**PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA**

**MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH MOHAMED BOUDIAF UNIVERSITY - M'SILA**

**FACULTE DES SCIENCES** | **DOMAINE : SCINCES DE LA NATURE ET DE LA VIE**

**DEPARTEMENT DE MICROBIOLOGIE & BIOCHIMIE** | **FILIERE : SCIENCES ALIMENTAIRES**

**N° : …….** | **OPTION : QUALITE DES PRODUITS ET SECURITE ALIMENTAIRE**

**Mémoire présenté pour l'obtention**

**Du diplôme de Master professionnel**

**Par** : Zerrouki Fatima

# Intitulé

# Genomic analysis of Coronaviruses

**Soutenu devant le jury composé de :**

| | | |
|---|---|---|
| Dr. Abdenassar HARRAR | Université de M'sila | Président |
| Dr. Mohammed Abdalla KHODJA | Université de M'sila | Rapporteur |
| Dr. Samia Bouaziz | Université de M'sila | Examinateur |

**Année universitaire : 2024/2025**

# Remerciement

# Remerciement

I would like to express my sincere gratitude to Mr. Mohamed Abdallah KHODJA , my supervisor, for his valuable guidance, continuous support, and encouragement throughout this work. His scientific expertise and dedication have been instrumental in the successful completion of this thesis.

My deepest thanks also go to Mr. Abdenassar Harrar, President of the jury, for honoring me by accepting to review this work and for his insightful and constructive comments.

I am equally grateful to my lab instructor Samia Bouaziz, examiner, for taking the time to evaluate this dissertation and for his thoughtful observations that have helped improve the quality of the manuscript.

To all of them, I extend my heartfelt appreciation.

# Dedication

# Dedication

I dedicate this work to those who have given me endless love, strength, and inspiration throughout my life.

To my beloved mother and father, whose sacrifices and unwavering support have shaped who I am today — this is for you.

To my dear sisters, Iman Louiza and Nadjat, and my brothers, Oussama meftah ,seddam and Habib, thank you for your constant encouragement and presence by my side.

And to my lifelong friend, MadiHa, whose friendship has been a steady light through every season of life.

With all my love,

Fatima Zerrouki

# Table of contents

**List of Figures**

**List of Tables**

**Abstract**

# List of Figures

# List of tables

# Summary

**Abstract**

This study presents a bioinformatics-based approach to the comparative genomic analysis of selected coronaviruses, including SARS-CoV, SARS-CoV-2, and other animal-related strains such as bat, pangolin, porcine, and feline coronaviruses. Using the BioEdit software, we carried out sequence retrieval, editing, alignment, nucleotide translation, and mutation detection.

In this study, a basic genetic analysis was conducted on several coronavirus strains originating from different sources, including both human and animal hosts. The samples comprised genomic sequences of human coronaviruses such as SARS-CoV and SARS-CoV-2, as well as animal coronaviruses isolated from bats, pangolins, felines, and pigs.

All genomic sequences were retrieved from the NCBI database and subsequently processed and analyzed using the BioEdit software. The analysis was limited to sequence comparison in order to examine similarities and differences among the selected coronavirus genomes, without performing advanced evolutionary analyses or complex statistical modeling.

This basic genetic analysis provides a preliminary overview of the degree of genetic relatedness among coronaviruses from different hosts and serves as an initial step toward understanding potential genetic relationships between human and animal coronaviruses.

**Keywords:** Coronavirus, SARS-CoV-2, BioEdit, Genomic analysis, Sequence alignment, Genetic databases, Bioinformatics tools.

**Résumé :**

Cette étude présente une approche bioinformatique basée sur l'analyse génomique comparative de coronavirus sélectionnés, incluant le SARS-CoV, le SARS-CoV-2, ainsi que d'autres souches d'origine animale telles que les coronavirus de chauves-souris, de pangolins, de porcs et de félins. À l'aide du logiciel BioEdit, nous avons effectué la récupération des séquences, leur édition, leur alignement, la traduction nucléotidique ainsi que la détection des mutations.

Dans cette étude, une analyse génétique de base a été réalisée sur plusieurs souches de coronavirus provenant de différentes sources, incluant des hôtes humains et animaux. Les échantillons comprenaient des séquences génomiques de coronavirus humains tels que le SARS-CoV et le SARS-CoV-2, ainsi que des coronavirus animaux isolés chez les chauves-souris, les pangolins, les félins et les porcs.

Toutes les séquences génomiques ont été extraites de la base de données NCBI, puis traitées et analysées à l'aide du logiciel BioEdit. L'analyse s'est limitée à la comparaison des séquences afin d'examiner les similitudes et les différences entre les génomes de coronavirus sélectionnés, sans réaliser d'analyses évolutives avancées ni de modélisation statistique complexe.

Cette analyse génétique de base fournit un aperçu préliminaire du degré de parenté génétique entre les coronavirus provenant de différents hôtes et constitue une étape initiale vers la compréhension des relations génétiques potentielles entre les coronavirus humains et animaux.

**Mots-clés :** Coronavirus, SARS-CoV-2, BioEdit, Analyse génomique, Alignement de séquences, Bases de données génétiques, Outils bioinformatiques.

**الملخص :**

تم في هذا العمل إجراء تحليل جيني بسيط لعدة فيروسات كورونا ذات مصادر مختلفة تشمل الإنسان والحيوان. شملت العينات تسلسلات جينية لفيروسات كورونا بشرية مثل SARS-CoV وSARS-CoV-2، إضافة إلى فيروسات كورونا حيوانية معزولة من الخفافيش، آكل النمل الحرشفي(pangolin) ، القطط، والخنازير.

تم الحصول على جميع التسلسلات الجينية من قاعدة بياناتNCBI ، ثم تمت معالجتها وتحليلها باستخدام برنامج .BioEdit اقتصر التحليل على مقارنة التسلسلات الجينية بهدف دراسة أوجه التشابه والاختلاف بينها، دون إجراء تحليلات تطورية متقدمة أو نمذجة إحصائية معقدة.

يسمح هذا التحليل الجيني الأساسي بتكوين تصور أولي حول درجة التقارب الجيني بين فيروسات كورونا ذات المصادر المختلفة، ويُعد خطوة تمهيدية لفهم العلاقات الجينية المحتملة بين الفيروسات البشرية والحيوانية.

.

# Introduction

## General Introduction

The global outbreak of SARS-CoV-2, the virus responsible for COVID-19, has marked one of the most significant public health crises of the 21st century[1]. Since its emergence in late 2019, this novel coronavirus has triggered a scientific race to understand its genetic makeup, evolutionary origin, mechanisms of transmission, and potential targets for treatment and prevention. As a result, the genomic analysis of SARS-CoV-2 has become a cornerstone in efforts to control the pandemic.

Despite the availability of extensive viral genome databases and computational tools, multiple scientific challenges remain unresolved. Among the most pressing issues is the origin of SARS-CoV-2.

Did it arise directly from bats, or was there an intermediate host such as pangolins, swine, or domestic cats [2] **.**

Additionally, the virus's rapid mutation rate, emergence of new variants, and spike protein evolution raise complex questions regarding vaccine effectiveness and long-term immunity [3].

To address such issues, the integration of bioinformatics with molecular virology has become essential. Bioinformatics allows for the alignment, annotation, and comparison of viral genomes across different hosts and regions. Through these methods, researchers can trace the virus’s evolutionary trajectory, identify mutation hotspots, and construct phylogenetic trees to compare SARS-CoV-2 with other coronaviruses [4]. This computational approach forms the foundation of genomic surveillance, a critical strategy in predicting and preventing future pandemics.

This research is purely theoretical and aims to highlight how genomic data—when properly analyzed using bioinformatics tools—can provide answers to fundamental virological and epidemiological questions. The ultimate goal is to encourage the use of such methods in improving global health and precision medicine, particularly in the context of emerging zoonotic viruses.

Accordingly, this thesis is structured into the following three chapters:

- **Chapter 1: History of Coronaviruses: This** chapter presents an overview of the discovery and classification of coronaviruses, from the early identification of SARS-CoV to the emergence of SARS-CoV-2.
- **Chapter 2: Literature Review on Coronavirus:** Here, we examine current scientific findings and competing hypotheses regarding the genetic structure, mutation patterns, and origin theories of SARS-CoV-2.
- **Chapter 3: Bioinformatics Tools for Coronavirus Genome Analysis** :This chapter explores key computational methods and platforms used to align, compare, and interpret coronavirus genomes, including tools such as BioEdit, MEGA, and online databases like GISAID and GenBank.
- **Chapter 4 :** Results and Discussion.

# **Chapter I:** History and fundamental concept about coronavirus

## 1. The Nature of Viruses: Structure, Replication, and Host Interaction

The term virus originates from the Latin virus, meaning "poison." A virus is an acaryotic microorganism, a biological entity incapable of autonomous reproduction[5]. It requires a host cell to replicate, utilizing the host's cellular machinery for its propagation, thereby classifying it as an obligate intracellular parasite **(Belarbi, 2019).**

Viruses are infectious agents that cannot grow or divide independently. Their replication strictly depends on host cells. To reproduce, viruses must penetrate host cells and hijack their cellular machinery to synthesize viral proteins and replicate their genetic material. Newly formed viral particles within the host cell are subsequently released, infecting other cells in a continuous cycle[5].

All living organisms are susceptible to viral infections. Currently, approximately 1,035 types of viruses have been identified, classified into multiple families based on the nature of their genome. Viral genomes can exist as either DNA (deoxyribonucleic acid) or RNA (ribonucleic acid). For instance, SARS-CoV-2 is an RNA virus. The viral genetic material is enclosed within a protective protein coat called the capsid. Some viruses, such as SARS-CoV-2, possess an additional lipid envelope surrounding the capsid[6].

In summary, the viral life cycle involves entering the cells of an animal or plant host, replicating in large numbers, depleting cellular resources, and exiting to infect new cells. Some viruses cause the destruction of infected cells and tissues, leading to diseases with potentially fatal outcomes. A virus with a 100% mortality rate in its host population would eventually cease to exist due to the lack of viable hosts for replication. In contrast, well-adapted viruses establish efficient replication without causing excessive harm. This is exemplified by coronaviruses that have coexisted with bats for millennia.

## 2. Epidemiology

The outbreak originated in the city of Wuhan, China, in late December 2019 and rapidly spread worldwide.

The pandemic resulted in approximately 850,000 deaths globally, with over 25.5 million confirmed cases of infection [7]**.**

## 3. Chronology of Events

Based on info from the World Health Group **(WHO)**[7]**:**

- ✓ **December 31, 2019:** China tells of many sick with lung illness in Wuhan, Hubei.
- ✓ **January 1, 2020:** They shut the Huanan seafood place.
- ✓ **January 5, 2020:** WHO puts out its first news on the new virus.
- ✓ **January 30, 2020:** WHO says the virus mess is a big world health worry; at the same time, China shows 7,711 sick, 170 dead. The virus reaches all parts of China.
- ✓ **March 11, 2020:** WHO names COVID-19 a world illness surge.

The WHO has been central in sharing info on the virus, following how it moves, and giving advice on how to stay safe and stop the spread.For many months, scientists and disease experts have guessed that the number of infections and deaths would go down, with better days seen from May 2022. By June 1, 2022, there were 529,036,622 known COVID-19 cases and more than 6,294,700 deaths reported[7] [8]**.**

## 4. Coronaviruses

Coronaviruses get their name from their crown-like shape, seen with a high-power microscope. They are round, covered viruses about 100 to 160 nm wide. They have a one-strand RNA with 27 to 32 big base pairs[8]**.**

The World Health Group **(WHO)** says that coronaviruses make up a big family of viruses able to make both people and animals sick. They can cause lung sick things from a simple cold to very bad illness[7] **.**Coronaviruses (CoVs) are part of the Orthocoronavirinae group in the Coronaviridae family[8]**.**

In people, some coronaviruses bring on light colds. But, some can cause bad sicknesses like Severe Acute Respiratory Syndrome **(SARS) from SARS-CoV-1,** and Middle East

Respiratory Syndrome **(MERS).** The newest one found, SARS-CoV-2, makes the sickness we call Coronavirus Disease 2019 (COVID-19)[8].

