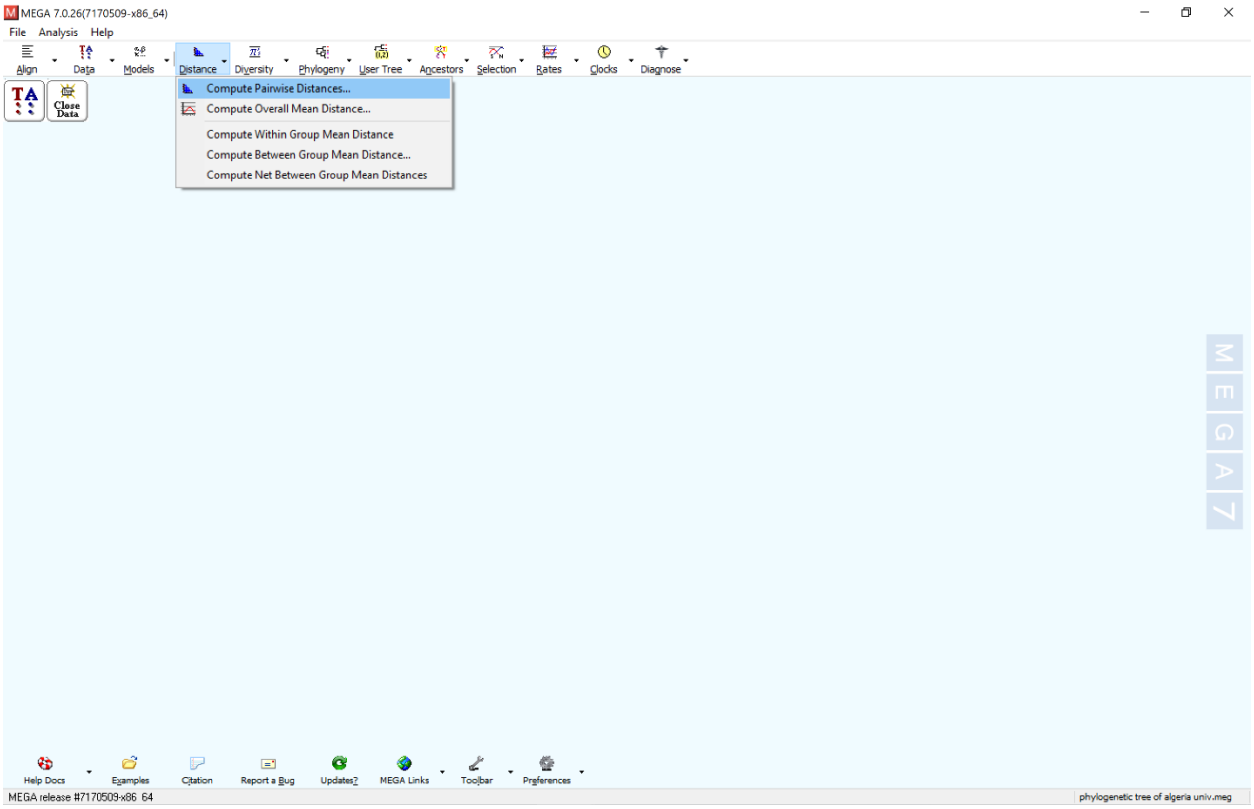
➢ The appearance of the GAPs confirms the achievement of the alignment.

Step 18. Save the alignment in MEGA format: Data_Export Alignment_MEGA Format.

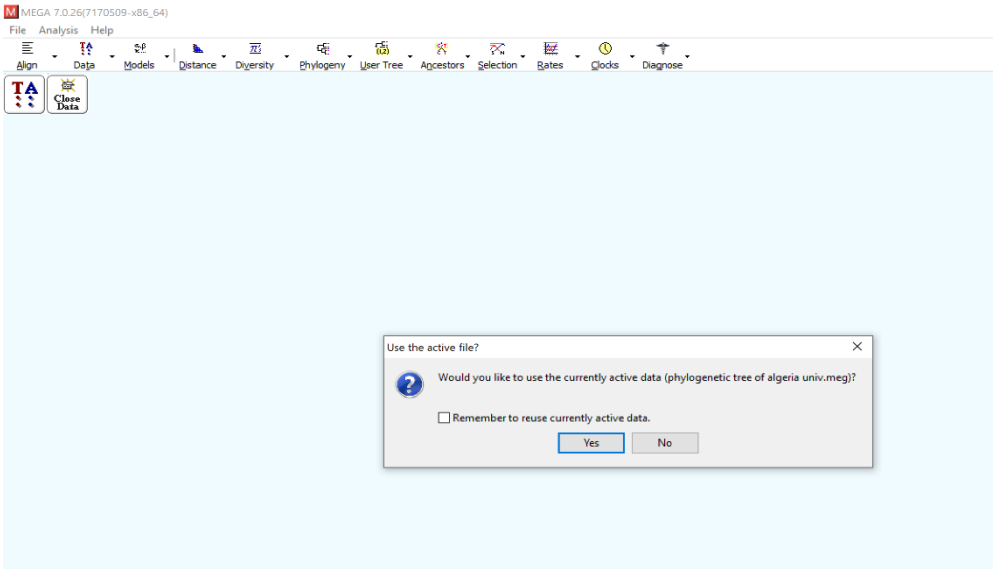


Step 19. Finally, open the MEGA file and view your alignment.

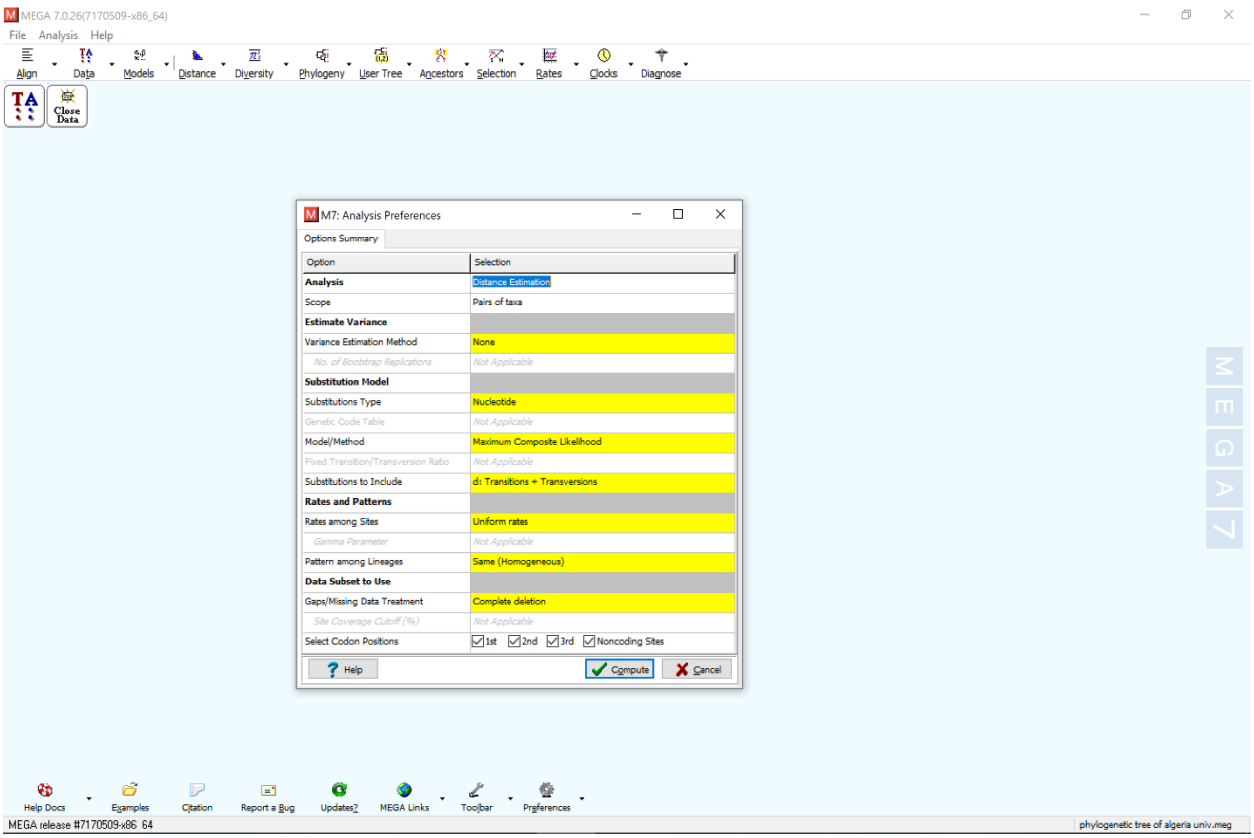
1.4.3. Estimating Scalable Distances Using Pairwise Distance



