Molecular Detection & Genetic Characterization of Coronavirus from Bat sample in Nipah Prone Areas of Bangladesh

By
Aurchi Komolika
ID: 20136041

A thesis submitted to the Department of Mathematics and Natural Sciences (MNS) for partial fulfillment of the requirements for the degree of B.Sc. in Biotechnology.

Department of Mathematics and Natural Sciences
BRAC University
December, 2024

Declaration

It is hereby declared that:

1. The thesis submitted is our own original work while completing our degree at BRAC University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material that has been accepted or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Aurchi Komolika, 20163041

Approval

The thesis/project titled "Molecular Detection & Genetic Characterization of Coronavirus from Bat sample in Nipah Prone Areas of Bangladesh" submitted by **Aurchi Komolika (20136041)** has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Biotechnology on 5th December 2024.

Examining Committee:

Supervisor: _____

Dr. Mahboob Hossain, PhD
Professor, Mathematics and Natural Sciences
BRAC University

Co-Supervisor:  _____

Dr. Mohammed Ziaur Rahman, PhD
Senior Scientist and Head,
One Health Laboratory and Program for Respiratory Infections
Infectious Diseases Division, icddr,b, Dhaka-1212, Bangladesh
Email: mzrahman@icddrb.org

Program Director: _____

Dr. Munima Haque, PhD
Associate Professor, Mathematics and Natural Sciences
BRAC University

Departmental Head: _____

Md. Firoze H. Haque, PhD
Associate Professor and Chairperson, Mathematics and Natural Sciences
BRAC University

Abstract

Bats are considered key natural carriers of Nipah virus and various zoonotic coronaviruses. In recent years, human activities have significantly contributed to the emergence of zoonotic diseases, leading to virus outbreaks like SARS-CoV-2. This virus results from human contact that promotes interactions between wild animals, livestock, and humans. The *Coronaviridae* family includes viruses with positive-sense RNA genomes, and SARS-CoV-2, classified as a β-CoV, shares genetic similarities with bat-derived coronaviruses. Both SARS-CoV and MERS-CoV originated from bats, with intermediate hosts playing a role in their transmission to humans. Additionally, the Nipah virus, belonging to the *Paramyxoviridae* family, also utilizes bats as hosts, raising concerns about cross-species transmission and potential viral recombination. Viral coinfection can be influenced by both host and virus factors, which affect transmission dynamics and pathogenicity. Investigating the interactions and host switching among bat species that carry these viruses is crucial, as it may lead to the emergence of new viral strains or changes in viral pathogenicity. The objective of this study was to categorically detect and analyze bat coronavirus samples from Nipah-prone areas as they have a higher probability of viral transmission, coinfection and cross-species transmission leading to recombination mutation in the RdRp region. A total of 1625 bat samples collected from five nipah-prone areas in Bangladesh were observed, where 16 samples tested positive for *Coronaviridae* targeting the conserved region of RdRp (RNA Dependent RNA Polymerase) by semi-nested PCR. 10 out of the 16 samples were successfully retrieved after Sanger sequencing through a specific protocol. Among them, nine of the positive samples were obtained by the pathogen discovery protocol suggested by **(Watanabe et al., 2010)** (Protocol004-Used for bat coronavirus identification) and one (OHL-BAT-003) through **(Maganga et al., 2014)** (Protocol-003-Used for human coronavirus identification). While the Coronavirus Typing Tool (CTT) assisted in identifying reference sequences for eight samples, the two samples designated OHL-BAT-001 and OHL-BAT-004 remain unidentified following molecular analysis. Mutational analysis for these two strains was conducted to identify mutations in the RdRp region. Further correlation was established between RdRp mutations and host switch, cross-species transmission, and multiple Bat-CoV coexistence by constructing a tanglegram (co-phylogenetic analysis). For future prospects of such investigative studies, full genome sequencing should be done to confirm the

novel mutations and viral recombination due to cross-species transmission. This approach would enhance our understanding of the molecular evolution and zoonotic potential of bat coronaviruses in Nipah-prone areas

Keywords: Bat-Coronaviruses, Genetic characterization, Coronaviridae, Nipah virus, RdRp gene, Bat, Bangladesh.

Dedicated To

My friends and family who have been a constant support system throughout the varied conquests of life

Acknowledgements

I am deeply thankful to the Almighty Allah, the Most Gracious and the Most Merciful, for bestowing upon me the patience and strength necessary to embark on this remarkable journey.

I would like to extend my heartfelt appreciation to my supervisor, Prof. Dr. **Dr. Mahboob Hossain,** Professor, Mathematics and Natural Sciences, BRAC University, Dhaka. His exemplary mentorship and steadfast encouragement have proven to be invaluable throughout my journey. His constant support and profound inspiration have propelled me forward at each step, and for that, I am profoundly grateful. Additionally, I wish to express my sincere appreciation to my co-supervisor, **Dr. Mohammed Ziaur Rahman**, a Senior Scientist and Head of the One Health Laboratory and the Program for Respiratory Infections at the Infectious Diseases Division of the International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b). His genuine interest in my research and his continuous guidance throughout this thesis journey have been immensely helpful. I am thankful for the opportunity he provided me with, which has been key to my successful completion of the thesis. My sincere thanks also goes to **Dr. Munima Haque,** Program Director, Associate Professor, Department of Mathematics and Natural Sciences, BRAC University for allowing me to have this opportunity. I must also recognize **Md. Mahfuzur Rahman**, Senior Research Officer and team lead of the Pathogen Discovery and Genomics Research Group at the One Health Laboratory, icddr,b. His remarkable patience and willingness to address my countless inquiries have consistently inspired me. I am fortunate to have him as a mentor. From the very commencement of my experience at icddr,b, I have considered Research Fellow **Momtaz Zamila Bukharid** to be more than just a colleague; she has been like a friend to me. She provided both emotional and mental support throughout my journey. Her unwavering commitment to mentoring and supporting me has been a tremendous asset. She has always been available to offer guidance and recommendations, and her encouraging words have motivated me to pursue my research with renewed enthusiasm. I feel extremely fortunate to have had her companionship during this incredible chapter of my life. **Moumita Chakrabarty**, a Senior Research Assistant at the One Health Laboratory, icddr,b, also

holds a special place in my heart. She consistently communicates with kindness and encouragement. I want to express my sincere gratitude to her for her uplifting remarks, which have kept me motivated and focused during this extraordinary journey. Moreover, I am incredibly thankful to **Shahreen Rahman**, Research Officer at the One Health Laboratory, icddr,b. Her guidance challenged me to step beyond my comfort zone and inspired me to strive for excellence in achieving my thesis objectives. I am deeply grateful for her invaluable advice especially in professional capacities. Finally, I wish to express my sincere appreciation to **Samiur Rahim**, a Research Officer at the One Health Laboratory, icddr,b. He has been a tremendous source of knowledge, encouraging me to embrace challenges and assisting me whenever necessary. His vital skills and knowledge of bioinformatics have been crucial in helping me overcome numerous obstacles.