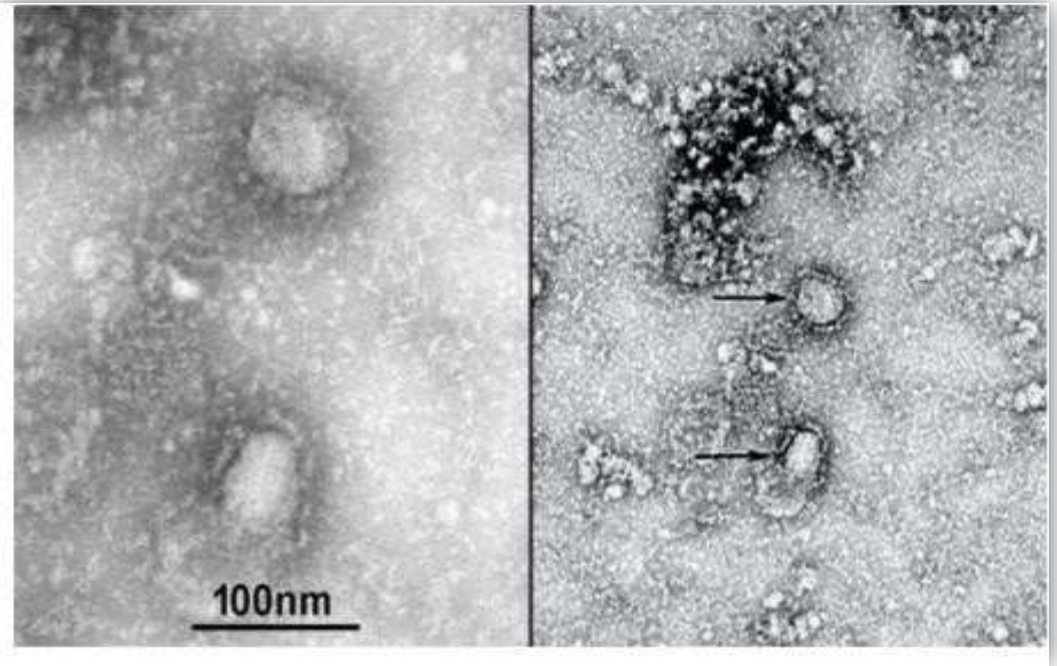



**Figure 1:** Coronavirus observed under an electron microscope (×100,000**).( Photo via China's Center for Disease Control) and**[9]**.**

## 5. SARS-CoV-2 Viruses

New and old viruses are a big risk to health all over the world. Near the end of 2019, China shared news about many people in Wuhan with bad lung sickness, now known as COVID-19. They had signs like hot body and hard breath, and had lung virus.

With the whole gene map, it was clear that a new virus, named 2019-nCoV by WHO, was the cause. Later, the virus name was set as SARS-CoV-2 by the ICTV **(Elakeb, N.H., 2021).**

The three big known viruses of this type are SARS-CoV (SARS), MERS-CoV (MERS), and SARS-CoV-2 (COVID-19)[9]. Even though they are from the same group, these viruses are not the same in build, start, and the sickness they bring **(Seshalyna, 2020)**

**Table 1:** First Appearances of Coronavirus Families Worldwide **(Seshalyna, 2020).**

| **Virus** | **Year of Emergence** | **Country** | **Disease** |
|---|---|---|---|
| **SARS-CoV** | 2002–2004 | China | Severe Acute Respiratory Syndrome (SARS) |
| **MERS-CoV** | 2012 | Middle East | Middle East Respiratory Syndrome (MERS) |
| **SARS-CoV-2** (COVID-19) | 2019 | China | Coronavirus Disease 2019 (COVID-19) |

### 6. Origin of the SARS-CoV-2 Virus

The disease COVID-19 is from SARS-CoV-2, a new virus that may have come from Wuhan, China, and could link to a live animal place**.** At the start, it was thought bats may have held the virus first, as close matches were seen in the genes between SARS-CoV-2 and a bat virus, BetaCoV/RaTG13/2013[10]**.**

We don't yet know the in-between host that passed the virus to people[3]. Some reports say, "Touching the in-between animal hosts, or eating wild animals might have been the main way the virus moved. But, the true starts and ways it moves are still not clear (Fig. 2).

More deep looks into the matter found that SARS-CoV-2 is an RNA virus[9]. It links to the same virus group that caused the Severe Acute Respiratory Syndrome (SARS, 2003) and Middle East Respiratory Syndrome (MERS, 2012)[10].

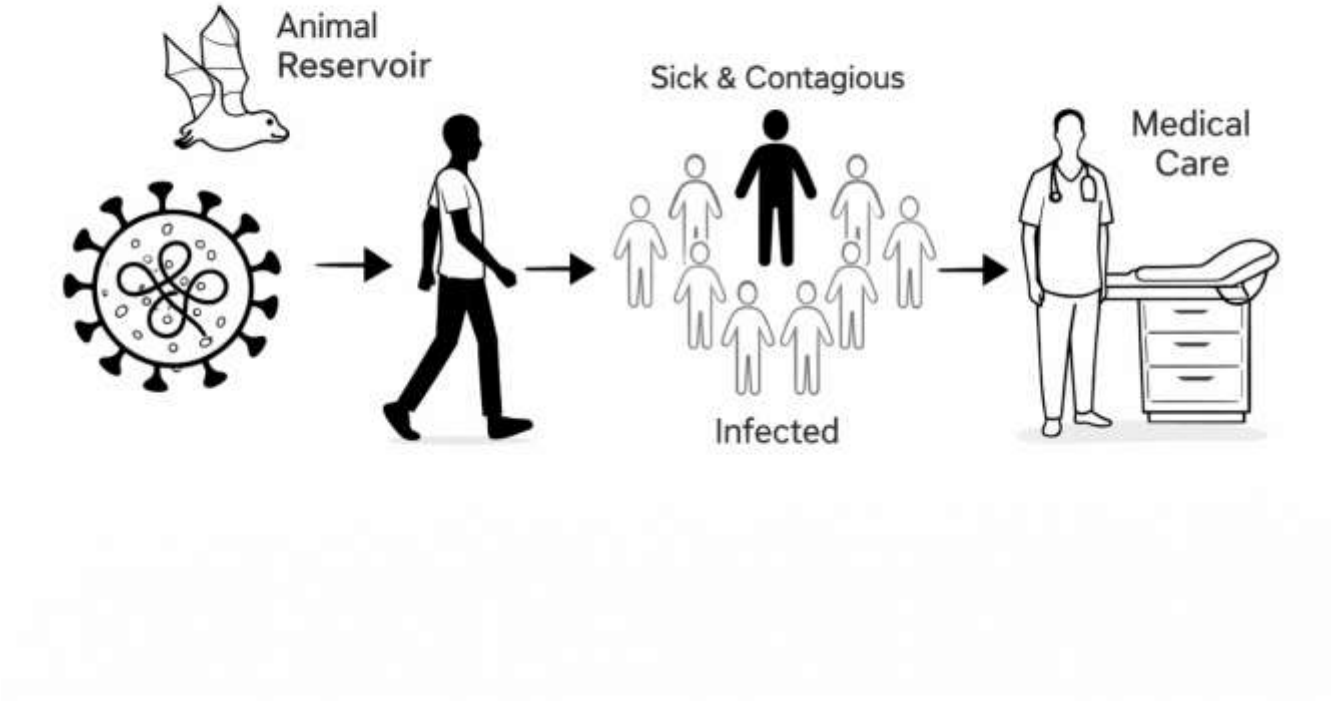



**Figure 2:** The emergence and transmission of coronaviruses to humans[11].

### 7. Structure of SARS-CoV-2[12]

The SARS-CoV-2 virus is microscopic and invisible to the naked eye. It is a spherical, enveloped viral particle with a diameter ranging from 60 to 220 nm, and its surface is covered with spike-like protrusions forming a crown-like structure, as indicated by its name **(Figure 3.A).**

The viral envelope consists of a lipid bilayer derived from the host cell. Embedded within this envelope are four types of structural proteins **(Figure 3.B):**

- **Spike (S) proteins:** Responsible for binding to host cell receptors, specifically the ACE2 receptor, which is present on epithelial cells of the bronchi and pulmonary alveoli **(Naoum, S., 2021).**
- **Membrane (M) proteins:** Contain three transmembrane domains and determine the virus's shape.
- **Envelope (E) proteins:** Play a role in the assembly and release of virions from infected cells.
- **Nucleocapsid (N) proteins:** Bind to the viral genome, forming a ribonucleoprotein complex.

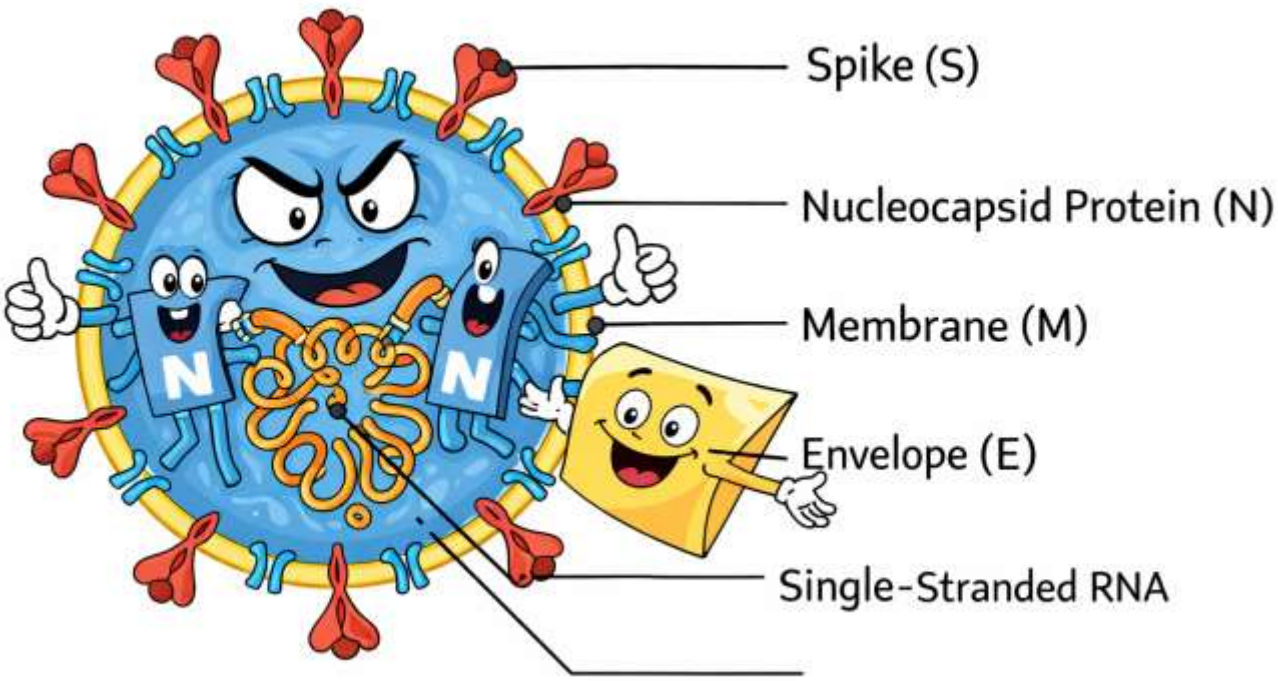



**Figure 3:** Representation of the SARS-CoV-2 virus[12]**.**

**8. Genome of SARS-CoV-2[12]**

**A. The Genome**

SARS-CoV-2 is an enveloped, positive-sense single-stranded RNA virus (+ssRNA) with a genome size of approximately 29.9 kb [4]**.** Two-thirds of the viral genome encodes a large replicase polyprotein (ORF1a and ORF1b), which is translated into two precursor polyproteins.