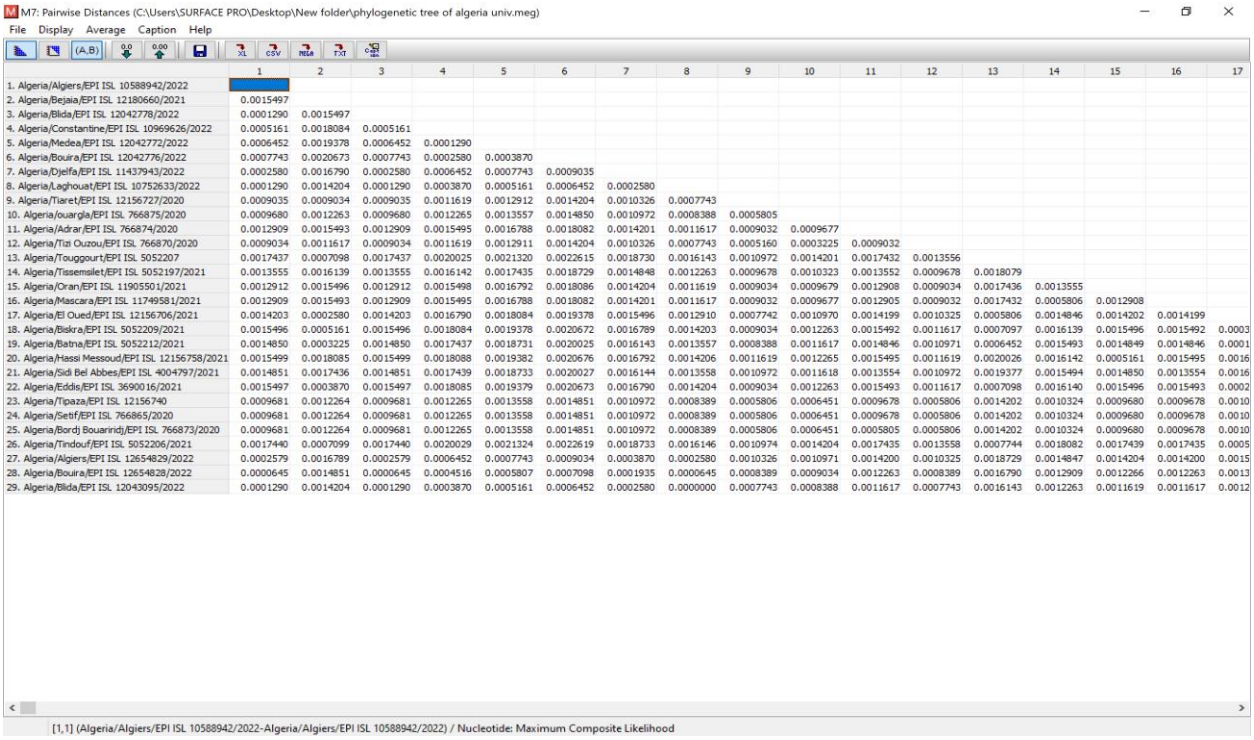
Step 20. Activate the Matrix Distance Explorer program and choose the Compute action Pairwise Distance.



Step 21. Confirm active data usage: Yes.

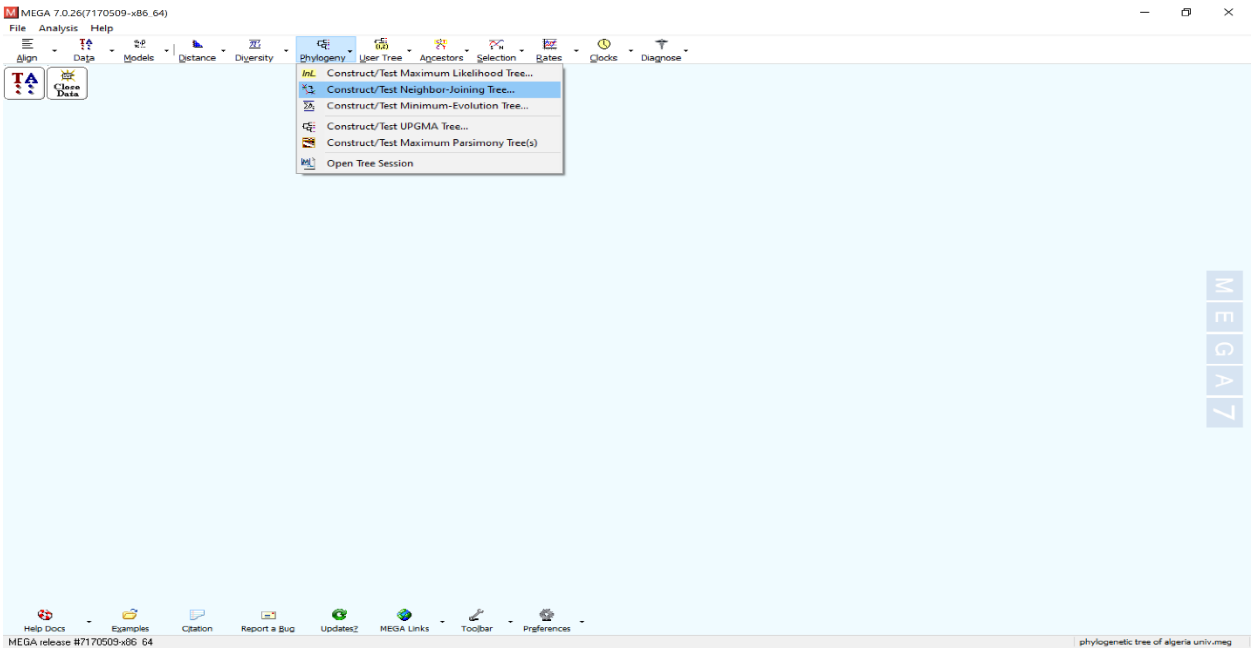


Step 22. Click on Compute to launch the distance matrix.

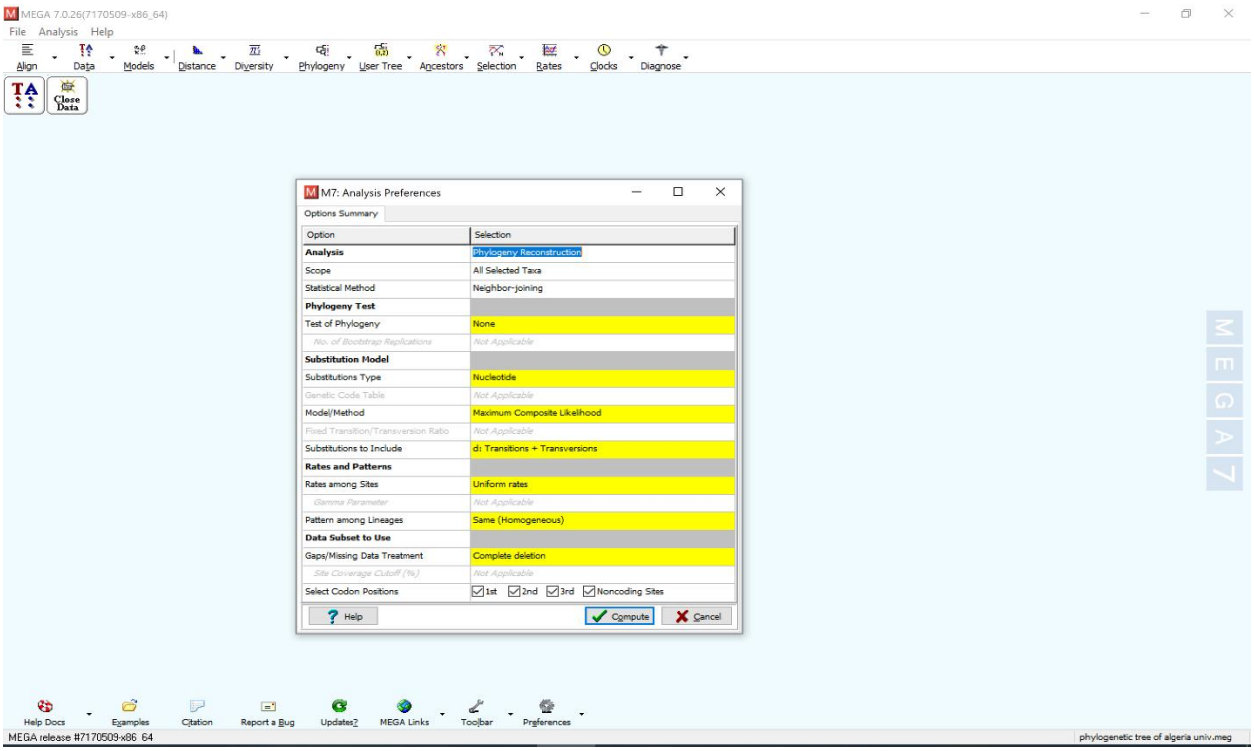


Step 23. Finally, you get the distance matrix.

1.4.4. Construction of the phylogenetic tree

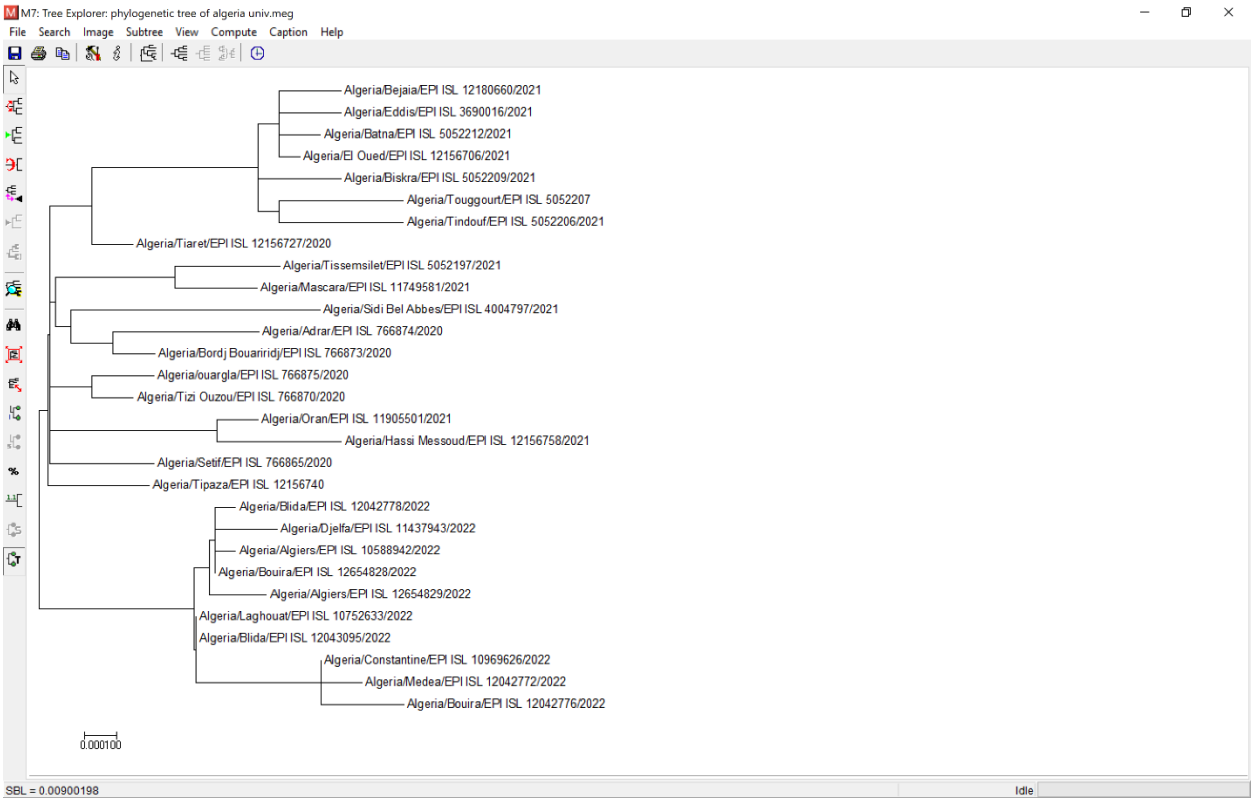


Step 24. Activate the Tree Explorer program and select the construction method of the NJ tree or others.