I genuinely feel fortunate to have had such remarkable mentors and supportive seniors throughout this journey. In summary, my heartfelt gratitude extends to all those who have played pivotal roles in my academic endeavors. Their unwavering support, mentorship, and encouragement have significantly shaped my research experience and personal growth. I recognize the profound impact they have had on my journey, and I feel truly blessed to have had such wonderful individuals by my side every step of the way.

Contents:

List of Figures:

List of Tables:

List of Abbreviations:

SARS- Severe acute respiratory syndrome

MERS - Middle East respiratory syndrome

CoV - coronavirus

cryo-ET - cryo-electron tomography

ER - endoplasmic reticulum

VLP - virus-like particle

NTD - N-terminal domain

CTD - C-terminal domain

ORF - open reading frame

UTR - untranslated region

RdRP - RNA-dependent RNA polymerase

ExoN - exoribonuclease domain

ACE2 - angiotensin-converting enzyme

ERGIC - ER-to-Golgi intermediate compartment

ARDS - acute respiratory distress syndrome

ADE - antibody-dependent enhancement

RSCU - relative synonymous codon usage

PCR - polymerase chain reaction

RBD - receptor-binding domain

BLAST - Basic Local Alignment Search Tool

NCBI - National Center for Biotechnology Information

Chapter – 1

Introduction

In recent years, human activity has made emerging zoonotic diseases a major driver of virus outbreaks and other health and socio-economic incidents. A recent illustration of these outbreaks is the Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is a direct result of virus transmission and is specifically human contact based, conditions that facilitate continual interaction between wild animals, livestock and humans **(Jones, Patel et al., 2008)**. The family *Coronaviridae* comprises viruses that possess positive-sense RNA genomes ranging from 22 to 36 kb, expressed through a nested array of 3′ co-terminal subgenomic mRNAs. Members of the subfamily *Orthocoronavirinae* are defined by their enveloped virions, which have a diameter of 80 to 160 nm and feature spike proteins. The orthocoronaviruses, including the severe acute respiratory syndrome coronavirus and the Middle East respiratory syndrome-related coronavirus, are highly pathogenic to humans and have been responsible for the SARS and MERS epidemics over the past twenty years. **(Woo, P.C. et al., 2023).** Of the four genera of coronaviruses (α, β, γ, δ), human coronaviruses (HCoVs) fall under α-CoV (HCoV-229E and NL63) and β-CoV (MERS-CoV, SARS-CoV, HCoV-OC43, and HCoV-HKU1). SARS-CoV-2 is classified as a β-CoV and demonstrates significant genetic similarity to two bat-derived coronaviruses, bat-SL-CoVZC45 and bat-SL-CoVZXC21. Despite this, its genome shares characteristics typical of coronaviruses. Both SARS-CoV and MERS-CoV have origins in bats, and it is likely that SARS-CoV-2 does as well. Previous findings have indicated that intermediate hosts may have facilitated the virus's emergence in humans, with civet cats serving as intermediaries for SARS-CoV and dromedary camels for MERS-CoV **(Malik Y. A. 2020).** Therefore bats serve as a very important host in the transmission of coronaviruses creating necessities for the characterization of bat coronaviruses.

Another virus, part of the *Paramyxoviridae* family known as the Nipah virus also utilizes the bats as host for susceptibility, cross-species transmission and spillover events. Due to both viruses originating from the same host, the events of cross-transmission or molecular evolution can be correlated and might have an impact on the recombination of the viral genome, especially in the case of *Coronaviridae* since their mutational rates are likely higher. The impacts of viral coinfection can primarily be influenced by two major elements which are virus factors and host factors. Host factors during coinfection alter the body’s environment, which in turn affects how viruses are transmitted and their pathogenicity. On the other hand, virus factors in coinfection modify the intracellular environment, impacting the viral life cycle both directly and indirectly

(Du, Y. et al., 2022). Therefore, investigating phenomena such as viral coinfection, and host switching in bat species which are carriers of both viruses is significant and might lead to the emergence of novel viral strains or altered viral pathogenicity.

As one of the most successful and diverse mammalian groups on the planet, bats comprise roughly 1,435 different species spread across the globe. Zoonotic viruses with a high impact on public health or agriculture were discovered on *Pteropus* spp. through 20 years including Ebola, Marburg, SARS, Hendra and Nipah virus. The Rousettus (fruit bats) or flying foxes act as a host to the coronavirus family and are often a particular interest of study **(Larsen, Gryseels et al., 2022)**. As only 543 bat species of a global diversity of 1,435 have been sampled for coronaviruses, this suggests that much higher levels of bat coronavirus diversity exist. Bats are known to be hosts for ancestral lineages of beta coronaviruses that have given rise to emerging public health threats such as SARS-CoV, MERS-CoV and SARS-CoV-2. The origins of at least five of the seven coronaviruses that can transmit from humans to humans come from bat coronaviruses. The other two human CoVs (HCoV-OC43 and HCoV-HKU1) are also believed to be zoonotic, likely spilling over from rodents and cattle into humans. Moreover, it has been suggested that bats were the primary host for such zoonotic transmission **(Ruiz-Aravena et al.,2022)**. There is a significant variety of coronaviruses present in bats, as over 60 species (exceeding 4,000 individual sequences) have been identified across 13 of the 19 recognized mammalian subgenera of Alphacoronavirus and Betacoronavirus.

Bats experimentally infected with coronaviruses have shown mild tissue damage, such as rhinitis and interstitial pneumonia, but typically display no clinical symptoms. Evidence of the virus or its RNA has been detected in their respiratory tracts and intestines. Our knowledge of bats' immune responses—both adaptive and innate—to coronaviruses is limited. While some studies monitor other pathogens in bats, there is little serological data available for coronaviruses. Their antibody responses may be weak and short-lived, and therefore, there is minimal information on the innate immune responses specific to coronaviruses. In experiments involving Egyptian fruit bats (*Rousettus aegyptiacus*) infected with bat SARS-like coronavirus WIV1, which was initially isolated from a Chinese rufous horseshoe bat (*Rhinolophus sinicus*), viral replication was limited, and no bats displayed significant signs of illness. Of the 12 bats studied, only 2 showed seroconversion, with enzyme-linked immunosorbent assays revealing no neutralizing antibodies.