These polyproteins are subsequently cleaved into sixteen non-structural proteins (NSPs) essential for viral replication **.** The remaining third of the genome primarily encodes the structural proteins, including four membrane glycoproteins[4], [13]: Spike (S), Hemagglutinin-Esterase (HE), Membrane (M), and Envelope (E), as well as the Nucleocapsid (N) protein.

**2. Viral Structure[13]**

The nucleocapsid has a helical conformation, formed by the N protein complexed with the viral RNA. It is enclosed within a phospholipid envelope that embeds the viral surface glycoproteins (S, HE, M, and E).

The Spike (S) protein plays a crucial role in viral entry by binding to the Angiotensin-Converting Enzyme 2 (ACE2) receptor on host cells. It consists of two subunits:

- S1 subunit, which contains the Receptor Binding Domain (RBD) responsible for ACE2 recognition.
- S2 subunit, which mediates the fusion of the viral and host cell membranes, facilitating viral entry into the cell **.**

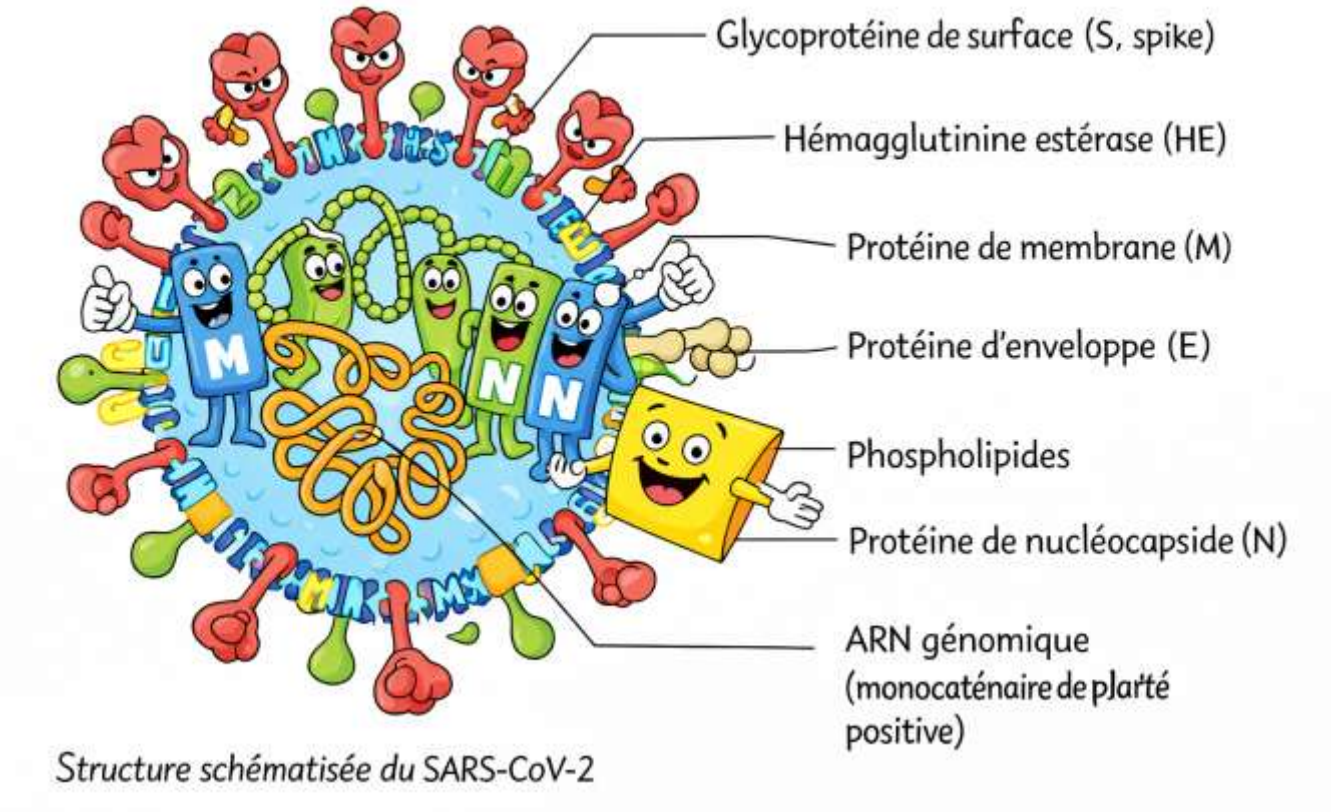



**Figure 4: Schematic structure of SARS-CoV-2.**[12], [13]

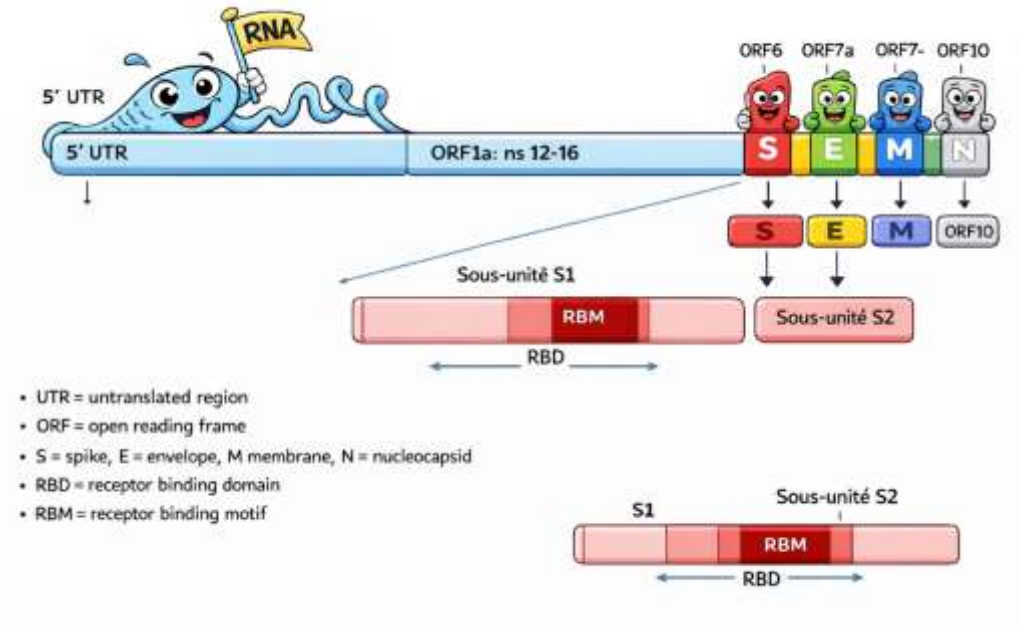



**Figure 5:** Schematic structure of the SARS-CoV-2 genome (29,903 nucleotides)[12], [13].

The primary receptor for SARS-CoV-2 is Angiotensin-Converting Enzyme 2 (ACE2), which is widely expressed on the surface of various human cells. It is predominantly found in epithelial cells of the nasal cavity, trachea, bronchi, and bronchial serous glands, as well as in alveolar cells of the lungs, monocytes, and alveolar macrophages. Additionally, ACE2 is expressed in endothelial cells and smooth muscle cells of blood vessels, enterocytes of the small intestine, epithelial cells of renal tubules, and neurons. This widespread distribution explains the multisystemic impact of SARS-CoV-2 infection in humans[14]**.**

**Conclusion**

Given these points, it is clear that proficiency in programming and bioinformatics—especially tools for genetic sequence analysis—has shifted from an optional skill to an essential requirement in contemporary virology and epidemiological studies. These technologies form the foundation for addressing unresolved questions about viruses like SARS-CoV and SARS-CoV-2, including their mutation patterns, transmission pathways, and origins. Computational genomics allows researchers to reassess and test competing theories on the zoonotic sources of SARS-related viruses, whether from bats, cats, or pigs **(Zhou et *al.*, 2020; Lam et al., 2020).** Bioinformatics further enables the reconstruction of coronavirus evolution, the identification of critical mutations

(e.g., in the Spike protein), and precise cross-strain comparisons**.**

These genomic insights also bridge disciplines, such as food security, since many zoonotic viruses emerge from animals in the human food chain. Understanding these links is vital not only for pandemic prevention but also for safeguarding global food systems**.**

Crucially, integrating bioinformatics and genomics advances precision medicine—an approach that tailors treatments to individual patients using genetic, environmental, and lifestyle data. Unlike conventional methods, precision medicine delivers targeted therapies based on molecular profiles, improving efficacy and minimizing adverse effects.

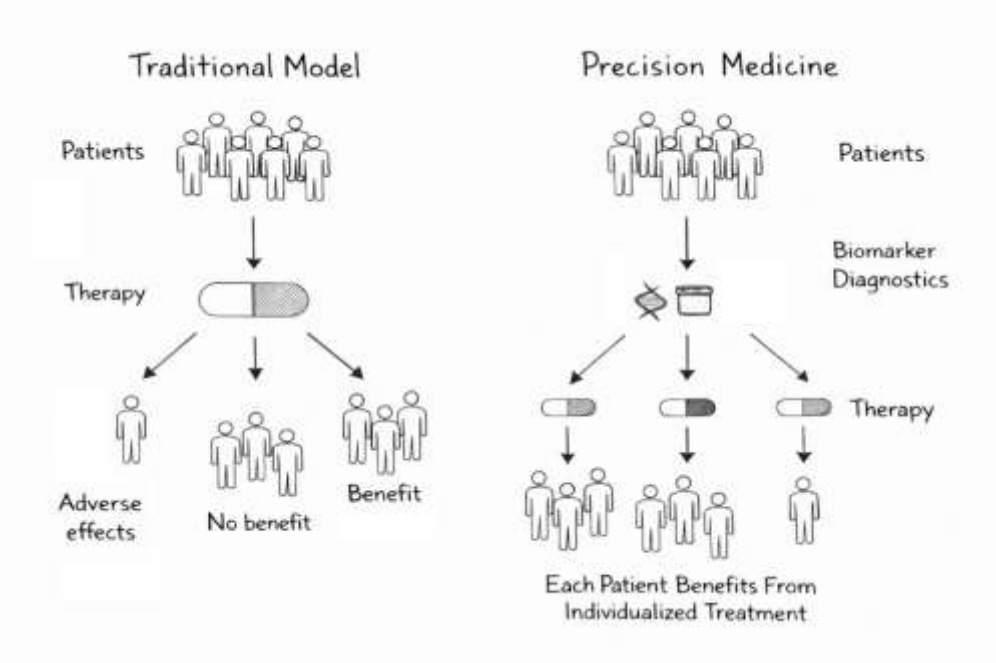



**Figure 6:** Comparison Between Traditional Medicine and Precision Medicine Approaches

Investing in bioinformatics training, particularly for early-career scientists and medical professionals, provides the tools to predict, prevent, and personalize responses to evolving pathogens. Such preparedness is indispensable not just for future outbreaks but also for strengthening global health strategies across virology, epidemiology, and biotechnology.