Step 25. Click on Compute.

➢ The resulting phylogenetic tree is shown below.



Step 26. The phylogenetic tree.

Results and discussion

2. Results and discussion

2.1. The study of SARS Cov-2 in different regions of Algeria

The table 4 below represents the number of sequences in Algeria taken from GISAID database during the month of May. A total of 264 isolates found from 27 Cities.

Table 4. Number of SARS Cov-2 isolated in Algeria from GISAID database.

Wilaya	Number of sequences
Algiers	120
Bouira	38
Blida	25
Laghouat	23
Medea	8
Ouargla	8
El Oued	6
Bejaia	5
Constantine	4
Sidi Bel Abbes	4
Oran	3
Tipaza	2
Tissemsilet	2
Djelfa	2
Ain Saleh	2
Hassi Messoud	1
Tiaret	1
Mascara	1
Batna	1
Biskra	1
Touggourt	1
Tindouf	1
Eddis	1
Adrar	1
Bordj Bouariridj	1
Tizi Ouzou	1
Setif	1
Total	**264**

A total of 27 genome sequences were obtained from GISAID, collected on May 2022 and the table below summarized some details of the sequences used in this study representing by region/ accession number/ date/ reference.

Table 5. Genomic sequences from different regions of Algeria (Region/Accession number/Date/Reference).

N°	Regions	Accession number	Date	Reference
1	Algiers	EPI ISL 10588942	2022-01-31	GISAID.
2	Bejaia	EPI ISL 12180660	2021-09-18	GISAID.
3	Blida	EPI ISL 12042778	2022-03-02	GISAID.
4	Batna	EPI ISL 5052212	2021-09-01	GISAID.
5	Adrar	EPI ISL 766874	2020-06-21	GISAID.
6	Blida	EPI ISL 12043095	2022-02-06	GISAID.
7	Algiers	EPI ISL 12654829	2022-03-27	GISAID.
8	Bouira	EPI ISL 12654828	2022-03-22	GISAID.
9	Bouira	EPI ISL 12042776	2022-02-27	GISAID.
10	Bordj Bouariridj	EPI ISL 766873	2020-06-22	GISAID.
11	Biskra	EPI ISL 5052209	2021-08-14	GISAID.
12	Constantine	EPI ISL 10969626	2022-01-06	GISAID.
13	Djelfa	EPI ISL 11437943	2022-02-19	GISAID.
14	Eddis	EPI ISL 3690016	2021-07-28	GISAID.
15	El Oued	EPI ISL 12156706	2021-07-29	GISAID.
16	Hassi messaoud	EPI ISL 12156758	2021-07-06	GISAID.
17	Laghouat	EPI ISL 10752633	2022-02-05	GISAID.
18	Medea	EPI ISL 12042772	2022-02-21	GISAID.
19	Mascara	EPI ISL 11749581	2021-05-17	GISAID.
20	Oran	EPI ISL 11905501	2021-05-03	GISAID.
21	Ouargla	EPI ISL 766875	2020-06-17	GISAID.
22	Sidi bel Abbes	EPI ISL 4004797	2021-03-17	GISAID.
23	Setif	EPI ISL 766865	2020-06-19	GISAID.

24	Tipaza	EPI ISL 12156740	2020-12-24	GISAID.
25	Tindouf	EPI ISL 5052206	2021-08-12	GISAID.
26	Tizi Ouzou	EPI ISL 766870	2020-06-15	GISAID.
27	Tissemsilet	EPI ISL 5052197	2021-07-08	GISAID.
28	Touggourt	EPI ISL 5052207	2021-07-30	GISAID.
29	Tiaret	EPI ISL 12156727	2020-05-28	GISAID.

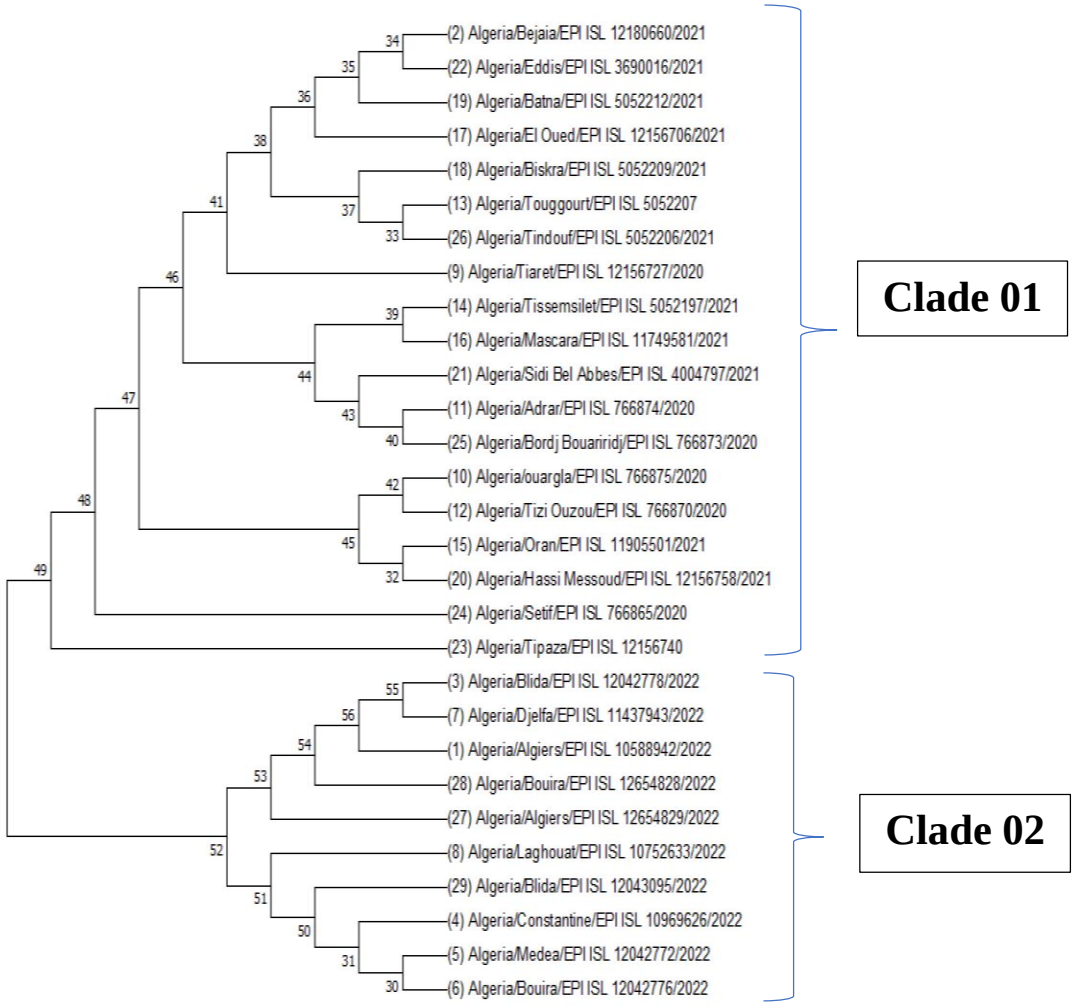


Figure 15. Complete genome phylogenetic tree of SARS Cov-2 strains from different area of Algeria. Isolates are represented with country/accession number of origin and the year of isolation.

The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The optimal tree with the sum of branch length = 0.00900198 is shown. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al., 2004). and are in the units of the number of base substitutions per site. The analysis involved 29 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 15510 positions in the final dataset. Evolutionary analyses were conducted in MEGA 7.0 (Kumar et al., 2016).

The phylogenetic tree (Figure 15) revealed the presence of 02 clades representing SARS Cov-2 isolates from different areas in Algeria and a lot of subclusters.

According to Zeghbib et al (2021), The phylogeographic results are in accordance with the phylogenetic analysis, emphasizing the importance of local travels and social contact in the spread of the disease. The spread of the virus is primarily dependent on social contact, the awareness of the community, and the respectful compliance regarding social distancing, seeing as how lower infection cases in relatively high population density cities were observed and vice versa.

In the phylogenetic tree (Figure 15) we see that there are a lot of subclusters, also sequences are different and distant from each other even if they are from the same country which is Algeria.

We used 2 sequences from Algiers (EPI ISL 12654829 and EPI ISL 10588942) and they're both clustered in the same clade (02) which means they are similar with an evolutionary distance of 0.0002579. Same thing for the two sequences of Bouira (EPI ISL 12042776and EPI ISL 12654828) and Blida (EPI ISL 12043095 and EPI ISL 12042778). Isolates from the same region are from the same strain.

According to Menasria and Aguilera (2022) the African region is characterized by the largest infectious disease burden and the weakest public health infrastructures, which can be explained by the fact that a large population is vulnerable due to conflict, poor socio-economic status, food insecurity and limited access to better health services.

According to Qu et al (2020) some studies have reported that coronavirus can remain in water or wastewater sources for days or weeks.