Similarly, when Jamaican fruit bats (*Artibeus jamaicensis*) were exposed to a human strain of MERS-CoV, only one of ten bats developed neutralizing antibodies. Additionally, moderate lung damage was observed, and the expression of innate antiviral genes (MX1, CCL5, and ISG56) was modestly increased **(Ruiz-Aravena et al., 2022)**. Recombination among large coronavirus genomes is a frequent occurrence; it generates increased genetic diversity, enhances viral evolution, and raises the likelihood of changes in cell tropism, host range, and pathogenicity. Around two-thirds of the coronavirus genome is dedicated to coding for an RNA-dependent RNA polymerase and other non-structural proteins essential for replication. At the same time, the remaining third comprises four structural proteins—the spike, envelope, membrane, and nucleocapsid proteins—as well as various accessory proteins. Coronaviruses replicate their genomes using RNA-dependent RNA polymerase, which is known to be error-prone, leading to mutations during the replication process. Nonetheless, the three largest viral families within the order Nidovirales—Coronaviridae, Roniviridae, and Mesoniviridae—all possess a 3′–5′ exoribonuclease which enhances the fidelity of RNA replication, a function that may be crucial for maintaining adequate fitness in such large genomes **(Ruiz-Aravena et al., 2022)**. Variability in replication among different host species may result in differences in sequence variation, potentially affecting the emergence of novel variants.

For RNA viruses such as *Coronaviridae,* the RNA-dependent RNA polymerase (RdRp) gene is an important target due to its critical function in RNA synthesis and the lack of a similar gene in host organisms. Consequently, RdRp is a primary focus for antiviral agents like Remdesivir **(Gordon et al., 2020)**, which was explored as a possible treatment for COVID-19. Considering the rapid mutations that coronaviruses undergo, we aimed to analyze the RdRp gene sequences of bat coronaviruses from various Nipah virus prone geographical locations to identify emergence of any novel sequence and establish a correlation about RdRp mutations. In this study, we discovered and characterized mutations in the RdRp gene sequence, detected cross species transmission which revealed susceptible bat species with mutational strains isolated from distinct regions of Bangladesh. Our findings strongly indicate that mutations are prevalent in the bat coronavirus genome, highlighting the need for new strategies to target this virus in bat. **(Chand, G. et al., 2020)**

Chapter – 2

Literature Review

1. Coronavirus

Coronaviruses, a diverse group of enveloped, single-stranded RNA viruses that are part of the Coronaviridae family, have gained global attention recently due to their capacity to cause severe respiratory illnesses in humans. The appearance of new coronaviruses presents significant challenges to public health systems worldwide, underscoring the critical need for a more profound comprehension of their genetic variability and origins. This review of literature seeks to offer an overview of coronaviruses, delve into the emergence of new strains, examine the role of bats as reservoirs for these viruses, and stress the significance of understanding genetic diversity to enhance pandemic readiness.

1.1 The family Coronaviridae

Coronaviruses have been acknowledged for many years as pathogens that impact a broad range of animals, such as birds and mammals. Coronaviruses (CoVs) belong to the order Nidovirales, the family Coronaviridae, and the subfamily Orthocoronavirinae. They possess the largest RNA virus genomes, which range in size from 26 to 32 kilobases (kb) **(Malik Y. A. 2020)**. The Coronaviridae family is composed of four genera: Alphacoronavirus, Betacoronavirus, Gammacoronavirus, and Deltacoronavirus. Alpha-coronaviruses and Betacoronaviruses mainly infect mammals, whereas Gamma-coronaviruses and Deltacoronaviruses mainly infect birds.

Human coronaviruses were first detected in the mid-1960s and were initially thought to cause mild respiratory infections resembling the common cold. However, the seriousness of coronavirus infections became evident with the emergence of two highly pathogenic strains: severe acute respiratory syndrome coronavirus (SARS-CoV) during 2002-2003 and Middle East respiratory syndrome coronavirus (MERS-CoV) in 2012. Both SARS-CoV and MERS-CoV are zoonotic viruses, indicating they originated in animals before spreading to humans **(Su,S., et al., 2016)**. These outbreaks emphasize the potential of coronaviruses to induce severe respiratory illnesses and emphasize the importance of vigilance in monitoring and managing these pathogens. Bat coronaviruses typically fall under the genera of alpha and beta coronavirus.

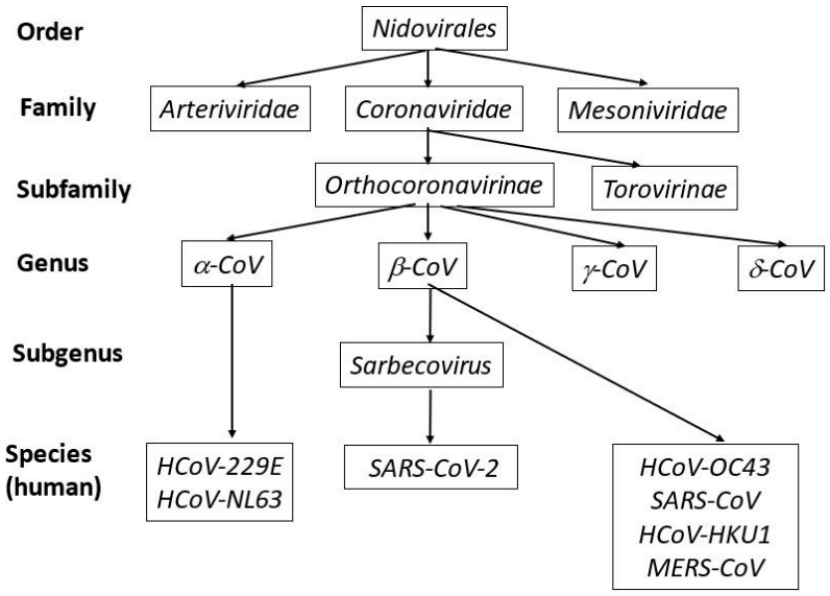


Fig. 2.1.1: Classification of coronavirus (Copyright Malaysian J Pathol, 2020)

1.2 Emergence of Novel Coronaviruses

The emergence of novel coronaviruses is a dynamic and continuous process influenced by a variety of factors such as genetic mutations, modifications in ecosystems, and human activities. These viruses possess the capacity to breach species barriers and adapt to different hosts, which can lead to sporadic outbreaks and potential pandemics. It is essential to comprehend the underlying mechanisms driving the appearance of novel coronaviruses to foresee and manage future outbreaks.The emergence of new coronaviruses typically involves spillover events, where the virus crosses from its original animal host to infect humans. These events can occur through direct contact with infected animals, consumption of contaminated animal products, or exposure to virus-laden environmental surfaces. Once introduced to humans, new coronaviruses can rapidly spread through person-to-person transmission, causing local outbreaks and, occasionally, global pandemics.

1.3 The Role of Bats as Reservoirs for Coronaviruses

Bats are recognized as a principal reservoir for a broad range of coronaviruses, playing a pivotal role in the ecology and evolutionary dynamics of these viruses. Belonging to the order Chiroptera, bats are the sole mammals capable of sustained flight. Their distinctive biology,

characterized by the capacity to harbor significant viral loads without displaying clinical symptoms, renders them well-suited hosts for a diverse spectrum of viruses, including coronaviruses.