# **Chapter 2:** Literature Review on Coronavirus

## Introduction

The emergence of SARS-CoV in 2002 and its re-emergence in the form of SARS-CoV-2 in 2019 have raised urgent questions about the origin of coronaviruses and the mechanisms of their transmission to humans. These questions cannot be fully addressed through classical virology alone. Instead, they require an interdisciplinary approach that combines biological sciences with computational tools, in what is now widely referred to as bioinformatics[6], [14].

Bioinformatics enables researchers to analyze massive genetic datasets, reconstruct evolutionary relationships, and identify mutation patterns with unprecedented precision. The integration of genetics, molecular biology, and computer science has proven crucial in tracing the phylogenetic history of SARS-CoV-related viruses and in hypothesizing how such pathogens may have jumped from animals to humans.

Numerous studies suggest that bats serve as natural reservoirs of SARS-like coronaviruses[15]**,** while other animals, including civets, pangolins, swine, and domestic cats, may act as intermediate hosts in zoonotic transmission[2]**.** Despite intensive research, the precise origin of SARS- CoV-2 remains a topic of scientific debate, with multiple hypotheses still under investigation.

Therefore, mastering bioinformatics tools is no longer optional for virologists or public health researchers. These tools are essential for:

- Re-examining and verifying existing hypotheses about the origin of SARS-CoV.
- Investigating the potential role of various animal species in viral transmission.
- Laying the foundation for predictive models in pandemic preparedness and precision medicine.

In this chapter, we explore how the synergy between biology and informatics provides a powerful framework to address the fundamental question: How did SARS-CoV and SARS-CoV-2 emerge in the human population, and from what source?

## 1. From Animals to Humans: The Origin of "SARS-CoV"

The zoonotic transmission of SARS-CoV (Severe Acute Respiratory Syndrome Coronavirus) from animal reservoirs to humans is supported by robust multidisciplinary evidence, including comparative genomic analyses, epidemiological tracing, and ecological investigations[10], [15].

Phylogenetic studies reveal close genetic relationships between SARS-CoV and coronaviruses circulating in bat and pangolin populations, suggesting cross-species spillover events. Epidemiological data from early outbreaks further indicate direct or intermediate animal contact as a likely transmission route. Additionally, ecological studies highlight the role of wildlife trade and habitat encroachment in facilitating viral emergence.

Collectively, these findings provide a compelling scientific foundation for the animal-to-human origin hypothesis of SARS-CoV, underscoring the importance of One Health approaches in pandemic prevention[11].

The emergence of the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) in 2002 marked a pivotal moment in the study of emerging infectious diseases. Mounting virological, epidemiological, and ecological evidence supports the theory that SARS-CoV originated in wild animals before crossing the species barrier to infect humans — a process commonly referred to as zoonotic spillover[3]**.**

### 1.1. Zoonotic Origin and Reservoir Host: Bats[15]

Extensive genomic and phylogenetic analyses have established bats (order Chiroptera) as the primary evolutionary reservoir for SARS-related coronaviruses (SARSr-CoVs). Among these, viruses detected in Rhinolophus spp. (horseshoe bats) exhibit particularly high genomic homology with human SARS-CoV strains. Notably, the bat coronavirus RaTG13 demonstrates 96.1% whole-genome sequence identity with SARS-CoV-2, with pronounced conservation in key functional domains, including the receptor-binding motif (RBM) of the spike protein**.** This striking genetic congruence, coupled with the absence of a direct epidemiological link to alternative intermediate hosts, provides compelling evidence for a bat progenitor origin.

Metagenomic surveillance across Southeast Asian bat populations has further revealed an extensive diversity of SARSr-CoVs, characterized by frequent

recombination events and adaptive mutations in the spike protein's receptor-binding domain (RBD) **.** Such features enhance viral fitness for cross- species transmission by maintaining affinity for angiotensin-converting enzyme 2 (ACE2) orthologs across mammalian hosts. These findings collectively underscore the role of bats as both a natural reservoir and an evolutionary crucible for SARS- like coronaviruses, with spillover risk amplified by ecological overlap at the human-wildlife interface.

### 1.2. Intermediate Host: Civet Cats[10], [16], [17]

Although bats constitute the primordial reservoir for SARS-related coronaviruses (SARSr-CoVs), epidemiological evidence suggests that direct bat-to-human transmission is improbable due to ecological separation and insufficient viral adaptation to human ACE2 receptors. Instead, phylogenetic and outbreak tracing data implicate palm civets (Paguma larvata) as critical intermediate hosts during the 2002–2003 SARS-CoV emergence. This conclusion is supported by the isolation of civet-derived viral strains (e.g., SZ3 and HC/SZ/61/03) exhibiting >99.8% nucleotide sequence identity with early human outbreak strains across structural genes, including the spike (S) protein **.** Notably, civet- associated variants contained key mutations in the receptor-binding domain (RBD), such as N479K and T487S, which enhanced ACE2 binding affinity in humans [11].

Experimental and field observations further substantiate the intermediary role of civets. Viral challenge studies demonstrated robust viral replication in civet respiratory and gastrointestinal tracts, correlating with high viral loads in naturally infected market animals**.** Pathological analyses revealed acute pulmonary lesions in civets, suggesting active infection rather than passive carriage. Critically, the high-density confinement of civets in live-animal markets created an ideal environment for viral amplification and adaptive evolution. Such conditions likely facilitated the selection of RBD mutations that overcame host barriers, culminating in zoonotic spillover[17]**.** These findings collectively highlight the synergistic roles of host biology (civet ACE2 compatibility), viral evolution (RBD optimization), and anthropogenic factors (wildlife trade practices) in bridging zoonotic transmission.

### 1.3. Summary of Transmission Pathway Hypothesis

The hypothesized transmission pathway of SARS-CoV involves a zoonotic origin beginning with bats, which are considered the natural reservoir of the virus. From bats, the virus is believed to have infected civet cats, serving as an intermediate host where the virus could amplify and adapt. Finally, the virus spilled over to humans, leading to outbreaks of infection. This transmission chain highlights the importance of interspecies interactions in the emergence of zoonotic diseases **.**

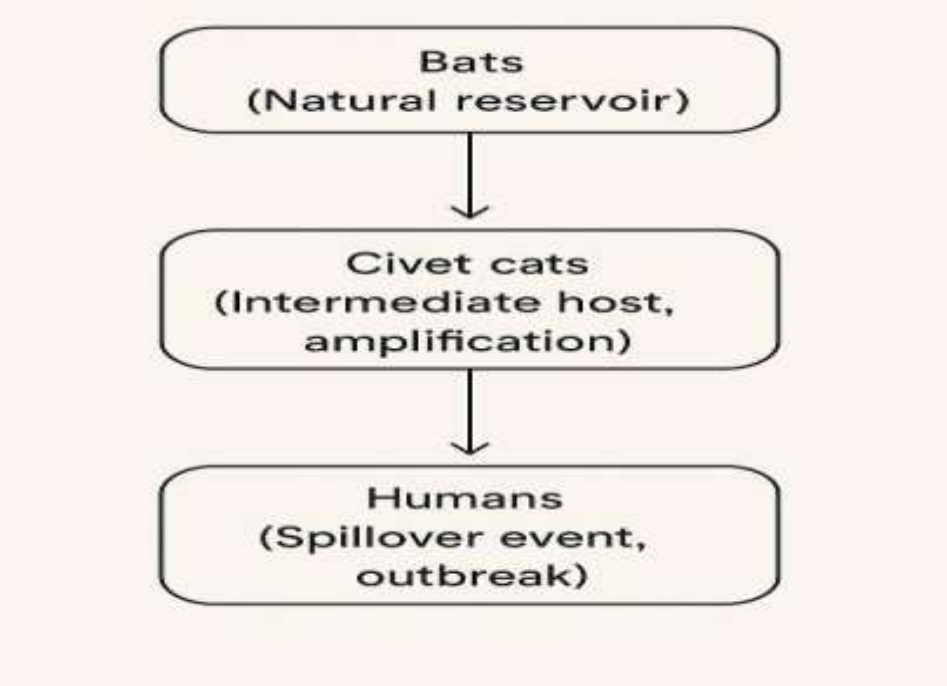



**Figure 7:** Transmission Pathway of SARS-CoV from Bats to Humans

## 2. Proposed Zoonotic Transmission Route of SARS-CoV: From Pigs to Humans

Swine [18] have long been recognized as potential "mixing vessels" for viral reassortment, particularly concerning influenza A viruses, due to their susceptibility to both avian and human viral strains. However, their involvement in the transmission dynamics of SARS-CoV appears to be more limited and less understood. Although the possibility of cross-species transmission cannot be entirely ruled out, current virological evidence indicates that porcine cells exhibit restricted permissiveness to SARS-CoV infection. This limited susceptibility is mainly attributed to inefficient binding between the viral spike (S) protein and the porcine angiotensin-converting enzyme 2 (ACE2) receptor.

Experimental in vitro analyses have shown that primary porcine airway epithelial cells support significantly lower levels of SARS-CoV replication—often not exceeding 1 $\log_{10}$ $TCID_{50}$/mL—when compared to human or civet cell cultures**.**

## 2.1. SARS-CoV and Pigs: Assessing the Zoonotic Link

The potential for pigs to act as vectors for zoonotic coronaviruses has been explored due to their known role as "mixing vessels" in the emergence of novel influenza viruses**.** However, regarding SARS- CoV, the current scientific consensus strongly indicates that pigs were neither direct nor indirect sources of human infection.

Epidemiological and molecular evidence points to bats as the natural reservoir of SARS-CoV and civet cats as the intermediate hosts responsible for transmitting the virus to humans **.** Experimental studies have shown that porcine cells are not permissive to efficient SARS-CoV replication, primarily due to low-affinity interactions between the viral spike (S) protein and the porcine ACE2 receptor**.**

Despite this, pigs are known to host several porcine-specific coronaviruses, which primarily affect their respiratory and gastrointestinal systems but do not infect humans under current evidence. Among these:

- Porcine Epidemic Diarrhea Virus (PEDV) was first identified in England in 1971 and formally characterized as a distinct coronavirus in 1978.
- Porcine Respiratory Coronavirus (PRCV) was first detected in Belgium in 1984 during outbreaks of mild respiratory illness in swine[19]**.**

### 2.1.1. Overview of Major Porcine Coronaviruses

Several coronaviruses have been identified in swine populations, primarily belonging to the Alphacoronavirus and Deltacoronavirus genera. Among the Alphacoronaviruses, the Porcine Epidemic Diarrhea Virus (PEDV), first identified in the 1970s, causes severe diarrhea and dehydration in piglets**.** Another important Alphacoronavirus is the Transmissible Gastroenteritis Virus (TGEV), known for its high mortality rates in young pigs **.**

The Porcine Respiratory Coronavirus (PRCV), a variant of TGEV, emerged in the 1980s and primarily affects the upper respiratory tract (Pensaert et al., 1986). More recently, the Swine Acute Diarrhea Syndrome Coronavirus (SADS-CoV) has been detected in China, linked to bat-origin spillover events **(Zhou et *al*.,).**

The Porcine Delta Coronavirus (PDCoV), part of the Deltacoronavirus genus, was discovered in Hong Kong in 2012 and later associated with outbreaks in the U.S.**.** Despite their diversity, none of these porcine coronaviruses have been proven to infect humans.

### 2.1.2. Emerging Porcine Coronaviruses: Zoonotic Potential and Surveillance Implications

Recent virological investigations have identified two porcine coronaviruses with demonstrated cross-species potential, warranting enhanced surveillance under the One Health framework:

**1. Swine Acute Diarrhea Syndrome Coronavirus (SADS-CoV)**

First isolated during 2016–2017 outbreaks in Guangdong pig farms, this alphacoronavirus shares 96%–98% genomic identity with bat HKU2-like coronaviruses, confirming a zoonotic origin**.**

While causing high mortality (>90%) in neonatal piglets through severe enteritis, in vitro studies reveal concerning tropism for human cell lines, including primary hepatocytes and airway epithelia**.** The virus utilizes the human aminopeptidase N (hAPN) receptor for entry, a conserved pathway across mammalian species. Despite these molecular prerequisites, no natural human infections have been epidemiologically confirmed to date.