According to Indseth et al (2021) large cities often have large proportions of immigrants, and urban living could be a factor in the spread of COVID-19.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Algeria/Algiers/EPI ISL 10588942/2022																	
2. Algeria/Bejaia/EPI ISL 12180660/2021	0.0015497																
3. Algeria/Blida/EPI ISL 12042778/2022	0.0001290	0.0015497															
4. Algeria/Constantine/EPI ISL 10969626/2022	0.0005161	0.0018084	0.0005161														
5. Algeria/Medea/EPI ISL 12042772/2022	0.0006452	0.0019378	0.0006452	0.0001290													
6. Algeria/Bouira/EPI ISL 12042776/2022	0.0007743	0.0020673	0.0007743	0.0002580	0.0003870												
7. Algeria/Djelfa/EPI ISL 11437943/2022	0.0002580	0.0016790	0.0002580	0.0006452	0.0007743	0.0009035											
8. Algeria/Laghouat/EPI ISL 10752633/2022	0.0001290	0.0014204	0.0001290	0.0003870	0.0005161	0.0006452	0.0002580										
9. Algeria/Tiaret/EPI ISL 12156727/2020	0.0009035	0.0009034	0.0009035	0.0011619	0.0012912	0.0014204	0.0010326	0.0007743									
10. Algeria/ouargla/EPI ISL 766875/2020	0.0009680	0.0012263	0.0009680	0.0012265	0.0013557	0.0014850	0.0010972	0.0008388	0.0005805								
11. Algeria/Adrar/EPI ISL 766874/2020	0.0012909	0.0015493	0.0012909	0.0015495	0.0016788	0.0018082	0.0014201	0.0011617	0.0009032	0.0009677							
12. Algeria/Tizi Ouzou/EPI ISL 766870/2020	0.0009034	0.0011617	0.0009034	0.0011619	0.0012911	0.0014204	0.0010326	0.0007743	0.0005160	0.0003225	0.0009032						
13. Algeria/Touggourt/EPI ISL 5052207	0.0017437	0.0007098	0.0017437	0.0020025	0.0021320	0.0022615	0.0018730	0.0016143	0.0010972	0.0014201	0.0017432	0.0013556					
14. Algeria/Tissemsilet/EPI ISL 5052197/2021	0.0013555	0.0016139	0.0013555	0.0016142	0.0017435	0.0018729	0.0014848	0.0012263	0.0009678	0.0010323	0.0013552	0.0009678	0.0018079				
15. Algeria/Oran/EPI ISL 11905501/2021	0.0012912	0.0015496	0.0012912	0.0015498	0.0016792	0.0018086	0.0014204	0.0011619	0.0009034	0.0009679	0.0012908	0.0009034	0.0017436	0.0013555			
16. Algeria/Mascara/EPI ISL 11749581/2021	0.0012909	0.0015493	0.0012909	0.0015495	0.0016788	0.0018082	0.0014201	0.0011617	0.0009032	0.0009677	0.0012905	0.0009032	0.0017432	0.0005806	0.0012908		
17. Algeria/El Oued/EPI ISL 12156706/2021	0.0014203	0.0002580	0.0014203	0.0016790	0.0018084	0.0019378	0.0015496	0.0012910	0.0007742	0.0010970	0.0014199	0.0010325	0.0005806	0.0014846	0.0014202	0.0014199	
18. Algeria/Biskra/EPI ISL 5052209/2021	0.0015496	0.0005161	0.0015496	0.0018084	0.0019378	0.0020672	0.0016789	0.0014203	0.0009034	0.0012263	0.0015492	0.0011617	0.0007097	0.0016139	0.0015496	0.0015492	0.0003
19. Algeria/Batna/EPI ISL 5052212/2021	0.0014850	0.0003225	0.0014850	0.0017437	0.0018731	0.0020025	0.0016143	0.0013557	0.0008388	0.0011617	0.0014846	0.0010971	0.0006452	0.0015493	0.0014849	0.0014846	0.0001
20. Algeria/Hassi Messoud/EPI ISL 12156758/2021	0.0015499	0.0018085	0.0015499	0.0018088	0.0019382	0.0020676	0.0016792	0.0014206	0.0011619	0.0012265	0.0015495	0.0011619	0.0020026	0.0016142	0.0005161	0.0015495	0.0016
21. Algeria/Sidi Bel Abbes/EPI ISL 4004797/2021	0.0014851	0.0017436	0.0014851	0.0017439	0.0018733	0.0020027	0.0016144	0.0013558	0.0010972	0.0011618	0.0013554	0.0010972	0.0019377	0.0015494	0.0014850	0.0013554	0.0016
22. Algeria/Eddis/EPI ISL 3690016/2021	0.0015497	0.0003870	0.0015497	0.0018085	0.0019379	0.0020673	0.0016790	0.0014204	0.0009034	0.0012263	0.0015493	0.0011617	0.0007098	0.0016140	0.0015496	0.0015493	0.0002
23. Algeria/Tipaza/EPI ISL 12156740	0.0009681	0.0012264	0.0009681	0.0012265	0.0013558	0.0014851	0.0010972	0.0008389	0.0005806	0.0006451	0.0009678	0.0005806	0.0014202	0.0010324	0.0009680	0.0009678	0.0010
24. Algeria/Setif/EPI ISL 766865/2020	0.0009681	0.0012264	0.0009681	0.0012265	0.0013558	0.0014851	0.0010972	0.0008389	0.0005806	0.0006451	0.0009678	0.0005806	0.0014202	0.0010324	0.0009680	0.0009678	0.0010
25. Algeria/Bordj Bouariridj/EPI ISL 766873/2020	0.0009681	0.0012264	0.0009681	0.0012265	0.0013558	0.0014851	0.0010972	0.0008389	0.0005806	0.0006451	0.0005805	0.0005806	0.0014202	0.0010324	0.0009680	0.0009678	0.0010
26. Algeria/Tindouf/EPI ISL 5052206/2021	0.0017440	0.0007099	0.0017440	0.0020029	0.0021324	0.0022619	0.0018733	0.0016146	0.0010974	0.0014204	0.0017435	0.0013558	0.0007744	0.0018082	0.0017439	0.0017435	0.0005
27. Algeria/Algiers/EPI ISL 12654829/2022	0.0002579	0.0016789	0.0002579	0.0006452	0.0007743	0.0009034	0.0003870	0.0002580	0.0010326	0.0010971	0.0014200	0.0010325	0.0018729	0.0014847	0.0014204	0.0014200	0.0015
28. Algeria/Bouira/EPI ISL 12654828/2022	0.0000645	0.0014851	0.0000645	0.0004516	0.0005807	0.0007098	0.0001935	0.0000645	0.0008389	0.0009034	0.0012263	0.0008389	0.0016790	0.0012909	0.0012266	0.0012263	0.0013
29. Algeria/Blida/EPI ISL 12043095/2022	0.0001290	0.0014204	0.0001290	0.0003870	0.0005161	0.0006452	0.0002580	0.0000000	0.0007743	0.0008388	0.0011617	0.0007743	0.0016143	0.0012263	0.0011619	0.0011617	0.0012

Figure 16. Estimates of Evolutionary Divergence between Sequences (GISAID Database).

2.2. The study of SARS Cov-2 in Algeria and different geographies in the world

To investigate the spread of SARS CoV-2 in Algeria, we performed a thorough analysis of 3 complete SARS-CoV-2 sequences from Algeria (twenty-nine) in addition to 25 sequences sampled worldwide (5 countries per continent) summarized in table 6.

Table 6. Genomic sequences from databases with corresponding continent/ country/accession number/date/reference.

N°	Continents	Countries	Accession number	Date	References
1		Algeria/Constantine	EPI_ISL_10969626	2022-01-06	GISAID.
2		Algeria/Algiers	EPI_ISL_10969631	2022-01-04	GISAID.
3		Algeria/Oran	EPI_ISL_13080457	2022-03-15	GISAID.

4	***Africa***	Egypt	EPI_ISL_12607862	2022-02-14	GISAID.
5		South Africa	EPI_ISL_12644895	2022-04-13	GISAID.
6		Morocco	EPI_ISL_12590769	2022-02-19	GISAID.
7		Kenya	EPI_ISL_12563849	2022-01-02	GISAID.
8	***America***	Colombia	EPI ISL 12690664	2022-04-21	GISAID.
9		USA/ California	EPI_ISL_12656559	2022-05-01	GISAID.
10		Mexico	EPI_ISL_12657413	2022-04-08	GISAID.
11		Canada	EPI_ISL_12643863	2022-01-19	GISAID.
12		Brazil	EPI_ISL_12591193	2022-03-10	GISAID.
13	***Europe***	Denmark	EPI_ISL_12632568	2022-05-06	GISAID.
14		Ireland	EPI_ISL_12646983	2022-03-28	GISAID.
15		Belgium	EPI_ISL_12588379	2022-05-04	GISAID.
16		Austria	EPI_ISL_12578415	2022-04-21	GISAID.
17		Poland	EPI_ISL_12648152	2022-02-15	GISAID.
18	***Asia***	China	EPI_ISL_11905848	2022-03-05	GISAID.
19		Vietnam	EPI_ISL_12647969	2022-04-21	GISAID.
20		Thailand	EPI_ISL_12657710	2022-02-25	GISAID.
21		Iraq	EPI_ISL_12604532	2022-02-09	GISAID.
22		Lebanon	EPI_ISL_11327317	2022-02-17	GISAID.
23	***Oceania***	New Zealand	EPI_ISL_12546394	2022-03-29	GISAID.
24		Australia	EPI_ISL_12645184	2022-04-30	GISAID.
25		Guam	EPI_ISL_12573223	2022-04-06	GISAID.
26		Solomon Islands	EPI_ISL_12489928	2022-04-19	GISAID.
27		American Samoa	EPI_ISL_12380301	2022-04-01	GISAID.