Bats exhibit a remarkable diversity of coronaviruses, with hundreds of distinct strains identified across different bat species and regions. The current diversity of coronaviruses, including those found in bats, has arisen from processes such as mutation, recombination, and the ability of hosts to become infected or co-infected. Certain bat families appear to have evolved alongside specific subgroups of betacoronaviruses over millions of years; for instance, rhinolophid bats with SARS-related sarbecoviruses, vespertilionid bats with MERS-related merbecoviruses, and pteropodidae bats with nobecoviruses. Host switching seems to be most prevalent within the rhinopholid–Sarbecovirus lineage, allowing these viruses to jump successfully between host species. This variability in bat coronaviruses might provide them with opportunities to adapt to new hosts, develop different tissue tropisms, and exhibit differences in how transmissible or severe infections are in those hosts. Once a virus establishes itself in a new host population, evolutionary pressures are likely to favor lineages that become more fit in that environment, leading to increased transmissibility, as seen with SARS-CoV-2 in humans **(Ruiz-Aravena et al., 2022)**.

1.4 Understanding Genetic Diversity for Pandemic Preparedness

The ongoing danger from emerging coronaviruses, including the current pandemic due to SARS-CoV-2, highlights the importance of studying the genetic diversity of these viruses. Genetic diversity -the differences among populations or among individuals within a population, including genetic differences through the inheritance differences of the nucleic acid sequence between offspring– is a broad concept that correlates with the range of populations between and within a species. For coronaviruses, genetic diversity covers the range of genetic variation between viral strains found circulation in the bat host population and other reservoirs. Genetic diversity is crucial for multiple reasons. For instance, it provides critical information on the evolutionary history and source of the novel coronaviruses, occurring as a result of the virus transmitting from non-reservoir hosts to humans. Thus, tracking the lineage of a virus allows for

knowing reservoirs and pathways of transmission and easily identifies the ecological and environmental cause of viral evolution.

Second, genetic diversity impacts the antigenic features of coronaviruses that make them evade host defence and result in differences in disease spectra and transmission dynamics. Point mutations in the viral genome may lead to the generation of new ones, straining the existing anti-viral immunity and complicating vaccination and drugs to treat new flaviviruses. Third, genetic diversity is important for designing effective diagnostic and monitoring tools for emerging coronaviruses because it enables researchers to determine specific genetic markers linked with various viral strains. They may then produce unique molecular evaluations to detect coronaviruses and distinguish between them. This may lead to early detection and elimination.

In brief, the appearance of novel coronaviruses poses a severe threat to the global public health community. It is vital to gain a thorough understanding of the genetic diversity and origin of novel coronaviruses. The ecologic and behaviour-related discoveries of how bats serve as the key coronavirus reservoir are critical for reducing the risk associated with future spillover. Understanding the genetic diversity of coronaviruses would considerably improve overall preparedness for pandemics and formulate more efficient policies to block and prevent new infectious diseases from emerging.

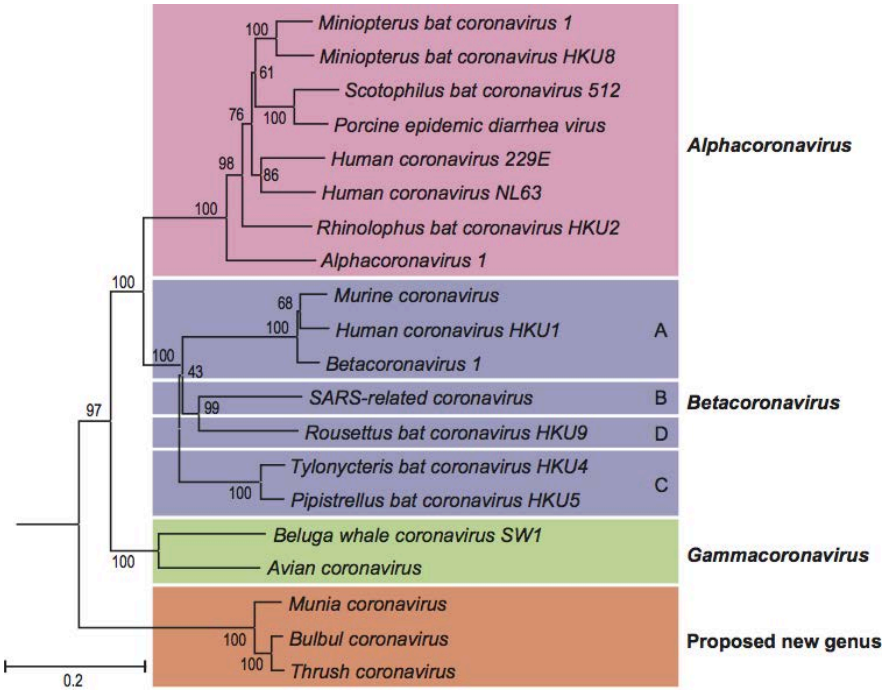


Fig. 2.1.2: Phylogenetic relationship in genera of *Coronaviridae* family (Copyright ICTV, 2011)

2. Virion

2.1 Morphological characteristics of Coronavirus

Coronaviruses (CoVs) are notable for their dissimilar morphological features, which are important aspects of viral infection, replication, and pathogenesis. CoVs are enveloped viruses with a spherical or pleomorphic form with a diameter of 60 to 140 nanometers. CoV virions have club-shaped glycoprotein spikes on their exterior that give them a corona-like appearance when viewed by electron microscopy **(Dhama et al., 2020)**. The coronaviral genome encodes four major structural proteins, specifically the spike (S), membrane (M), envelope (E) and nucleocapsid (N) protein, all of which are encoded within the 3′ end of the genome. The viral entry occurs when the S protein interacts with the surface receptors on the host cell, initiates fusion and subsequently releases its contents into the cells. The M protein is the most prevalent protein and contributes to the structure of the viral envelope. E is also one of the smallest major structural proteins, as it is involved in viral assembly and budding. N protein is the only viral

structural protein that binds to the RNA genome which is also involved in viral assembly and budding **(Malik Y. A. 2020)**.