**2. Porcine Deltacoronavirus (PDCoV)**

Initially detected in **Hong Kong swine (2012)** before spreading globally, PDCoV exhibited unexpected zoonotic capacity in a 2021 study reporting PCR-positive samples from febrile Haitian children**.** Key findings include:

Sequence Evidence: Identical spike protein mutations in human and porcine strains

- **Controversies:** Lack of serological confirmation and possible sample contamination debates

- Experimental Data: Demonstrated replication competence in human intestinal organoids

While neither virus currently represents an established public health threat, their evolutionary trajectories warrant monitoring given three risk factors:

- **Genetic promiscuity:** Frequent recombination events in swine enteric coronaviruses
- **Agricultural practices:** High-density farming as an amplifier of viral evolution
- **Interface exposure:** Increasing human-livestock contact in endemic regions

**Conclusion**

In conclusion, while pigs have not yet been confirmed as significant hosts for SARS-related coronaviruses in humans, their biological compatibility with various viruses, including influenza and some coronaviruses, underscores the importance of continued vigilance**.** The documented genetic plasticity of porcine coronaviruses—through mutation and recombination—highlights the potential for future zoonotic emergence**.** Therefore, ongoing surveillance programs, molecular monitoring, and robust early warning systems in swine populations are vital strategies to mitigate the risk of interspecies spillover and ensure global health preparedness.

The comparative analysis of SADS-CoV and PDCoV underscores the diversity of porcine coronaviruses in terms of origin, pathogenicity, and zoonotic potential. SADS-CoV, which originated from bat coronaviruses, caused high-mortality outbreaks in piglets and raised public health concerns due to its ability to infect human cells in vitro **(Zhou et *al*., 2018).** PDCoV, although less virulent, has demonstrated possible zoonotic transmission; notably, a novel PDCoV strain was detected in febrile Haitian children, raising questions about cross-species transmission **.**

The emergence and characterization of Feline Coronavirus (FCoV) illustrate key principles in viral evolution and pathogenesis. Initially recognized in the 1960s in association with feline infectious peritonitis (FIP), FCoV was later classified into two distinct biotypes: the relatively benign Feline Enteric Coronavirus (FECV) and the lethal FIP virus (FIPV) . This dichotomy underscores the capacity of coronaviruses to undergo pathogenic divergence within a single host species. The timeline of FCoV research highlights milestones in veterinary virology, from clinical identification (1963–1966) to etiological confirmation (1970s) and biotype classification (1981). These findings emphasize the importance of continued surveillance for coronavirus variants in animal populations, as such studies not only address veterinary health concerns but also contribute to our broader understanding of coronavirus biology and interspecies transmission potential **.**

# **Chapter 3:** Bioinformatics Tools for Coronavirus Genome Analysis

## Introduction

**BioEdit** is a widely utilized software application designed for the editing, alignment, and analysis of biological sequences. It is particularly valued in the fields of virology, molecular biology, and genomics due to its intuitive interface and versatile toolset. Researchers working with coronavirus genomes, such as SARS-CoV and SARS-CoV-2, frequently employ BioEdit for tasks including multiple sequence alignment, mutation tracking, and primer design. Although BioEdit is no longer actively updated, it remains a preferred option for preliminary genomic assessments and educational purposes due to its lightweight design and ease of use [20](Hall, 1999). Its integration with external tools such as ClustalW makes it especially useful for aligning viral genome segments and analyzing evolutionary relationships.

## 1.Step-by-Step Workflow for Genomic Analysis of SARS-CoV/SARS-CoV-2 Using BioEdit

### 1.1.Download and Install BioEdit

**BioEdit** is a legacy sequence alignment software compatible with Windows OS. Although no longer under active development, it remains available through archival websites and continues to serve as a foundational tool in molecular biology for handling and editing nucleotide or protein sequences (Hall, 1999). Installation involves downloading the executable installer and following the standard setup process.

### 1.2.Retrieval of SARS-CoV or SARS-CoV-2 Genomic Sequences

SARS-related coronavirus sequences can be accessed via public genomic databases such as NCBI GenBank (https://www.ncbi.nlm.nih.gov/genbank/) and GISAID (https://www.gisaid.org/), which host curated and annotated sequence data. Researchers typically retrieve sequences in FASTA format to ensure compatibility with alignment and analysis software (**Shu & McCauley, 2017).**

### 1.3.Importing Sequences into BioEdit

To begin analysis, open BioEdit and import your sequence data by navigating to File → Open and selecting the relevant .fasta file. The sequences will populate the editing window, where they can be visualized, annotated, or aligned.

### 1.4. Sequence Editing and Alignment

Users may perform manual edits (e.g., correcting ambiguous bases) or use BioEdit's built-in integration with ClustalW for multiple sequence alignment (Alignment → ClustalW Multiple Alignment). This is particularly useful when comparing homologous regions across different strains, such as the Spike (S) protein of SARS-CoV and SARS-CoV-2.

### 1.5. Nucleotide Translation to Amino Acids

To examine mutations at the protein level, nucleotide sequences can be translated using the Nucleic → Translate Frame function. This is essential for identifying key amino acid changes, especially within functional domains like the receptor-binding domain (RBD) of the S protein, which is crucial for host receptor interaction[21].

### 1.6. Mutation Analysis and Annotation

Regions of interest (e.g., coding sequences or conserved motifs) can be annotated or highlighted manually. Comparative analysis with a reference strain, such as Wuhan-Hu-1 (NC_045512), facilitates the identification of single nucleotide polymorphisms (SNPs) or insertions/deletions (indels)[4].

### 1.7. Exporting Results for Downstream Applications

Once alignments are finalized, BioEdit allows exporting in multiple formats—FASTA, GenBank, or NEXUS—for further use in phylogenetic tools like MEGA, BEAST, or PhyML. Tables summarizing mutations can also be generated for reporting or statistical analysis.

## 2. Applications of BioEdit in SARS-CoV Research

BioEdit offers several functionalities that are particularly useful in the genomic investigation of SARS-CoV and related coronaviruses. The table below summarizes its main applications in virological research:

**Table 2:** Main applications of BioEdit in the genomic analysis of SARS-CoV-2

| Research Task | Relevant BioEdit Function |
|---|---|
| **Sequence alignment** | Multiple sequence alignment using ClustalW or manual adjustment |
| **Mutation analysis** | Visual comparison of nucleotide or codon sequences |
| **Open Reading Frame (ORF) identification** | Use of built-in translation tools to identify coding regions |
| **Spike (S) gene analysis** | Translation and alignment of amino acid sequences to detect mutations in key domains (e.g., RBD) |
| **Data export for phylogenetic tools** | Export in FASTA or NEXUS format for downstream use in MEGA, PhyML, or IQ-TREE |

These functions allow researchers to investigate genetic variability, identify critical mutations, and perform comparative genomic analysis across different coronavirus strains. For instance, aligned and annotated sequences from BioEdit can be exported to MEGA X, IQ-TREE, or other phylogenetic platforms to construct evolutionary trees and assess relationships between viral isolates[22].

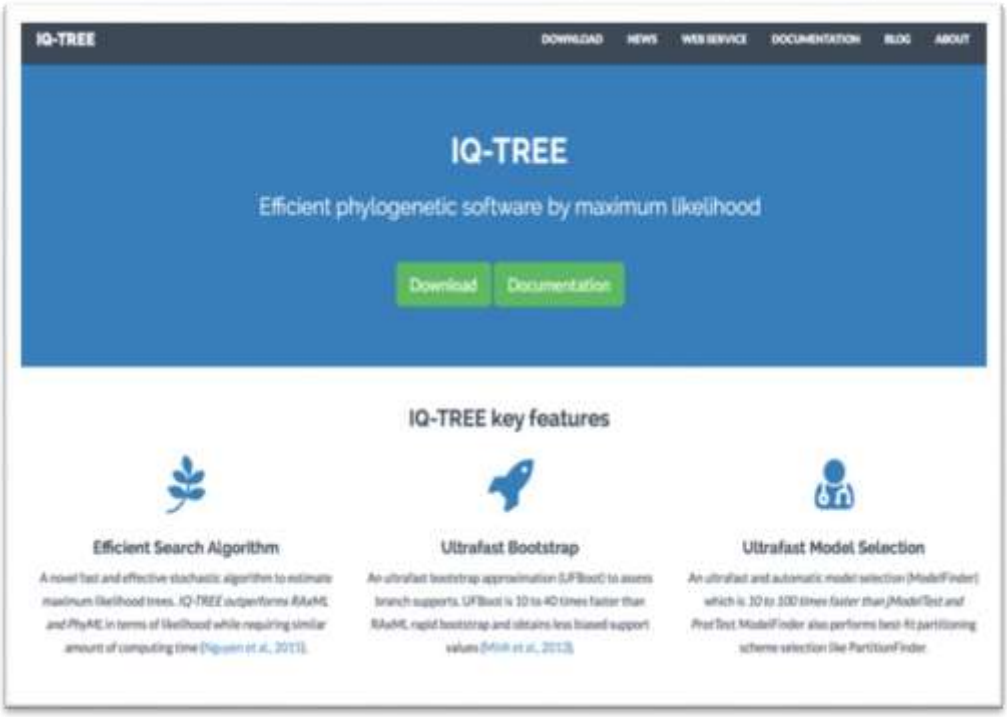



**Figure 8 :**IQ-TREE Web Portal A Next-Generation Platform for Phylogenetic Analysis[23]

### 3.Practical Application of BioEdit Software in SARS-CoV Genomic Analysis

**BioEdit** is a free biological sequence editor that is widely used in virology, molecular biology, and genomics to edit, align, and analyze nucleotide and protein sequences. It is especially helpful for comparing viral genomes, as those of SARS-CoV and SARS-CoV-2[20]. This software has an easy-to-use interface that lets you do a number of important bioinformatics tasks, like multiple alignment, reading frame translation, mutation identification, and exporting sequences for later phylogenetic analyses[20].

### 3.1.Steps of genomic analysis with BioEdit

#### 3.1.1.Downloading and installing BioEdit

Can obtain BioEdit from archive sites or specialist repositories (Hall, 1999). It works with Windows. It's necessary to check that the version you installed supports alignment functions through ClustalW.

### 3.1.2.Obtention des séquences génomiques virales

The sequences of SARS-CoV and SARS-CoV-2 can be obtained from public databases such as:

•GenBank NCBI : https://www.ncbi.nlm.nih.gov/genbank/

•GISAID SARS-CoV-2 : https://www.gisaid.org

The sequences are generally saved in FASTA format (.fasta or .fa), a format recognized by BioEdit.

### 3.1.3.Importation of sequences to BioEdit

Once the sequences are downloaded, the user can open them via the File → Open option. The sequences then appear in the sequence editor, ready for alignment or analysis.

### 3.1.4.Alignement

BioEdit lets you do manual sequence alignment or utilize the built-in ClustalW tool (Alignment → ClustalW Multiple Alignment), as long as the ClustalW program is installed correctly. This stage is very important for finding similarities and differences between different strains, such SARS-CoV and SARS-CoV-2.

### 3.1.5.Translation of nucleotides into amino acids

To analyze mutations in key proteins (e.g., Spike protein), BioEdit offers the translation function using Nucleic → Translate Frame. This allows for the transition from the nucleotide level to the protein level, facilitating the study of functional domains..