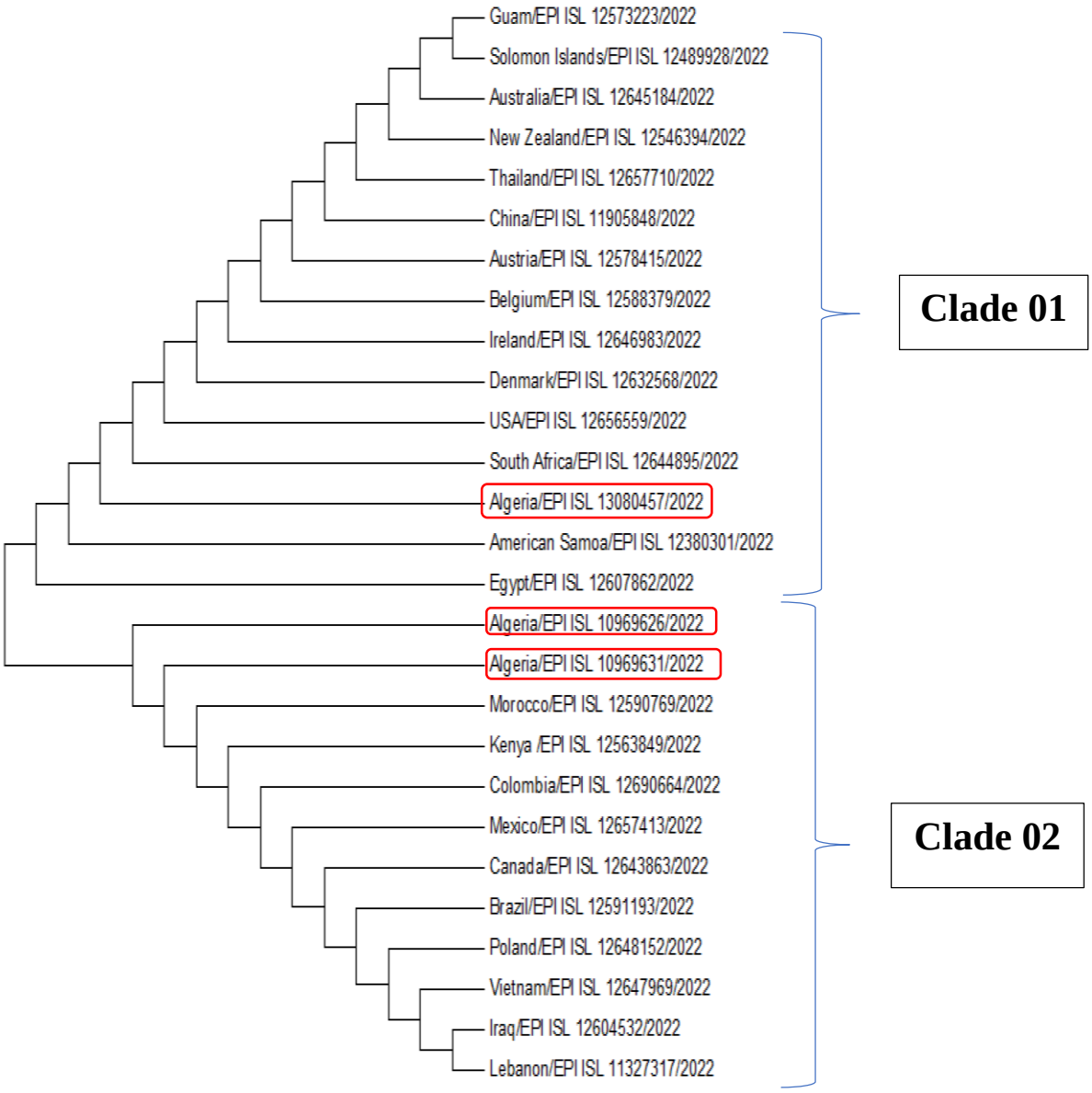


Figure 17. Complete genome phylogenetic tree of SARS Cov-2 strains from Algeria and different geographies in the world. Isolates are represented with country/accession number of origin and the year of isolation.

The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The optimal tree with the sum of branch length = 0.01207654 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite

Likelihood method (Tamura et al., 2004) and are in the units of the number of base substitutions per site. The analysis involved 27 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 335 positions in the final dataset. Evolutionary analyses were conducted in MEGA 7.0 (Kumar et al., 2016).

The phylogenetic tree (Figure 17) revealed the presence of 02 clades representing SARS Cov-2 isolates from different countries per continent with the accession number to GISAID and the year of identification (2022). However, it is found that the three isolates from Algeria are located in two different and distant clades.

Two Algerian isolates (EPI_ISL_10969626/2022 and EPI_ISL_10969631/2022) are clustered in the same clade (Clade 02) with sequences from Morocco (EPI_ISL_12590769), Kenya (EPI_ISL_12563849), Colombia (EPI ISL 12690664), Canada (EPI_ISL_12643863), Brazil (EPI_ISL_12591193), Mexico (EPI_ISL_12657413), Poland (EPI_ISL_12648152), Vietnam (EPI_ISL_12647969), Iraq (EPI_ISL_12604532) and Lebanon (EPI_ISL_11327317).

Whereas the third Algerian isolate (EPI_ISL_13080457) is situated in Clade 01 with American Samoa (EPI_ISL_12380301), South Africa (EPI_ISL_12644895), USA (EPI_ISL_12656559), Denmark (EPI_ISL_12632568), Ireland (EPI_ISL_12646983), Belgium (EPI_ISL_12588379), Austria (EPI_ISL_12578415), China (EPI_ISL_11905848), Thailand (EPI_ISL_12657710), New Zealand (EPI_ISL_12546394), Australia (EPI_ISL_12645184), Solomon Islands (EPI_ISL_12489928), Guam (EPI_ISL_12573223).

The first two Algerian isolates (EPI_ISL_10969631/ EPI_ISL_10969626) located in the clade 02 are identical to each other with evolutionary distance of 0.000000. Even if the third isolate (EPI_ISL_13080457) is not situated with them in the same clade but it is closely interrelated with a very small evolutionary distance of 0.009522 for both isolates. These phylogenetic results of the Algerian isolates clearly show that it there is not a single strain of SARS Cov-2 in Algeria, therefore different origins and evolutionary course of the virus are suggested.

According to Indseth et al (2021) immigrants had higher rates of notified COVID-19 and related hospitalizations than non-immigrants. COVID-19 emerged in a tightly connected world by local and international population on the move, with more people moving for work, education and

family reasons, tourism and survival than ever before. Intense population movements, especially of tourists and business workers, have been key factors in the global spread of the epidemic. This explains that immigration is one of the factors in the spread of SARS CoV-2.

From the phylogenetic analysis the third Algerian isolate is distant from the other 2 isolates which means that there are different strains in Algeria.

According to Eslami & Mahrokh., (2020) the effect of environmental factors such as ambient temperature, humidity, etc., relating to the COVID-19 pandemic can be investigated. Also feeding domestic insects and beetles from the stool and its mechanical transmission can play a significant role in the transmission of the disease.

According to Bontempi (2021) analysis shows that regions with higher international trade activities (that are representative of living standard, economic dynamics, and globalization level) are more susceptible to be in contact with foreign populations, with an increased risk to import the virus in their communities and spread the diffusion. This explains the similarity of Algerian isolate with Vietnam (EPI_ISL_12647969) and Poland (EPI_ISL_12648152), Morocco (EPI_ISL_12590769).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Algeria/EPI_ISL_10969626/2022																	
2. Algeria/EPI_ISL_10969631/2022	0.0000000																
3. Algeria/EPI_ISL_13080457/2022	0.0090522	0.0090522															
4. Egypt/EPI_ISL_12607862/2022	0.0030036	0.0030036	0.0059983														
5. South Africa/EPI_ISL_12644895/2022	0.0090522	0.0090522	0.0000000	0.0059983													
6. Morocco/EPI_ISL_12590769/2022	0.0000000	0.0000000	0.0090522	0.0030036	0.0090522												
7. Kenya /EPI_ISL_12563849/2022	0.0000000	0.0000000	0.0090522	0.0030036	0.0090522	0.0000000											
8. Colombia/EPI_ISL_12690664/2022	0.0000000	0.0000000	0.0090522	0.0030036	0.0090522	0.0000000	0.0000000										
9. USA/EPI_ISL_12656559/2022	0.0090522	0.0090522	0.0000000	0.0059983	0.0000000	0.0090522	0.0090522	0.0090522									
10. Mexico/EPI_ISL_12657413/2022	0.0000000	0.0000000	0.0090522	0.0030036	0.0090522	0.0000000	0.0000000	0.0000000	0.0090522								
11. Canada/EPI_ISL_12643863/2022	0.0000000	0.0000000	0.0090522	0.0030036	0.0090522	0.0000000	0.0000000	0.0000000	0.0090522	0.0000000							
12. Brazil/EPI_ISL_12591193/2022	0.0000000	0.0000000	0.0090522	0.0030036	0.0090522	0.0000000	0.0000000	0.0000000	0.0090522	0.0000000	0.0000000						
13. Denmark/EPI_ISL_12632568/2022	0.0090522	0.0090522	0.0000000	0.0059983	0.0000000	0.0090522	0.0090522	0.0090522	0.0000000	0.0090522	0.0090522	0.0090522					
14. Ireland/EPI_ISL_12646983/2022	0.0090522	0.0090522	0.0000000	0.0059983	0.0000000	0.0090522	0.0090522	0.0090522	0.0000000	0.0090522	0.0090522	0.0090522	0.0000000				
15. Belgium/EPI_ISL_12588379/2022	0.0090522	0.0090522	0.0000000	0.0059983	0.0000000	0.0090522	0.0090522	0.0090522	0.0000000	0.0090522	0.0090522	0.0090522	0.0000000	0.0000000			
16. Austria/EPI_ISL_12578415/2022	0.0090522	0.0090522	0.0000000	0.0059983	0.0000000	0.0090522	0.0090522	0.0090522	0.0000000	0.0090522	0.0090522	0.0090522	0.0000000	0.0000000	0.0000000		
17. Poland/EPI_ISL_12648152/2022	0.0000000	0.0000000	0.0090522	0.0030036	0.0090522	0.0000000	0.0000000	0.0000000	0.0090522	0.0000000	0.0000000	0.0000000	0.0090522	0.0090522	0.0090522	0.0090522	
18. China/EPI_ISL_11905848/2022	0.0090522	0.0090522	0.0000000	0.0059983	0.0000000	0.0090522	0.0090522	0.0090522	0.0000000	0.0090522	0.0090522	0.0090522	0.0000000	0.0000000	0.0000000	0.0000000	0.0090522
19. Vietnam/EPI_ISL_12647969/2022	0.0000000	0.0000000	0.0090522	0.0030036	0.0090522	0.0000000	0.0000000	0.0000000	0.0090522	0.0000000	0.0000000	0.0000000	0.0090522	0.0090522	0.0090522	0.0090522	0.0000000
20. Thailand/EPI_ISL_12657710/2022	0.0090522	0.0090522	0.0000000	0.0059983	0.0000000	0.0090522	0.0090522	0.0090522	0.0000000	0.0090522	0.0090522	0.0090522	0.0000000	0.0000000	0.0000000	0.0000000	0.0090522
21. Iraq/EPI_ISL_12604532/2022	0.0029923	0.0029923	0.0120974	0.0060210	0.0120974	0.0029923	0.0029923	0.0029923	0.0120974	0.0029923	0.0029923	0.0029923	0.0120974	0.0120974	0.0120974	0.0120974	0.0029923
22. Lebanon/EPI_ISL_11327317/2022	0.0000000	0.0000000	0.0090522	0.0030036	0.0090522	0.0000000	0.0000000	0.0000000	0.0090522	0.0000000	0.0000000	0.0000000	0.0090522	0.0090522	0.0090522	0.0090522	0.0000000
23. New Zealand/EPI_ISL_12546394/2022	0.0090522	0.0090522	0.0000000	0.0059983	0.0000000	0.0090522	0.0090522	0.0090522	0.0000000	0.0090522	0.0090522	0.0090522	0.0000000	0.0000000	0.0000000	0.0000000	0.0090522
24. Australia/EPI_ISL_12645184/2022	0.0090522	0.0090522	0.0000000	0.0059983	0.0000000	0.0090522	0.0090522	0.0090522	0.0000000	0.0090522	0.0090522	0.0090522	0.0000000	0.0000000	0.0000000	0.0000000	0.0090522
25. Guam/EPI_ISL_12573223/2022	0.0090522	0.0090522	0.0000000	0.0059983	0.0000000	0.0090522	0.0090522	0.0090522	0.0000000	0.0090522	0.0090522	0.0090522	0.0000000	0.0000000	0.0000000	0.0000000	0.0090522
26. Solomon Islands/EPI_ISL_12489928/2022	0.0090522	0.0090522	0.0000000	0.0059983	0.0000000	0.0090522	0.0090522	0.0090522	0.0000000	0.0090522	0.0090522	0.0090522	0.0000000	0.0000000	0.0000000	0.0000000	0.0090522
27. American Samoa/EPI_ISL_12380301/2022	0.0090522	0.0090522	0.0000000	0.0059983	0.0000000	0.0090522	0.0090522	0.0090522	0.0000000	0.0090522	0.0090522	0.0090522	0.0000000	0.0000000	0.0000000	0.0000000	0.0090522