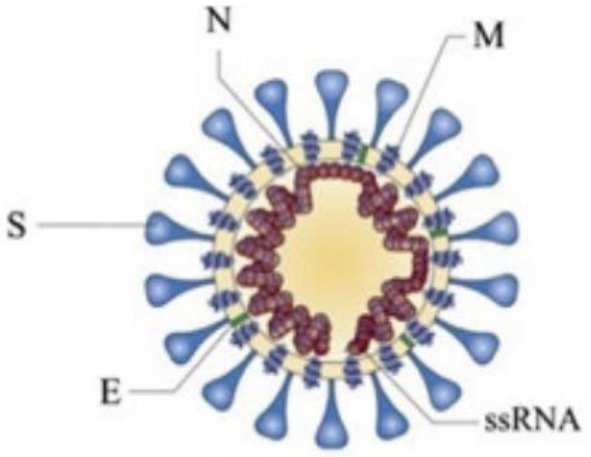


Fig. 2.2.1: Virion of coronaviridae (Viral zone), here, S- spike protein, M- membrane protein, E- envelope protein, N- nucleocapsid protein (Copyright Li et al., 2020)

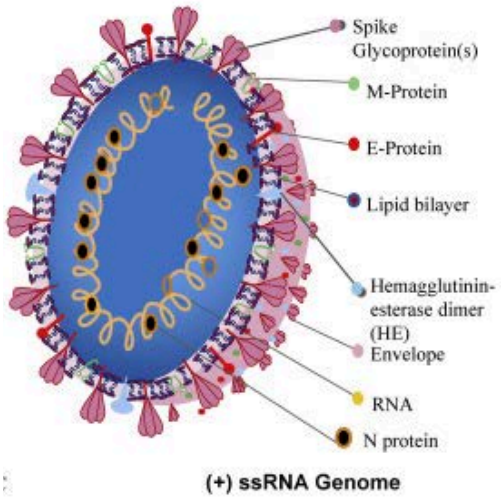


Fig. 2.2.2: Structural components of coronavirus (Copyright Journal of Molecular Biology, 2022)

2.2 Spike Glycoproteins

Infectivity and pathogenesis of coronaviruses mainly depend on their spike glycoproteins known as spike proteins or S proteins. Recently, it was discovered that the trimer structure of the SARS-CoV-2 spike glycoproteins using advanced imaging, cryo-EM, provides fundamental

information regarding the fusion machinery necessary for virus entry in host cells **(Walls et al.,2020)**. Further, the dynamic conformational shifts of S proteins during viral fusion using cryo-electron tomography (cryo-ET) have also been seen, which could help identify therapeutic targets **(Klein et al., 2021)**. The S protein enters the endoplasmic reticulum (ER) using an N-terminal signal sequence and displays extensive N-linked glycosylation. The spike-like structure is composed of homotrimers of the virus-encoded S protein. The S glycoprotein, a trimeric class I fusion protein, mediates attachment to the host receptor. For most coronaviruses, S is cleaved by a furin-like protease in the host cell into two monomeric polypeptides called S1 and S2. The S protein consists of S1 and S2, with the former comprising most of the large receptor-binding domain of the spike and the latter making up the stalk **(Malik Y. A. 2020)**.

2.3 Envelope and Membrane Proteins

Coronaviruses contain envelope proteins which are structural proteins which play an essential role in the assembly and stability of the viral envelope. High-resolution cryo-EM analyses by Schoeman and Fielding allowed the imaging of the organization of both E and M proteins within the viral envelope.

The M protein, which weighs approximately 25–30 kDa and features three transmembrane domains, is the most prevalent structural protein and determines the configuration of the viral envelope. It consists of a small glycosylated ectodomain at the N-terminus and a significantly larger C-terminal endodomain that penetrates 6–8 nm into the viral particle. The M protein functions as a dimer and displays two distinct conformations, allowing it to bind with the nucleocapsid and create membrane curvature **(Malik Y. A. 2020)**. The M protein and the N protein linkage stabilizes the nucleocapsid (the N protein-RNA complex) and the internal core of the virus, ultimately aiding in the completion of viral assembly.

The E protein, weighing approximately 8–12 kDa, is the smallest among the structural proteins. It is a transmembrane protein featuring a N-terminal ectodomain and a C-terminal endodomain that possesses ion channel functionality. Throughout the viral replication cycle, E is produced in large amounts within the infected cell; however, only a small fraction becomes part of the virus envelope. The bulk of this protein is involved in the processes of viral assembly and budding

(Malik Y. A. 2020). Together, the M and E proteins form the viral envelope, and their interaction is crucial for the generation and release of virus-like particles (VLPs).Based on the findings, we can observe the role of these proteins in the virion's morphology and interaction with the host and viral pathogenesis and vaccination.

2.4 Nucleocapsid Protein

The nucleocapsid, carrying the ribonucleoprotein, is located inside the viral envelope. The nucleocapsid is a helical structure made up of genomic RNA and nucleoproteins. Cryo-EM reconstructions established the helical package of the coronavirus pool, which underlined the importance of the role of RNA packaging and genome regulation **(Gui et al., 2017)**. The N protein is unique in its ability to bind to the RNA genome. The nucleocapsid protein has two domains which are the N-terminal domain (NTD) and the C-terminal domain (CTD) **(Malik Y. A. 2020)**. It has been seen that effective RNA binding depends on the contributions from both domains. Additionally, the N protein plays a role in the assembly and budding of the virus, leading to the formation of complete virions. Despite its critical role in structural continuity, the N protein appears to play such diverse critical roles in viral RNA synthesis, genomic growth and host disease control.

2.5 Non-Structural Protein

The ORF1ab gene codes for 16 different nonstructural proteins which play an important role in replication and transcription of the virus. Nsp12, nsp7 and nsp8 comprise together to form the RdRp complex. The RdRp enzyme is responsible for replication of the viral genome and thus is responsible for mutation, recombination and novel sequences of the coronavirus. The non structural proteins with their functions are illustrated in Table 2.5.1..

Table 2.5.1: Non Structural proteins of Coronavirus (Copyright Malaysian J Pathol, 2020)

Protein	Functions
nsp1	Promotes cellular mRNA degradation and blocks host cell translation, resulting in innate immune response block
nsp2	Function unknown, binds to prohibitin proteins
nsp3	Large, multi-domain transmembrane protein, its activities- ● Ubl1 and Ac domains, interact with N protein ● ADRP activity, promotes cytokine expression ● PLPro/Deubiquitinase domain, cleaves viral polyprotein & blocks host innate immune response ● Ubl2, NAB, G2M, SUD, Y domains, unknown functions
nsp4	Potential transmembrane scaffold protein, important for proper structure of DMVs
nsp5	Mpro, cleaves viral polyprotein
nsp6	Potential transmembrane scaffold protein
nsp7	Forms hexadecameric complex with nsp8, may act as processivity clamp for RNA polymerase
nsp8	Forms hexadecameric complex with nsp7, may act as processivity clamp for RNA polymerase, also as a primase
nsp9	RNA binding protein
nsp10	Cofactor for nsp16 and nsp14, forms heterodimer with both and stimulates ExoN and 2-O-MT activity
nsp11	Function not known yet
nsp12	RdRp
nsp13	RNA helicase, 5’ triphosphatase
nsp14	N7 MTase and 3′-5′ exoribonuclease, ExoN; N7 MTase adds 5′ cap to viral RNAs, ExoN activity is necessary for proofreading of viral genome
nsp15	Viral endoribonuclease
nsp16	2′-O-MT; shields viral RNA from MDA5 recognition

2.6 Genomic Organization

Apart from morphology, it is crucial to explore the genomic organization and replication strategies of coronaviruses to identify the pathogenesis of this virus. Indeed, an extensive description of the genome organization reveals the open reading frames (ORF) that encode structural and non-structural virion proteins essential for transcription and replication **(Masters, P. S., 2006)**.