### 3.1.6.Analysis of mutations and variations

The user can manually mark areas of interest, like the receptor-binding domain (RBD), and then compare the sequences to a reference strain, like Wuhan-Hu-1 for SARS-CoV-2.

### 3.1.7.Exporting the results

Can export the alignments or changed sequences in FASTA, GenBank, or NEXUS formats, which makes it easier to use them with phylogenetic analysis applications like MEGA X or IQ-TREE[22].

## 4.Functional Applications of BioEdit in SARS-CoV Genomic Research

BioEdit is a widely utilized tool in genomic studies, particularly for sequence editing and alignment. Its integration with other software like ClustalW and MEGA makes it highly suitable for molecular investigations of SARS-CoV and SARS- CoV-2.The following table summarizes core functionalities relevant to coronavirus research, especially in comparative genomics and phylogenetic inference.

**Table 3:** Core Analytical Tasks Enabled by BioEdit for SARS-CoV-2 Genomic Studies

| Analytical Task | BioEdit Feature |
|---|---|
| **Multiple Sequence Alignment** | Uses ClustalW interface or manual adjustment to align nucleotide or protein sequences **(Hall, 1999).** |
| **Mutation Detection** | Enables visual inspection and codon-based comparisons across SARS-CoV variants. |
| **Open Reading Frame (ORF) Identification** | Translation tools assist in detecting potential coding regions (frames) within viral genomes. |
| **Spike (S) Protein Analysis** | Translation and alignment tools facilitate the study of amino acid variation in the spike gene. |
| **Data Export for Phylogenetics** | Export options to FASTA, NEXUS, or GenBank formats for use in MEGA, PhyML, or IQ-TREE software (Tamura et al., 2021). |

# **Chapter 04:** Results and Discussion

This chapter outlines the key steps for conducting genomic analysis, including how to work with BioEdit and retrieve DNA sequences from genomic databases such as NCBI.

**1.Downloading BioEdit Software**

To download and install BioEdit, begin by searching for "BioEdit software" on Google. From the search results, click on the link labeled "BioEdit 7.2", which will direct you to the official download page. Once there, locate and select the blue "Download" button to obtain the installation file. After the download completes, run the setup file and follow the on-screen instructions to install the software on your computer.

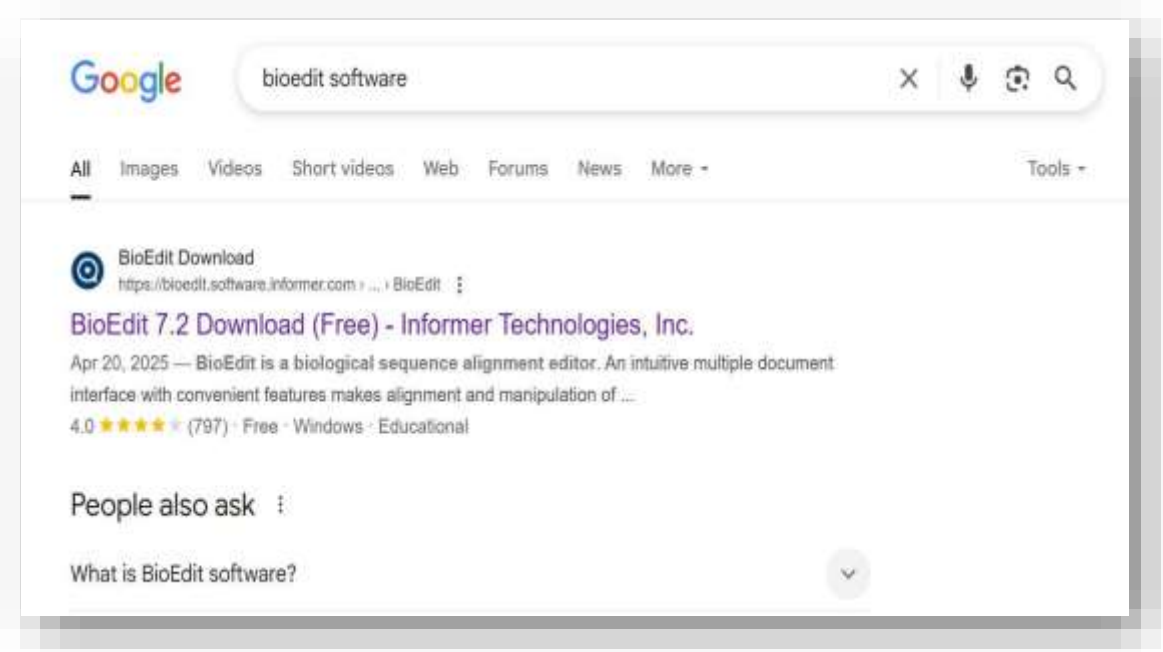



**Figure 09:** Search result on Google for "BioEdit software" showing the official download link (BioEdit 7.2) from Informer Technologies, Inc.

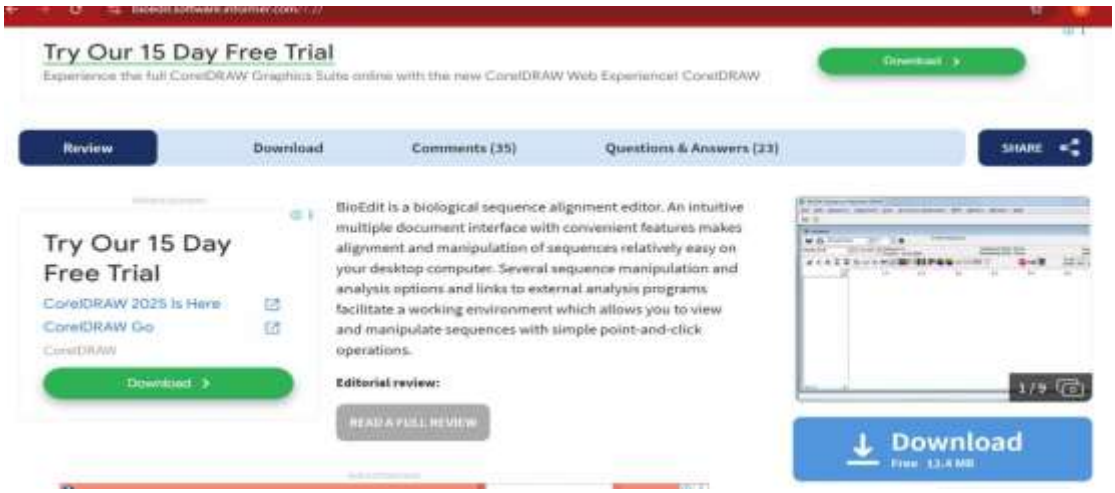



**Figure 10 :** Official BioEdit 7.2 download page on Informer Technologies, Inc. website, showing the download button and software description.

**2.Coronaviruses samples**

This study analyzed a selection of coronaviruses relevant to SARS-CoV and SARS-CoV-2 research, chosen for their diversity in host species, geographic distribution, and isolation timelines. The accompanying table 4 provides details on each virus, including its designation, natural host, spatiotemporal origin, and corresponding NCBI accession number for genomic reference.

**Table 4:** Description of the Coronaviruses Used in This Study

| Coronaviruses | The host | Year | NCBI Accession number |
|---|---|---|---|
| Wuhan-Hu-1 | Human SARS-COV2 | 2019-China | NC_045512.2 |
| Guangzhou SARS GZ02 | Human SARS-COV | 2003-China | AY390556.1 |
| Feline | Feline | 2005-USA | AY994055 |
| Bat RaTG13 | Bat | 2020-China | MN996532.2 |
| Pangolin | Pangolin | 2019-03-China | MT121216.1 |
| Porcine | Porcine | 1986-China | JN547228 |
| Bat Yunnan2011 | Bat | 2011-China | JX993988 |

The analyzed coronavirus strains include: Wuhan-Hu-1 (human SARS-CoV-2, China 2019, NC_045512.2); Guangzhou SARS GZ02 (human SARS-CoV, China 2003, AY390556.1);

Feline Coronavirus (USA cats 2005, AY994055); Bat RaTG13 (China 2020, MN996532.2); Pangolin Coronavirus (China 2019, MT121216.1); Porcine Coronavirus (China pigs 1986, JN547228); and Bat Yunnan2011 (China 2011, JX993988). These were selected for their relevance to SARS-CoV-2 research and evolutionary diversity.

**3.Main steps to obtain Coronaviruses samples**

To obtain coronavirus genomic sequences, researchers can use public genomic databases such as the NCBI website. By using the accession numbers found during the literature review, the desired sequence can be retrieved. For example, after accessing the NCBI homepage, the accession number “AY390556.1” can be entered into the search bar, as shown in the next figure.

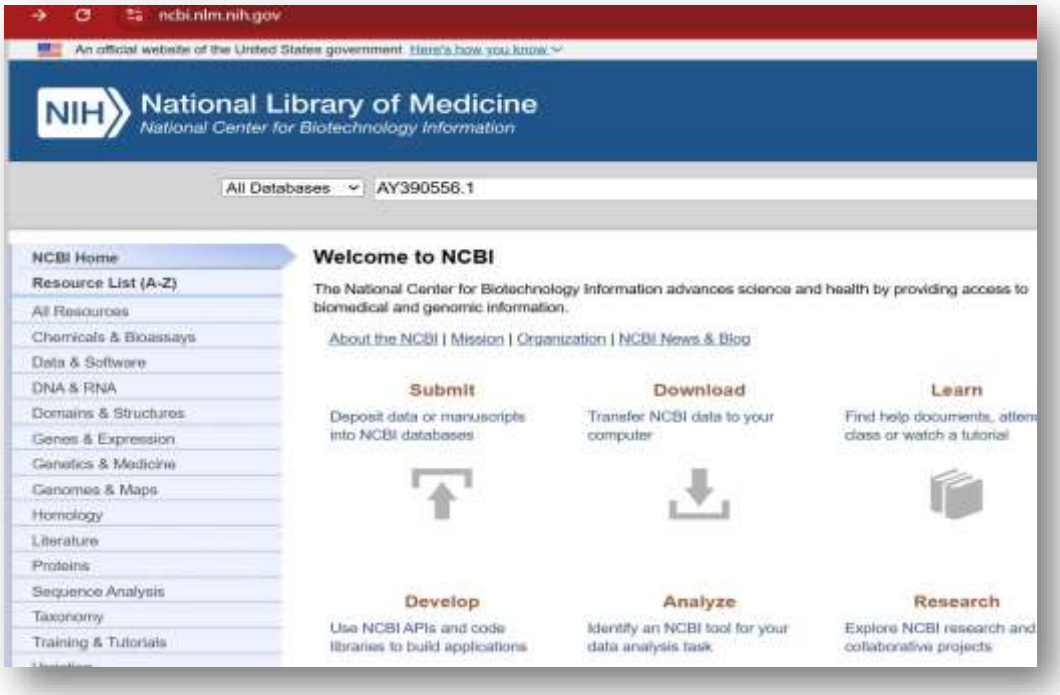



**Figure 11:** Homepage of the NCBI website with the accession number “AY390556.1” entered in the search bar.

The search yields five results. Click the first blue-highlighted entry labeled "SARS coronavirus GZ02" to access the sequence record for AY390556.1. The resulting page displays comprehensive metadata including sequence definition, collection date/location, related publication, and sample details.


**Figure 12:** Search results on the NCBI website showing the complete genome of SARS coronavirus GZ02 (accession number AY390556.1).