Figure 18. Estimates of Evolutionary Divergence between Sequences (GISAID Database).

Conclusion

Conclusion

As I know, the present study is one of the first studies about the comparative phylogenetic of Algerian strains on one hand, and with other countries of the world on the other hand.

I have collected the nucleotide sequences of the complete SARS CoV-2 genome, isolated from different countries of the world using the bioinformatics resources of GISAID, to construct the evolutionary distance matrix and the phylogenetic tree by the MEGA 7 software.

Two phylogenetic trees have been created based on 29 isolated strings from different areas in Algeria for the first tree and 27 sequences from the five continents Africa, Europe, America, Asia and Oceania for the second phylogenetic tree. Results showed that Algerian isolates have the same strains with those of four countries: Morocco, Kenya, Canada, Columbia, Brazil, Poland, Lebanon, Iraq, Mexico and Vietnam. In the other part they have very close relation with these countries: American Samoa, South Africa, USA, Ireland, Belgium, Denmark, Austria, China, Thailand, New Zealand, Australia, Solomon Islands and Guam.

The spread of COVID-19 is based on many factors among them immigration, tourism and commercial exchanges.

We conclude that the Algerian isolates were distributed in 2 clades which suggests that there are viral strains of different geographical origin and evolution in Algeria.

I recommend to analyze and sequence much more Algerian genomes of SARS Cov-2 to gain more insights into this virus evolution and understand the origin of the strains with the genetic mutations of isolates. I also suggest to follow up the new sequences added to GISAID database. It is important to remove low quality sequences and use only full-length ones to make this work succeed. Also you can do a phylogenetic analysis at the base of the gene coding for the S protein responsible for binding to the host cell receptors to better understand the pathogenicity of the virus.

List of references

References

(A)

- Adebali O, Bircan A, Circi D, Islek B, Kilinc Z, Selcuk B & Turhan B (2020). Phylogenetic analysis of SARS-CoV-2 genomes in Turkey. Turkish Journal of Biology, Special Issue on COVID-19, 146-156.
- Andrés E. Castilloa,Bárbara Parra,Paz Tapia,Alejandra Acevedo,Jaime Lagos,Winston Andrade,Loredana Arata,Gabriel Leal,Gisselle Barra,Carolina Tambley,Javier Tognarelli,Patricia Bustos,Soledad Ulloa,Rodrigo Fasce,Jorge Fernández .(2020) Phylogenetic analysis of the first four SARS-CoV-2 cases in Chile .Journal of medical virology.
- Angelini MM, Akhlaghpour M, Neuman BW, Buchmeier MJ (2013) severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles.
- Anna Z, Luc B, Frédéric L, Marie M, Jakub V, Olivier G (2021) Origin, evolution and global spread of SARS-CoV-2 Comptes Rendus. Biologies, Tome 344 no. 1, pp 57-75.

(B)

- Bao-Liang Zhong, Wei Luo, Hai-Mei Li, Qian-Qian Zhang, Xiao-Ge Liu, Wen-Tian Li, Yi Li Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey Int J Biol Sci. 2020; 16(10): 1745–1752.
- Bhardwaj K, Sun J, Holzenburg A, Guarino LA, Kao CC (2006) RNA recognition and cleavage by the SARS coronavirus endoribonuclease. J Mol Biol 361:243–256
- Bontempi E., et al. (2021) "Can commercial trade represent the main indicator of the COVID-19 diffusion due to human-to-human interactions? A comparative analysis between Italy, France, and Spain." Environmental Research.

(C)

- Case JB, Ashbrook AW, Dermody TS, Denison MR (2016) Mutagenesis of S-adenosyl-l-methionine-binding residues in coronavirus nsp14 N7-methyltransferase demonstrates differing requirements for genome translation and resistance to innate immunity. J Virol 90:7248–7256

- Cottam EM, Whelband MC, Wileman T (2014) Coronavirus NSP6 restricts autophagosome expansion. Autophagy 10:1426–1441

- Cornillez-Ty CT, Liao L, Yates JR, Kuhn P, Buchmeier MJ (2009) severe acute respiratory syndrome coronavirus nonstructural protein 2 interacts with a host protein complex involved in mitochondrial biogenesis and intracellular signaling. J Virol 83:10314–10318

- Cecilia Acuti M, Maria Elena F, Rosaria C, Francesca B, Lorenzo M, Lamberto M, SARS-CoV-2 pandemic: An overview, Advances in Biological Regulation, Volume 77,(2020)

- Christian J M, Claudine M, Olivier P & Julie Dawn T (2020) (Characterization of accessory genes in coronavirus genomes) Virology Journal

- Chen Y., Guo Y., Pan Y. (2020) Structure analysis of the receptor binding of 2019-nCoV. Biochem Biophys Res Commun. 2:1–6

- Couvreur P and Louvard D. (2021) "COVID-19 and drugs: Pathophysiology and TherapeuticApproaches. Comptes Rendus. Biologies.1-16.

- Cowling, B.J, Leung, G.M (2020). Epidemiological research priorities for public health control of the ongoing global novel coronavirus (2019-nCoV) outbreak. Euro Surveill.

(D)

- Decroly E, Debarnot C, Ferron F, Bouvet M, Coutard B, Imbert I, Gluais L, Papageorgiou N, Sharf A, Bricogne G, Ortiz-Lombardia M, Lescar J, Canard B (2011) Crystal structure and functional

analysis of the sars-coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex. PLoS Pathog 7 : e1002059

- Deng X, Baker SC (2018) An "Old" protein with a new story: coronavirus endoribonuclease is important for evading host antiviral defenses. Virology 517:157–163

- Decaro N., Mari V., Elia G., Addie D.D., Camero M., Lucente M.S., Martella V., Buonavoglia C (2010) Recombinant canine coronaviruses in dogs, Europe. *Emerg. Infect.* 16:41–47.

- Decroly E, Debarnot C, Ferron F, Bouvet M, Coutard B, Imbert I, Gluais L, Papageorgiou N, Sharf A, Bricogne G, Ortiz-Lombardia M, Lescar J, Canard B (2011) Crystal structure and functional analysis of the sars-coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex. PLoS Pathog 7: e1002059
- de Haan, C. A., M. de Wit, L. Kuo, C. Montalto-Morrison, B. L. Haagmans, S. R. Weiss, P. S. Masters, and P. J. Rottier. 2003. The glycosylation status of the murine hepatitis coronavirus M protein affects the interferogenic capacity of the virus in vitro and its ability to replicate in the liver but not the brain. Virology.
- Di W, Tiantian W, Qun L, Zhicong Y, (2020) The SARS-CoV-2 outbreak: What we know, International Journal of Infectious Diseases, Volume 94, Pages 44-48

(E)

- Eslami H, Ehrampoush MH, Falahzadeh H, Hematabadi PT, Khosravi R, Dalvand A, Esmaeili A, Taghavi M, Ebrahimi AA (2018) Biodegradation and nutrients removal from greywater by an integrated fixed-film activated sludge (IFAS) in different organic loadings rates.
- Eslami, H, and Mahrokh J (2020). "The role of environmental factors to transmission of SARS-CoV-2 (COVID-19)." Amb Express.10:1-8.

(F)

- Forster, Peter, et al (2020): "Phylogenetic network analysis of SARS-CoV-2 genomes." Proceedings of the National Academy of Sciences.117.17 9241-9243).