The coronavirus genome is the largest positive RNA strand about approximately 29.9 kb long. The coronavirus genome is structured as follows: it starts with a 5′ leader sequence, followed by the untranslated region (UTR), the replicase sequence, and then the structural protein coding genes: S (Spike), E (Envelope), M (Membrane), and N (Nucleocapsid). At the 3′ end, there is a UTR and a poly(A) tail, with additional accessory genes located within the structural gene regions. Generally, the four main structural proteins are essential for creating a complete viral particle in most coronaviruses **(Malik Y. A. 2020)**.

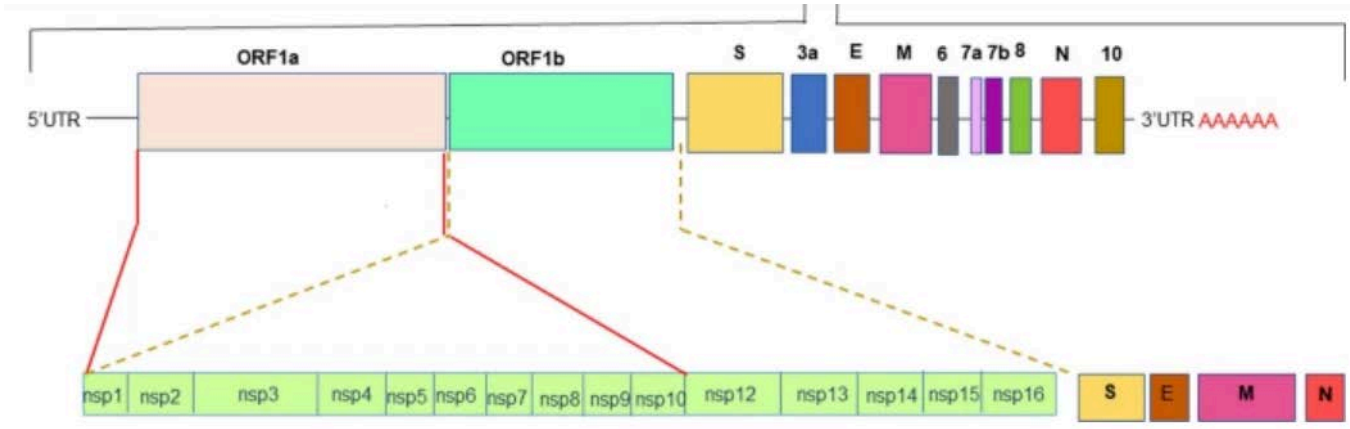


Fig. 2.2.3: Genomic structure of *Coronaviridae* family (Copyright Springer Nature)

The complete RNA genome consists of about 29,903 nucleotides and features a replicase complex located at the 5′ UTR, which includes ORF1a and ORF1b. The ORF1a region is responsible for encoding proteins nsp1 through nsp10, while ORF1b encodes proteins nsp1 through nsp16.

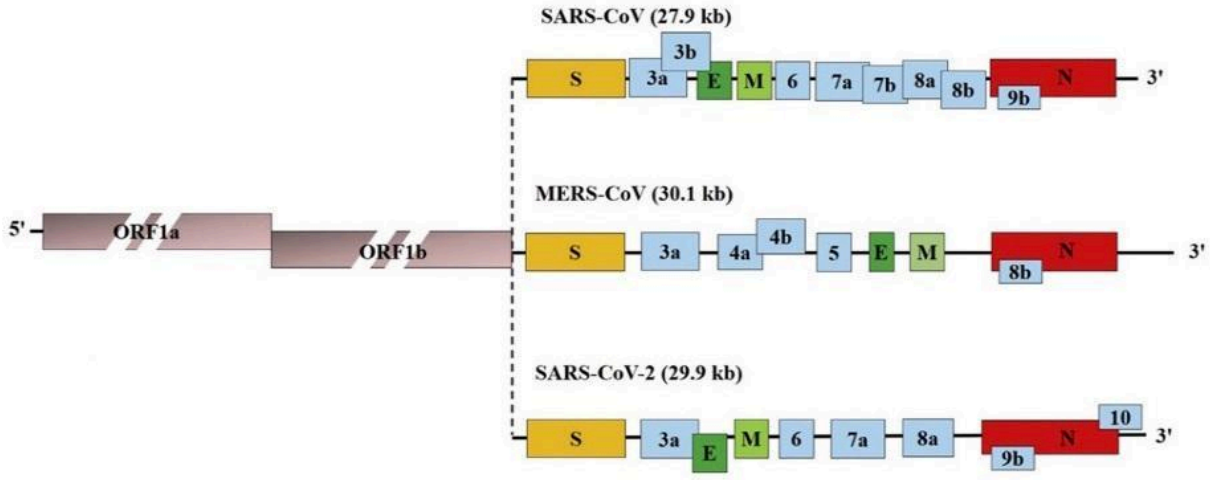


Fig. 2.2.4: Genome of SARS-CoV, MERS-CoV and SARS-CoV-2 with ORF1b gene, S/E/M/N genes (Copyright Li et al., 2020)

The alpha and beta coronaviruses all have similar genomic organization. The ORF1 gene codes for the RdRp enzyme. All three coronaviruses, SARS-CoV, MERS-CoV and SARS-CoV-2 have been found within bat species.

2.7 Transcription and Replication Mechanism

A complex interaction between viral replication complexes and cellular machinery is seen in coronavirus including mechanisms that take over cellular processes. In the case of coronavirus, the positive-strand RNA genome is a template for both replication and the synthesis of proteins. Upon entry via membrane fusion, the virus releases its positive-sense RNA into the cytoplasm. Coronaviruses (CoVs) regulate the expression of their proteins through a conserved mechanism called -1 programmed ribosomal frameshifting (PRF). The PRF in SARS-CoV-2 is very similar to that in SARS-CoV-1, differing by only one nucleotide, which does not affect the efficiency of PRF in SARS-CoV-2 **(Rastogi, M. et al., 2020)**. The viral genome consists of two open reading frames (ORFs), ORF1a and ORF1b, which are translated into non-structural proteins (nsps) within the cytoplasm. ORF1a results in a polypeptide known as pp1a, which is cleaved proteolytically to yield 10 nsps. Meanwhile, the PRF in SARS-CoV-2 enables ongoing translation into ORF1b, producing a larger polypeptide (pp1ab), which is then cleaved into 16

nsps. The cleavage of these polypeptides is performed by the viral proteases 3CLpro and Mpro **(Rastogi, M. et al., 2020)**.