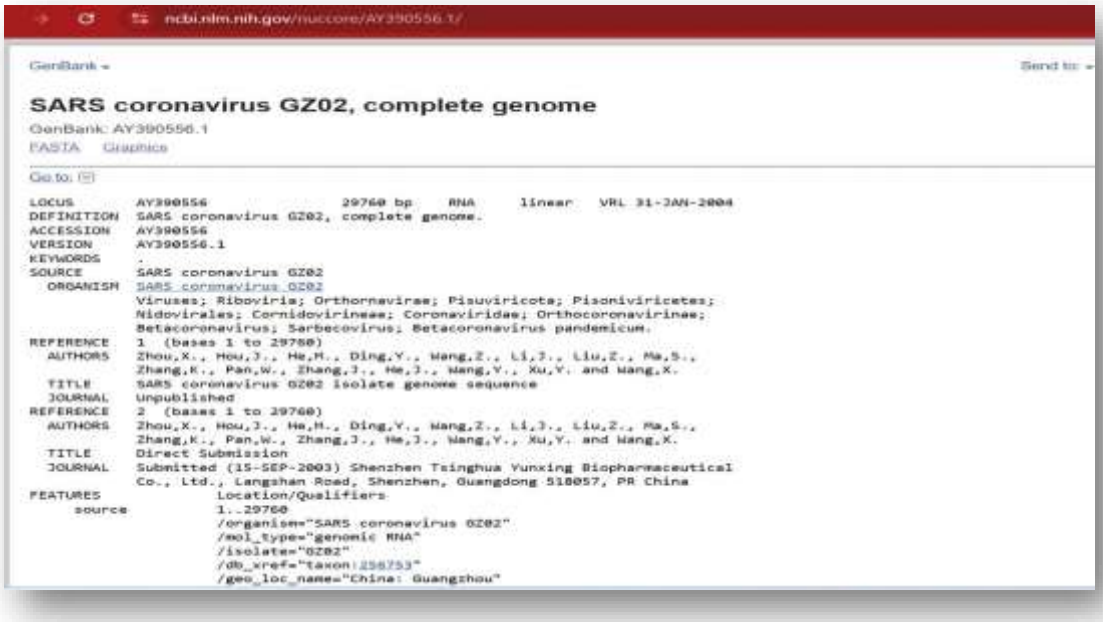



**Figure 13:** GenBank entry for SARS-CoV GZ02 complete genome (AY390556.1) with sequence annotations and metadata.

To download a copy of the DNA sequence in FASTA format, click on the "FASTA" option under the sequence section (as shown in positions 1, 2, 3, or 4 in the next figure).

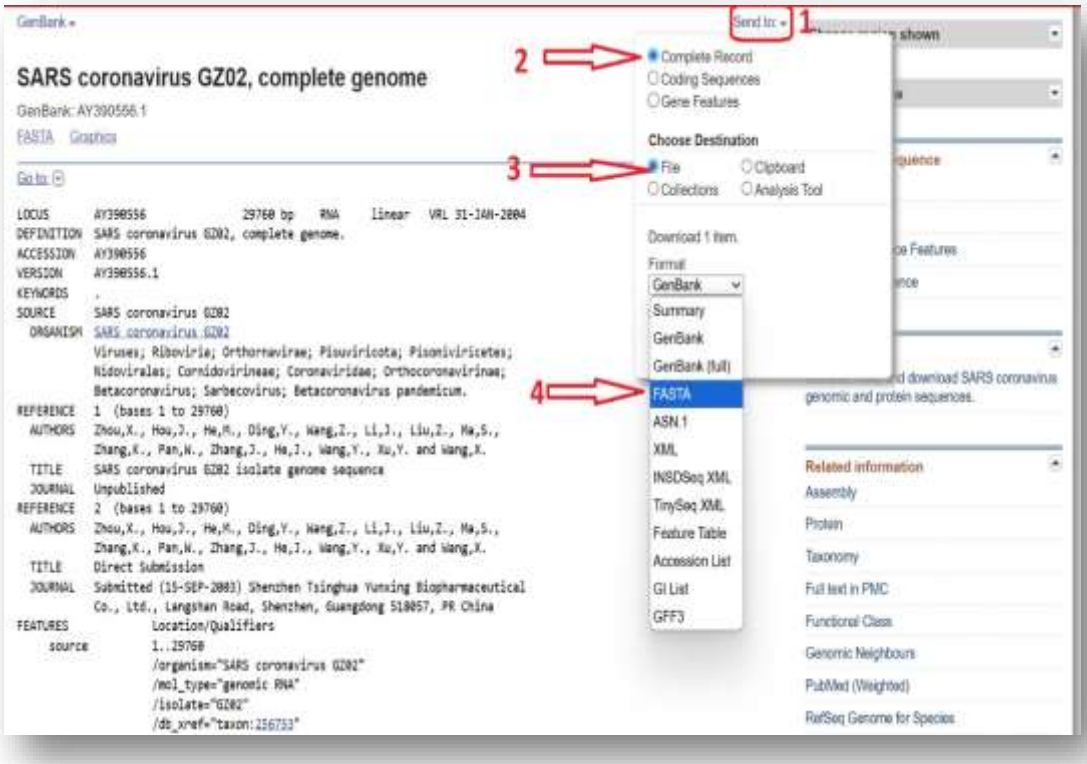



**Figure 14:** Downloading the DNA sequence in FASTA format by selecting the "FASTA" option from the sequence section (positions 1, 2, 3, or 4).

**4.Nucleotide composition**

Upon importing the curated sequences into BioEdit (version 7.2), researchers can execute fundamental bioinformatic analyses. The accompanying figure (Figure 14) provides a visual guide to deriving nucleotide composition statistics (A/T/G/C percentages) for all sequences specified in Table 5, including quality control steps and output interpretation.

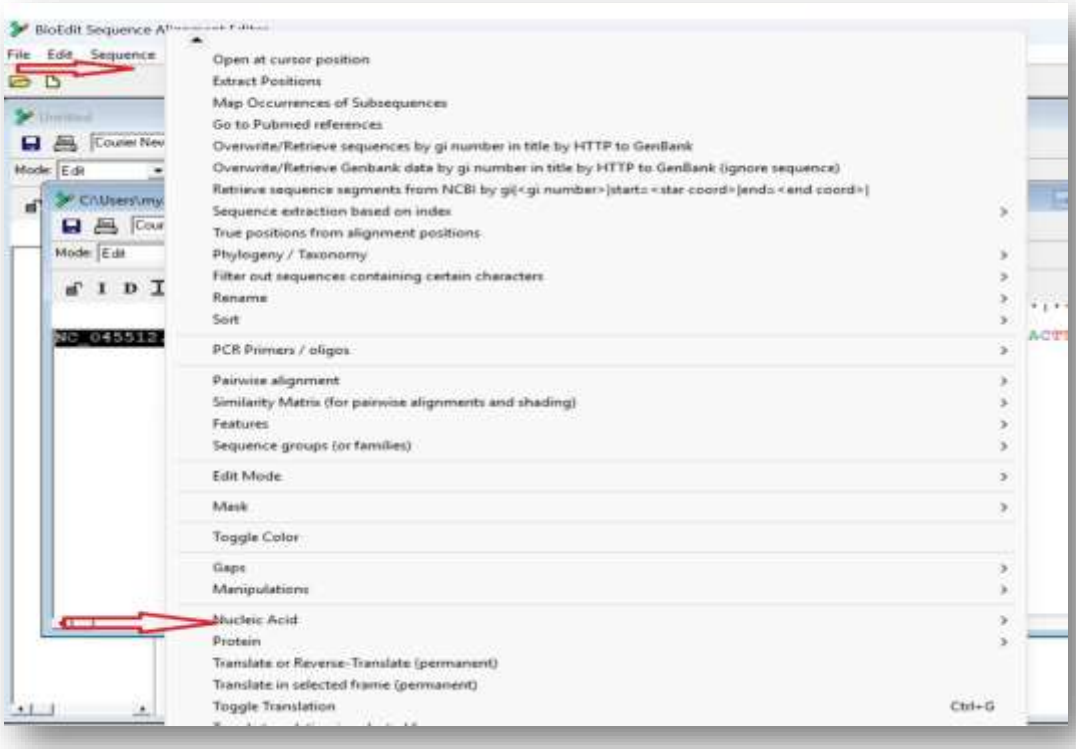
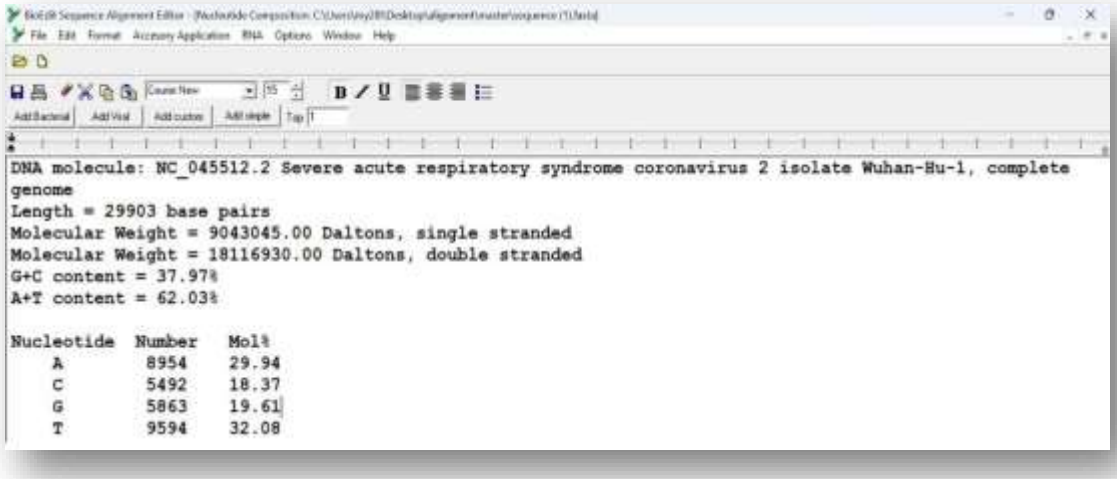



**Figure 15:** Steps to obtain nucleotide composition of selected samples using BioEdit.

As an example, the nucleotide composition of the Wuhan sample is illustrated in the following figure.



```
DNA molecule: NC_045512.2 Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete
genome
Length = 29903 base pairs
Molecular Weight = 9043045.00 Daltons, single stranded
Molecular Weight = 18116930.00 Daltons, double stranded
G+C content = 37.97%
A+T content = 62.03%

Nucleotide  Number   Mol%
    A        8954   29.94
    C        5492   18.37
    G        5863   19.61
    T        9594   32.08
```

**Figure 16:** Nucleotide composition of the Wuhan SARS-CoV-2 sample.


**Figure 17:** Nucleotide composition analysis of SARS-CoV-2 Wuhan-Hu-1 (NC_045512.2) in Blocklit Sequence Alignment Editor, showing G+C content (37.97%) and A+T content (62.03%) with detailed base distribution statistics.

The nucleotide composition of the remaining samples is summarized in the following table.

**Table 5:** Nucleotide composition (G+C and A+T content) of selected coronavirus genomes.

| Sample name | G+C content | A+T content |
|---|---|---|
| Wuhan-Hu-1 | 37.97% | 62.03% |
| Guangzhou SARS GZ02 | 40.81% | 59.19% |
| Feline | 38.14% | 61.86% |
| Bat RaTG13 | 38.03% | 61.97% |
| Pangolin | 38.19% | 61.81% |
| Porcine | 41.74% | 58.26% |
| Bat Yunnan2011 | 40.87% | 59.13% |

This table 6 quantifies the genomic base composition of various human and animal coronaviruses, presenting both G+C and A+T content percentages. Notably, the Wuhan-Hu-1 strain (SARS-CoV-2) exhibits a characteristic coronavirus profile with elevated A+T content (62.03%) and reduced G+C content (37.97%). Comparative analysis reveals the porcine

coronavirus possesses the highest G+C ratio (41.74%), a feature potentially linked to genomic stability and translational efficiency. Animal-derived strains—including Feline, Bat RaTG13, Pangolin, and Bat Yunnan2011—demonstrate conserved G+C content (38–41%), implying shared evolutionary constraints. These compositional variations, though subtle, may critically influence viral RNA secondary structure, replication dynamics, and host-specific adaptation mechanisms.