(G)

- Gao Y, Yan L, Huang Y, Liu F, Zhao Y, Cao L, Wang T, Sun Q, Ming Z, Zhang L, Ge J, Zheng L, Zhang Y, Wang H, Zhu Y, Zhu C, Hu T, Hua T, Zhang B, Yang X, Li J, Yang H, Liu Z, Xu W, Guddat LW, Wang Q, Lou Z, Rao Z (2020) Structure of the RNA-dependent RNA polymerase from COVID-19 virus. Science
- Gerwyn M, Chiara C. Bortolasci, Basant K. Puri, Lisa O, Wolfgang Marx, Adrienne O'Neil, Eugene Athan, Andre F. Carvalho, Michael M, Ken W, Michael B (2020). The pathophysiology of SARS-CoV-2: A suggested model and therapeutic approach, Life Sciences, Volume 258
- Gordon, D.E., Jang, G.M., Bouhaddou, M. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583, 459–468 (2020).
- Gorbalenya1, Alexander E. 2, Susan C. Baker3, Ralph S. Baric4, Raoul J. de Groot5, Christian Drosten6, Anastasia A. Gulyaeva1, Bart L. Haagmans7, Chris Lauber1, Andrey M Leontovich2, Benjamin W. Neuman8, Dmitry Penzar2, Stanley Perlman9, Leo L.M. Poon10, Dmitry Samborskiy2, Igor A. Sidorov, Isabel Sola11, John Ziebuhr (2020). Severe acute respiratory syndrome-related coronavirus: The species and its viruses – a statement of the Coronavirus Study Group
- Gorla, U. S., Rao, G. K. (2020). SARS-CoV-2: The Prominent Role of Nonstructural Proteins (NSPS) in COVID-19. Indian Journal of Pharmaceutical Education and Research. S381-S389
- Guo Y-R, Cao Q-D, Hong Z-S, Tan Y-Y, Chen S-D, Jin H-J, Tan K-S, Wang D-Y, Yan Y (2020). The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status. *Mil Med Res*

(H)

- Harrison, A. G., Lin, T., Wang, P. (2020). Mechanisms of SARS-CoV-2 Transmission and Pathogenesis. Trends in immunology.41(12) :1100-1115.

- Hackbart M, Deng X, Baker SC (2020) Coronavirus endoribonuclease targets viral polyuridine sequences to evade activating host sensors. Proc Natl Acad Sci USA 117:8094–8103

- Huang, Y, et al (2020) "Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19." Acta Pharmacologica Sinica 41.9 1141-1149.

- Huang C, Lokugamage KG, Rozovics JM, Narayanan K, Semler BL, Makino S (2011) SARS coronavirus nsp1 protein induces template-dependent endonucleolytic cleavage of mRNAs: viral mRNAs are resistant to nsp1-induced RNA cleavage. PLoS Pathog 7: e1002433

- Helmy YA, Fawzy M, Elaswad A, Sobieh A, Kenney SP, Shehata AA (2020). The COVID-19 Pandemic: A Comprehensive Review of Taxonomy, Genetics, Epidemiology, Diagnosis, Treatment, and Control. *Journal of Clinical Medicine*.; 9(4):1225

(I)

- Ivanov KA, Thiel V, Dobbe JC, van der Meer Y, Snijder EJ, Ziebuhr J (2004) Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. J Virol 78:5619–5632

- Indseth, T., Grøsland, M., Arnesen, T., Skyrud, K., Kløvstad, H., Lamprini, V., Telle, K., & Kjøllesdal, M. (2021). COVID-19 among immigrants in Norway, notified infections, related hospitalizations and associated mortality: A register-based study. *Scandinavian Journal of Public Health*, *49*(1), 48–56.

(J)

- Jang K-J, Jeong S, Kang DY, Sp N, Yang YM, Kim D-E (2020) A high ATP concentration enhances the cooperative translocation of the SARS coronavirus helicase nsP13 in the unwinding of duplex RNA. Sci Rep 10:4481
- Jia Z, Yan L, Ren Z, Wu L, Wang J, Guo J, Zheng L, Ming Z, Zhang L, Lou Z, Rao Z (2019) Delicate structural coordination of the severe acute respiratory syndrome coronavirus Nsp13 upon ATP hydrolysis. Nucleic Acids Res 47:6538–6550

(K)

- Kumar P, Gunalan V, Liu B, Chow VTK, Druce J, Birch C, Catton M, Fielding BC, Tan Y-J, Lal SK (2007) The nonstructural protein 8 (nsp8) of the SARS coronavirus interacts with its ORF6 accessory protein. Virology 366:293–303

- Kumar S., Stecher G., and Tamura K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets.Molecular Biology and Evolution 33:1870-1874.

- Kumar, S., Nyodu, R., Maurya, V.K., Saxena, S.K. (2020). Morphology, Genome Organization, Replication, and Pathogenesis of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). In: Saxena, S. (eds) Coronavirus Disease 2019 (COVID-19). Medical Virology: From Pathogenesis to Disease Control. Springer, Singapore
- Khailany RA, Safdar M, Ozaslan M (2020) Genomic characterization of a novel SARS-CoV-2. Gene Rep 19:100682

(L)

- Lei J, Kusov Y, Hilgenfeld R (2018) Nsp3 of coronaviruses: structures and functions of a large multi-domain protein. Antiviral Res 149:58–74

- Liu J, Zhou J, Yao J, Zhang X, Li L, Xu X, He X, Wang B, Fu S, Niu T, Yan J, Shi Y, Ren X, Niu J, Zhu W, Li S, Luo B, Zhang K (2020). Impact of meteorological factors on the COVID-19 transmission: A multi-city study in China. Sci Total Environ.; 726:138513.

- Liu S.L., Saif L.J., Weiss S.R., Su L (2020). No credible evidence supporting claims of the laboratory engineering of SARS-CoV-2. *Emerg. Microbes Infect.*
- Le TM, Wong HH, Tay FPL, Fang S, Keng C-T, Tan YJ, Liu DX (2007) Expression, post-translational modification and biochemical characterization of proteins encoded by subgenomic mRNA8 of the severe acute respiratory syndrome coronavirus. FEBS J 274:4211–4222
- Liu DX, Fung TS, Chong KK-L, Shukla A, Hilgenfeld R (2014) Accessory proteins of SARS-CoV and other coronaviruses. Antiviral Res 109:97–109
- Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al (2020) Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. N Engl J Med.

(M)

- Machhi, J., Herskovitz, J., Senan, A. M., Dutta, D., Nath, B., Oleynikov, M. D., Kevadiya, B. D. (2020). The natural history, pathobiology, and clinical manifestations of SARS-CoV-2 infections. Journal of Neuroimmune Pharmacology.1-28.
- Menasria, T.; Aguilera, M (2022). Genomic Diversity of SARS-CoV-2 in Algeria and North African Countries: What We Know So Far and What We Expect? Microorganisms.
- Ma Y, Wu L, Shaw N, Gao Y, Wang J, Sun Y, Lou Z, Yan L, Zhang R, Rao Z (2015) Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. Proc Natl Acad Sci USA 112:9436–9441
- Mu J, Xu J, Zhang L, Shu T, Wu D, Huang M, Ren Y, Li X, Geng Q, Xu Y, Qiu Y, Zhou X (2020) SARS-CoV-2-encoded nucleocapsid protein acts as a viral suppressor of RNA interference in cells. Sci China Life Sci 63:10

(N)

- Nelson CA, Pekosz A, Lee CA, Diamond MS, Fremont DH (2005) Structure and intracellular targeting of the SARS-coronavirus Orf7a accessory protein. Structure 13:75–85

(O)

- Oude Munnin, B.B., Nieuwenhuijse, D.F., Stein, M. *et al.* (2020). Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat Med* 26**,** 1405–1410

(Q)

- Qu G, Li X, Hu L, Jiang G (2020) An imperative need for research on the role of environmental factors in transmission of novel Coronavirus (COVID-19). Environ Sci Technol 54:3730–3732.

(R)

- Rajnik M, Cascella M, Cuomo A X, Dulebohn S. C, Di Napoli R. (2021). Features, Evaluation, and Treatment of Coronavirus (COVID-19). Uniformed Services University of the Health Sciences. 1-21

- Rajiv S, Mohd J, Abid H, Raju V, Shashi B, Devaki N (2020) Sustainability of Coronavirus on Different Surfaces, Journal of Clinical and Experimental Hepatology, Volume 10, Issue 4
- Rota P.A., Oberste M.S., Monroe S.S., Nix W.A., Campagnoli R., Icenogle J.P., Penaranda S., Bankamp B., Maher K., Chen M.-h (2003). Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science.

(S)

- Sallard, E., Halloy, J., Casane, D., van Helden, J., Decroly, É. (2020). Retrouver les origines du SARS-CoV-2 dans les phylogénies de coronavirus. Médecine/sciences.36(8-9) :783-796.

- Sakai Y, Kawachi K, Terada Y, Omori H, Matsuura Y, Kamitani W (2017) Two-amino acids change in the nsp4 of SARS coronavirus abolishes viral replication. Virology 510:165–174

- Shuo S, Gary W, Weifeng S, Jun L, Alexander C.K. Lai, Jiyong Z, Wenjun L, Yuhai B, George F. Gao (2016). epidemiology, Genetic Recombination and pathogenesis of corona viruses Trends Microbiol; 24(6): 490–502.

- Schaecher SR, Mackenzie JM, Pekosz A (2007) The ORF7b protein of severe acute respiratory syndrome coronavirus (SARSCoV) is expressed in virus-infected cells and incorporated into SARS-CoV particles. J Virol 81 :718–731

- Satarker, S., & Nampoothiri, M. (2020). Structural Proteins in Severe Acute Respiratory Syndrome Coronavirus-2. *Archives of medical research*, *51*(6), 482–491.

- Stroher, U., A. DiCaro, Y. Li, J. E. Strong, F. Aoki, F. Plummer, S. M. Jones, and H. Feldmann. 2004. severe acute respiratory syndrome-related coronavirus is inhibited by interferon-alpha. J. Infect.