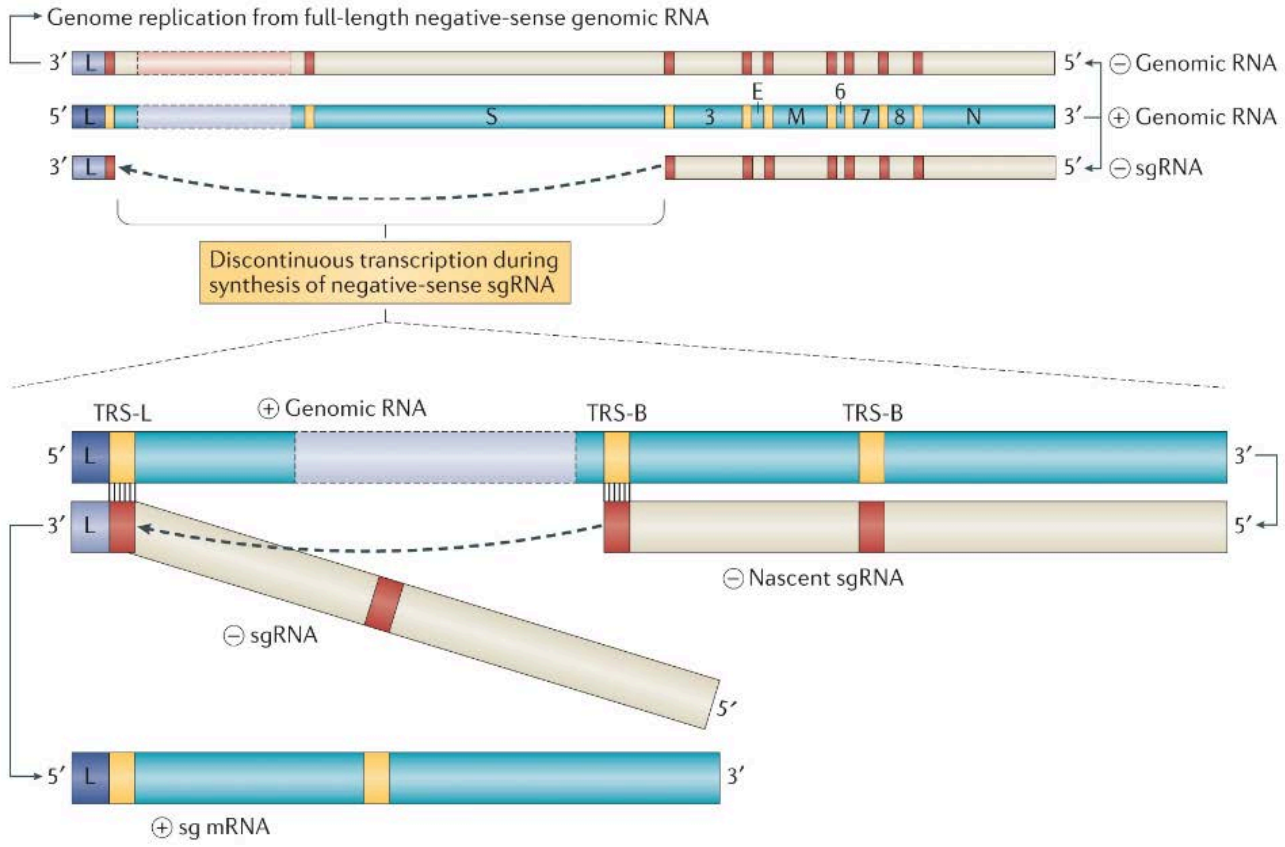


Fig. 2.2.5: Coronavirus replication and synthesis of genomic positive strand RNA (Copyright Nature Reviews Microbiology, 2020)

The viral genome replication and transcription are facilitated by RNA-dependent RNA polymerase (RdRP/nsp12). This enzyme, with cofactors nsp7 and nsp8, drives the synthesis of viral RNA. RNA viruses do not possess proofreading capabilities. However, it has been observed that an exoribonuclease domain (ExoN) present in nsp14 provides a proofreading function that helps safeguard SARS-CoV1 against mutations. The absence of ExoN results in decreased replicative accuracy. The replication complex produces full-length negative-sense RNA intermediates from the viral genome, which serve as templates for creating positive-sense genomic RNAs (gRNA) as well as sub-genomic RNAs (sgRNA). The nucleocapsid protein encapsulates the gRNA along with the S, M, and E proteins in the endoplasmic reticulum-Golgi intermediate compartment. Mature virion assembly occurs within the Golgi, and the vesicles

containing the virions combine with the plasma membrane so that the virus is released via exocytosis. SARS-CoV-2 generates nine sgRNAs (S, 3a, E, M, 6, 7a, 7b, 8, and N) that correspond to structural and accessory proteins. These sgRNAs are generated through a canonical mechanism of discontinuous transcription mediated by the Transcription Regulatory Sequence (TRS) **(Rastogi, M. et al.,2020)**.