**5.Pairwise sequence analysis**

Pairwise sequence alignment is a computational method that evaluates the degree of similarity between two biological sequences (DNA, RNA, or protein) by identifying matching regions. These alignments reveal potential functional conservation, structural homology, or evolutionary connections between sequences.

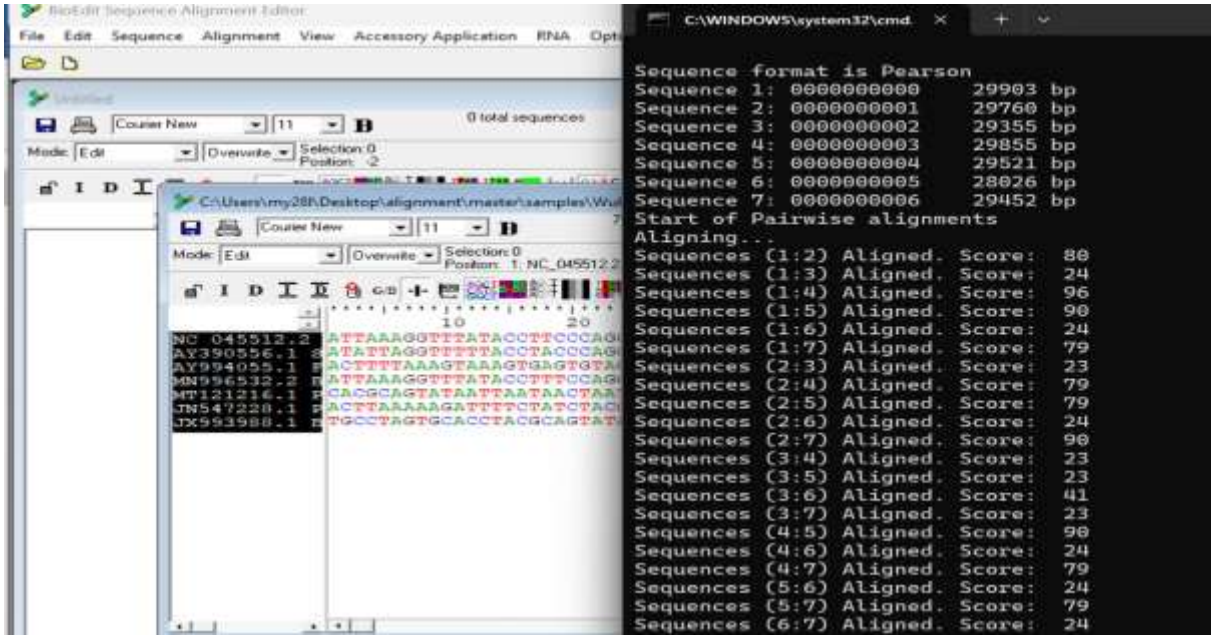



**Figure 19:** Pairwise sequence alignment illustrating similarities between two biological sequences, used to detect conserved regions, structural homology, or evolutionary relationships.

# General conclusion

This study offered a practical and methodological overview of how genomic analysis can be applied to better understand coronaviruses, particularly SARS-related strains. Using BioEdit, a legacy but still useful bioinformatics tool (Hall, 1999), we were able to carry out several essential steps in genomic data handling, including sequence importation, editing, alignment, translation, and mutation analysis.

First, we retrieved complete genome sequences of various coronaviruses—SARS-CoV, SARS-CoV-2, bat coronavirus RaTG13, pangolin CoV, feline CoV, porcine CoV, and others—from public genomic databases like NCBI GenBank and GISAID. These sequences were used to perform multiple sequence alignments with ClustalW via BioEdit, which enabled the comparison of conserved and divergent regions across the genomes. Such comparisons are crucial for identifying evolutionary changes, mutation hotspots, and potential functional domains like the Spike protein receptor-binding domain (RBD).

The translation of nucleotide sequences into amino acid sequences allowed us to better understand how specific genetic mutations can affect protein structure and function—especially in key viral proteins responsible for host entry, virulence, and immune evasion. Observing such changes between animal-origin strains (e.g., bat, pangolin, feline, and porcine CoVs) and human strains provides insight into zoonotic transmission and viral adaptation.

Moreover, we conducted an analysis of nucleotide composition, specifically G+C and A+T content, which can reflect evolutionary pressures and genome stability. For example, porcine coronavirus showed the highest G+C content, suggesting a potentially different evolutionary path compared to other strains like Wuhan-Hu-1.

Finally, while BioEdit itself does not support direct phylogenetic tree construction, it provides the foundation for such analyses by allowing the export of aligned sequences to external programs such as MEGA (Tamura et al., 2021) or IQ-TREE (Nguyen et al., 2015). Phylogenetic approaches, when combined with accurate alignment data, help researchers infer evolutionary relationships, track the origin of outbreaks, and anticipate future cross-species transmission events.

In conclusion, this work underlines the importance of bioinformatics tools in viral genomics. Even basic platforms like BioEdit can play a significant role in introductory research by enabling the analysis and interpretation of viral genome sequences. As viral evolution continues to shape public health responses, ongoing genomic surveillance and comparative analysis remain essential in anticipating

emerging threats and guiding vaccine or therapeutic development.

# Reference bibliography

[1] I. Popa, S. C. Ștefan, A. A. Olariu, Ș. C. Popa, and C. F. Popa, "Modelling the COVID-19 Pandemic Effects on Employees' Health and Performance: A PLS-SEM Mediation Approach," International Journal of Environmental Research and Public Health, vol. 19, no. 3, p. 1865, Feb. 2022, doi: 10.3390/ijerph19031865.

[2] T. T.-Y. Lam et al., "Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins," Nature, vol. 583, no. 7815, pp. 282–285, Jul. 2020, doi: 10.1038/s41586-020-2169-0.

[3] P. Zhou et al., "A pneumonia outbreak associated with a new coronavirus of probable bat origin," Nature, vol. 579, no. 7798, pp. 270–273, Mar. 2020, doi: 10.1038/s41586-020-2012-7.

[4] F. Wu et al., "A new coronavirus associated with human respiratory disease in China," Nature, vol. 579, no. 7798, pp. 265–269, Mar. 2020, doi: 10.1038/s41586-020-2008-3.

[5] S. J. Flint, V. R. Racaniello, G. F. Rall, T. Hatziioannou, A. M. Skalka, and S. J. Flint, Principles of Virology, Volume 2: Pathogenesis and Control, 5th ed. Hoboken, NJ, USA: Wiley, 2020.

[6] E. C. Holmes et al., "The origins of SARS-CoV-2: A critical review," Cell, vol. 184, no. 19, pp. 4848–4856, Sep. 2021, doi: 10.1016/j.cell.2021.08.017.

[7] D. Cucinotta and M. Vanelli, "WHO Declares COVID-19 a Pandemic," Acta Bio Medica Atenei Parmensis, vol. 91, no. 1, pp. 157–160, Mar. 2020, doi: 10.23750/abm.v91i1.9397.

[8] B. Hu, H. Guo, P. Zhou, and Z.-L. Shi, "Characteristics of SARS-CoV-2 and COVID-19," Nature Reviews Microbiology, vol. 19, no. 3, pp. 141–154, Mar. 2021, doi: 10.1038/s41579-020-00459-7.

[9] N. Zhu et al., "A Novel Coronavirus from Patients with Pneumonia in China, 2019," The New England Journal of Medicine, vol. 382, no. 8, pp. 727–733, Feb. 2020, doi: 10.1056/NEJMoa2001017.

[10] M. A. Khodja, A. Melouki, N. Benazi, and A. Dehimat, "Genomic analysis of SARS-CoV-2 using Biopython and SimPlot comparison," Journal of Bioscience and

Applied Research, vol. 0, no. 0, pp. 0–0, Dec. 2024, doi: 10.21608/jbaar.2024.332656.1098.

[11] M. A. Shereen, S. Khan, A. Kazmi, N. Bashir, and R. Siddique, "COVID-19 infection: Emergence, transmission, and characteristics of human coronaviruses," Journal of Advanced Research, vol. 24, pp. 91–98, Jul. 2020, doi: 10.1016/j.jare.2020.03.005.

[12] D. A. Jamison et al., "A comprehensive SARS-CoV-2 and COVID-19 review, Part 1: Intracellular overdrive for SARS-CoV-2 infection," European Journal of Human Genetics, vol. 30, no. 8, pp. 889–898, Aug. 2022, doi: 10.1038/s41431-022-01108-8.

[13] M. Saberiyan, E. Karimi, Z. Khademi, P. Movahhed, A. Safi, and A. Mehri-Ghahfarrokhi, "SARS-CoV-2: phenotype, genotype, and characterization of different variants," Cellular and Molecular Biology Letters, vol. 27, no. 1, p. 50, Dec. 2022, doi: 10.1186/s11658-022-00352-6.

[14] I. Jungreis, R. Sealfon, and M. Kellis, "SARS-CoV-2 gene content and COVID-19 mutation impact by comparing 44 Sarbecovirus genomes," Nature Communications, vol. 12, no. 1, p. 2642, May 2021, doi: 10.1038/s41467-021-22905-7.

[15] P. Zhou et al., "A pneumonia outbreak associated with a new coronavirus of probable bat origin," Nature, vol. 579, no. 7798, pp. 270–273, Mar. 2020, doi: 10.1038/s41586-020-2012-7.

[16] R. Lu et al., "Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding," The Lancet, vol. 395, no. 10224, pp. 565–574, Feb. 2020, doi: 10.1016/S0140-6736(20)30251-8.

[17] S. Wang et al., "A Crowned Killer's Résumé: Genome, Structure, Receptors, and Origin of SARS-CoV-2," Virologica Sinica, vol. 35, no. 6, pp. 673–684, Dec. 2020, doi: 10.1007/s12250-020-00298-z.

[18] L.-F. Wang and D. E. Anderson, "Viruses in bats and potential spillover to animals and humans," Current Opinion in Virology, vol. 34, pp. 79–89, Feb. 2019, doi: 10.1016/j.coviro.2018.12.007.

[19] M. Pensaert, P. Callebaut, and J. Vergote, “Isolation of a porcine respiratory, non-enteric coronavirus related to transmissible gastroenteritis,” The Veterinary Quarterly, vol. 8, no. 3, pp. 257–261, Jul. 1986, doi: 10.1080/01652176.1986.9694050.

[20] T. A. Hall, “BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT,” Nucleic Acids Symposium Series, vol. 41, no. 41, pp. 95–98, 1999.

[21] A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, and D. Veesler, “Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein,” Cell, vol. 181, no. 2, pp. 281–292.e6, Apr. 2020, doi: 10.1016/j.cell.2020.02.058.

[22] K. Tamura, G. Stecher, and S. Kumar, “MEGA11: Molecular Evolutionary Genetics Analysis Version 11,” Molecular Biology and Evolution, vol. 38, no. 7, pp. 3022–3027, Jun. 2021, doi: 10.1093/molbev/msab120.

[23] J. Trifinopoulos, L.-T. Nguyen, A. von Haeseler, and B. Q. Minh, “W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis,” Nucleic Acids Research, vol. 44, no. W1, pp. W232–W235, Jul. 2016, doi: 10.1093/nar/gkw256.