- Sudhir K, Glen S, Koichiro Tamura (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets, *Molecular Biology and Evolution*, Volume 33, Issue 7, Pages 1870–1874,
- Subissi L, Posthuma CC, Collet A, Zevenhoven-Dobbe JC, Gorbalenya AE, Decroly E, Snijder EJ, Canard B, Imbert I (2014) One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. Proc Natl Acad Sci USA 111: E3900–E3909
- Siu K-L, Yuen K-S, Castano-Rodriguez C, Ye Z-W, Yeung M-L, Fung S-Y, Yuan S, Chan C-P, Yuen K-Y, Enjuanes L, Jin D-Y (2019) severe acute respiratory syndrome coronavirus ORF3a protein activates the NLRP3 infammasome by promoting TRAF3-dependent ubiquitination of ASC. FASEB J 33:8865–8877

- Susan R. Weiss, Sonia Navas-Martin (Coronavirus Pathogenesis and the Emerging Pathogen Severe Acute Respiratory Syndrome Coronavirus Microbiol Mol Biol Rev. 2005 Dec; 69(4): 635–664. doi: 10.1128/MMBR.69.4.635-664.2005)

- Saitou N. and Nei M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4:406-425.

(T)

- te Velthuis AJ, van de Worm SH, Snijder EJ (2012) The SARS coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension. Nucleic Acids Res 40:1737–1747

- Tong, S. et al (2009). Detection of novel SARS-like and other coronaviruses in bats from Kenya. *Emerg. Infect. Dis.* 15, 482–485
- Tomar S, Johnston ML, St. John SE, Osswald HL, Nyalapatla PR, Paul LN, Ghosh AK, Denison MR, Mesecar AD (2015) Ligandinduced dimerization of middle east respiratory syndrome (MERS) coronavirus nsp5 protease (3CLpro) implications for nsp5 Regulation and The Development of Antivirals. J. Biol. Chem. 290:19403–19422

- Thomas, Sunil. (2020) "The structure of the membrane protein of SARS-CoV-2 resembles the sugar transporter semiSWEET." Pathogens and Immunity.5.1: 342.

- Tamura K, Nei M, and Kumar S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Sciences (USA) 101:11030-11035.

(V)

- Verdiá-Báguena C, Jose L. Nieto-Torres, Antonio A, Marta L. DeDiego, Jaume T, Vicente M. Aguilella, Luis E, (2012). Coronavirus E protein forms ion channels with functionally and structurally-involved membrane lipids,Virology,Volume 432, Issue 2, Pages 485-494,

- Velazquez-Salinas L, Zarate S, Eberl S, Gladue DP, Novella I, Borca MV (2020). Positive Selection of ORF1ab, ORF3a, and ORF8 Genes Drives the Early Evolutionary Trends of SARS-CoV-2 During the 2020 COVID-19 Pandemic. *Front Microbiol* 11:550674.
- Vella, F., Senia, P., Ceccarelli, M., Vitale, E., Maltezou, H., Taibi, R., ... & Ledda, C. (2020). Transmission mode associated with coronavirus disease 2019: a review. *Health*, *1*(2).

(W)

- Wang Y, Sun Y, Wu A, Xu S, Pan R, Zeng C, Jin X, Ge X, Shi Z, Ahola T, Guo D (2015) Coronavirus nsp10/nsp16 methyltransferase can be targeted by nsp10-derived peptide in vitro and in vivo to reduce replication and pathogenesis. J Virol 89:8416–8427
- Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Vessler D (2020) Structure, function, and antigenicity of the SARSCoV-2 spike glycoprotein. Cell 180:281–292
- Weiss, S. R., & Navas-Martin, S. (2005). Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. Microbiology and molecular biology reviews: *MMBR*, *69*(4), 635–664.
- WHO (2020) WHO Director-General's Opening Remarks at the Media Briefing on COVID-19
- WHO (2020) Water, sanitation, hygiene and waste management for COVID-19: technical brief, 03 March 2020. World Health Organization, Geneva.
- (World Health Organization / Corona virus disease (covid-19))
- Wong HH, Fung TS, Fang S, Huang M, Le MT, Liu DX (2018). Accessory proteins 8b and 8ab of severe acute respiratory syndrome coronavirus suppress the interferon signaling pathway by mediating ubiquitin-dependent rapid degradation of interferon regulatory factor 3. *Virology.* 515:165–175
- (WHO Rolling Updates on Coronavirus Disease (COVID-19) [(accessed on 14 April 2020)])
- Wurm, T., H. Chen, T. Hodgson, P. Britton, G. Brooks, and J. A. Hiscox (2001). Localization to the nucleolus is a common feature of coronavirus nucleoproteins, and the protein may disrupt host cell division. J. Virol. 75

(X)

- Xia S, Zhu Y, Liu M (2020). Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. Cell Mol Immunol.:1–3

(Y)

- Yashpal S M, Shubhankar S, Sudipta B, Khan S, Kuldeep D, Maryam D, Ruchi T & Wanpen C (2020) Emerging novel coronavirus (2019-nCoV)—current scenario, evolutionary perspective based on genome analysis and recent developments, Veterinary Quarterly, 40:1, 68-76
- Yadav R, Chaudhary J.K, Jain N, Chaudhary P.K, Khanra S, Dhamija P, Sharma A, Kumar A, Handu S (2021). Role of Structural and Non-Structural Proteins and Therapeutic Targets of SARS-CoV-2 for COVID-19. *Cells*, *10*, 821
- Yan X., Hao Q., Mu Y (2006). Nucleocapsid protein of SARS-CoV activates the expression of cyclooxygenase-2 by binding directly to regulatory elements for nuclear factor-kappa B and CCAAT/enhancer binding protein. *Int J Biochem Cell Biol* 38:1417–1428
- Yount B, K. M. Curtis, E. A. Fritz, L. E. Hensley, P. B. Jahrling, E. Prentice, M. R. Denison, T. W. Geisbert, and R. S. Baric. (2003). Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus. Proc. Natl. Acad. Sci. USA 100
- Yoshimoto; Francis K (2021) A Biochemical Perspective of the Nonstructural Proteins (NSPs) and the Spike Protein of SARS CoV Accepted: 30 January 2021 / Published online: 24 February 2021 The Protein Journal 40:260–295

(Z)

- Zhou ; X.-L. Yang; X.-G. Wang; B. Hu; L. Zhang; W. Zhang; et al (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin, *Nature*, Volume 579 ,pp. 270-273
- Zhao J, Falcon A, Zhou H, Netlan J, Enjuanes L, Brena PP, Perlman S (2009) severe acute respiratory syndrome coronavirus protein 6 is required for optimal replication. J Virol 83:2368–2373

- Zhao S, Ge X, Wang X, Liu A, Guo X, Zhou L, Yu K, Yang H (2015) The DEAD-box RNA helicase 5 positively regulates the replication of porcine reproductive and respiratory syndrome virus by interacting with viral Nsp9 in vitro. Virus Res 195:217–224

- Zhao, H., Lu, X., Deng, Y., Tang, Y., & Lu, J. (2020). COVID-19: asymptomatic carrier transmission is an underestimated problem. *Epidemiology and infection*, *148*, e116.

- Zeghbib S, Somogyi B, Zana, B, Kemenesi G, Herczeg R, Derrar F, Jakab, F (2021). The Algerian Chapter of SARS-CoV-2 Pandemic: An Evolutionary, Genetic, and Epidemiological Prospect. *Viruses*

Abstract

The world is facing a threat because of the outbreak of COVID-19, and therefore humanity must work to eliminate it. In this present study 51 sequences of whole genome of SARS Cov-2 from Algeria and different geographies all over the world were used from GISAID database. The nucleotide sequences were the subject of an alignment, an estimation of the evolutionary distances and the construction of a dendrogram by the bioinformatics tool MEGA 7. The results showed a significant genetic variability of SARS-CoV-2 in the world.

Key words: SARS CoV-2, phylogeny, evolutionary distances, genetic variability.

Résumé

Le monde est confronté à une menace à cause de l'épidémie du COVID-19, et donc l'humanité doit travailler pour l'éliminer. Dans cette présente étude, 51 séquences du génome complet du SARS CoV-2 d'Algérie et de différentes régions du monde ont été utilisées à partir de la base de données GISAID. Les séquences nucléotidiques ont fait l'objet d'un alignement, d'une estimation des distances évolutives et de la construction d'un dendrogramme par l'outil bio-informatique MEGA 7. Les résultats ont montré une importante variabilité génétique du SARS-CoV-2 dans le monde.

Mots clés: SARS CoV-2, phylogénie, distances évolutives, variabilité génétique.

ملخص

يواجه العالم تهديدًا بسبب تفشي كوفيد -19، وبالتالي يجب على البشرية العمل على القضاء عليه. في هذه الدراسة الحالية تم استخدام 51 سلسلة من الجينوم الكامل لـــ SARS CoV-2 من الجزائر ومن مناطق جغرافية مختلفة في جميع أنحاء العالم من قاعدة بيانات GISAID. كانت متواليات النيكليوتيدات موضوع محاذاة وتقدير للمسافات التطورية وبناء مخطط شجر بواسطة أداة المعلوماتية الحيوية MEGA 7. وأظهرت النتائج تباينًا جينيًا كبيرًا لـ SARS CoV-2 في العالم.

الكلمات المفتاحية: SARS CoV-2، علم الوراثة، المسافات التطورية، التباين الجيني.

First name and last name: **Date of dissertation: 06/07/2022**

BOULESBAA WIAM

Theme:

The phylogeny of SARS Cov-2 (an update of the year 2022)

Abstract :

The world is facing a threat because of the outbreak of COVID-19, and therefore humanity must work to eliminate it. In this present study 51 sequences of whole genome of SARS Cov-2 from Algeria and different geographies all over the world were used from GISAID database. The nucleotide sequences were the subject of an alignment, an estimation of the evolutionary distances and the construction of a dendrogram by the bioinformatics tool MEGA 7. The results showed a significant genetic variability of SARS-CoV-2 in the world.

Key words: SARS CoV-2, phylogeny, evolutionary distances, genetic variability

In front of The Jury:

- **President: Merzoug Amina (M.C.A) Mila University Center**
- **Examiner: Boucherit Hanane (M.C.B) Mila University Center**
- **Supervisor: Boubendir Abdelhafid (M.C.A) Mila University Center**