3. Life cycle of coronavirus

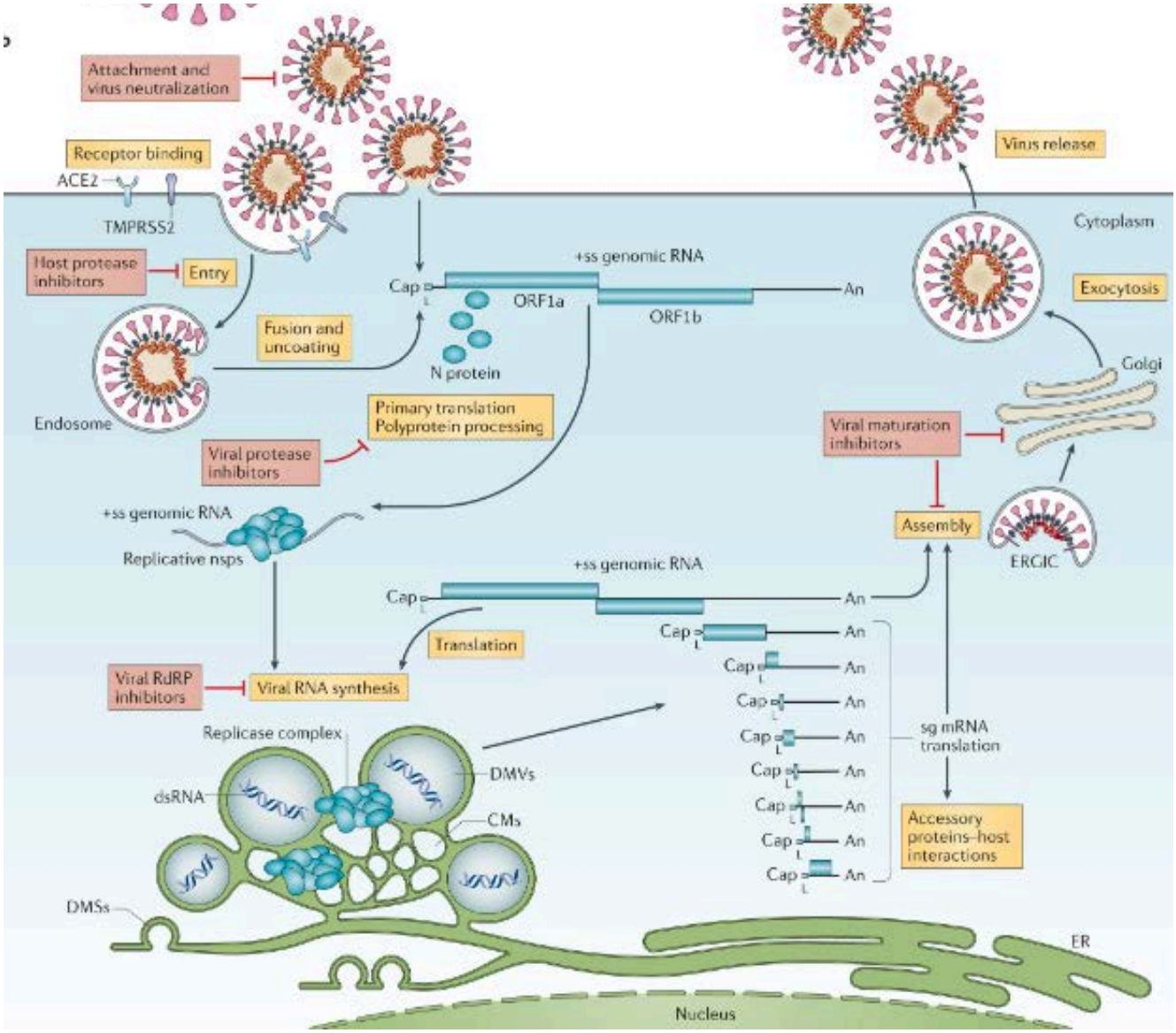


Fig. 2.3.1: Life cycle of Coronavirus (Copyright Nature Reviews Microbiology, 2020)

3.1 Viral Entry: Gateway to Infection

The initiation of coronavirus starts from when it is able to enter host cells, which is facilitated by interactions between the viral spike glycoproteins S and host cell receptors. In their work, **(Hoffmann et al., 2020)** showed that angiotensin-converting enzyme (ACE2) is the primary receptor for SARS-CoV-2 entry and internalization. In the cryo-EM studies, high-resolution cryo-EM structures of SARS-CoV-2 spike protein with human ACE2 were presented **(Wrapp et al., 2020)**. These electrophysiological revelations provide insight into how SARS-CoV-2 interacts with human ACE2 receptors, COVID-19 caused by the viral association, and some therapeutic opportunities. Attachment of the coronavirus to the host cell happens through interactions between its S protein and the receptor. The location of the receptor binding domains (RBD) within the S1 region of the S protein varies among different coronaviruses.

3.2 Endocytosis and Fusion

Receptor binding causes coronaviruses to be endocytosed, which occurs when viral particles are ingested by host cells through vesicular transport. After the S2 region of the S protein causes viral and cellular membrane fusion, the viral genome gets released into the host cell cytoplasm. An experiment conducted by **(Tortorici et al., 2019)** revealed how membrane fusion affects the structural conformation of the coronavirus S proteins and serves as a framework for understanding viral entry and fusion inhibitors.

3.3 Genome Replication and Transcription: Viral Hijacking of Host Machinery

Following entry into the host cell, coronaviruses use cellular enzymes to synthesize their genomic RNA and produce viral proteins. The genomic RNA of the virus accomplishes the transcription of subgenomic RNAs that are responsible for coding structural and non-structural proteins. **(Fehr & Perlman, 2015)** In a study described the viral replicase-transcriptase complex and showed that they interact with host cell co-factors. The genomic RNA undergoes immediate translation into two large open reading frames, ORF1a and ORF1b. It produces polyproteins pp1a and pp1ab, which are then processed both during and after translation into distinct non-structural proteins (nsps) that are part of the viral replication and transcription complex

(V'kovski, P. et al., 2021). The viral positive sense RNA is then replicated as illustrated before in the figure. All positive-sense sub-genomic RNAs are 3′ co-terminal with the full-length viral genome, resulting in a series of nested RNAs, a unique characteristic of the order Nidovirales. Both genomic and sub-genomic RNAs are generated through negative-strand intermediates. Furthermore, coronaviruses are recognized for their ability to recombine, utilizing both homologous and non-homologous recombination. This recombination capability is linked to the strand-switching function of RNA-dependent RNA polymerase (RdRp) **(Malik Y. A. 2020)**.

3.4 Assembly and Budding: From Viral Components to Infectious Progeny

The assembly of coronaviruses takes place in unique cellular compartments in which the viral structural proteins and genomic RNA are arranged to construct progeny virions. The investigation by **(Neuman et al., 2015)** revealed how coronavirus replication organelles were ultrastructurally assembled and arranged as assembly sites for morphologically characterized virions. The interaction of viral envelope proteins with host cell membranes is responsible for this budding process, which releases fully developed virions capable of infecting neighboring cells. Structural proteins that have been translated move into the endoplasmic reticulum (ER) and pass through the ER-to-Golgi intermediate compartment (ERGIC). In this area, they interact with newly synthesized genomic RNA that is encapsidated with N protein, leading to the formation of buds into the interior of secretory vesicular compartments. Ultimately, the virions are released from the infected cell via exocytosis **(V'kovski, P. et al., 2021)**.

3.5 Host Immune Response & Evasion Strategies

The coronavirus typically spreads to various organs after initially infecting the upper and lower respiratory tracts. In the case of SARS-CoV-2, the ACE2 receptors are highly expressed in type II alveolar epithelial cells, the airway epithelium, lung parenchyma, esophageal epithelial cells, enterocytes from the ileum and colon, myocardial cells, cholangiocytes in the liver, and the proximal tubule cells in the kidneys. The RNA of SARS-CoV-2 is recognized by endosomal TLR7 and TLR8, as well as RNA sensors such as MDA5 and RIG-1. Upon activation, this triggers a signaling cascade involving NF-kB transcription and AP-1, which induces the expression of genes responsible for producing pro-inflammatory cytokines like TNF-α, TGF-β,

IL-1β, IL-6, IL-8, IL-12, and IL-18, as well as chemokines such as CCL2, CCL3, CCL5, CXCL8, CXCL9, and CXCL10. This results in a cytokine storm that can directly, indirectly, or synergistically damage organs, which might lead to acute respiratory distress syndrome (ARDS) or multiple organ failure in affected patients **(Rastogi, M. et al., 2020)**. It has been reported that nsp9 and nsp10 of SARS-CoV-2 interact with NKRF to mediate IL-6/IL-8 signaling, resulting in uncontrolled activation and infiltration of neutrophils from peripheral circulation **(Li et al., 2020)**. Furthermore, antibody-dependent enhancement (ADE) directed against similar epitopes associated with S or N glycoproteins of SARS-CoV-1 may likewise lead to ARDS or multiple organ failure.

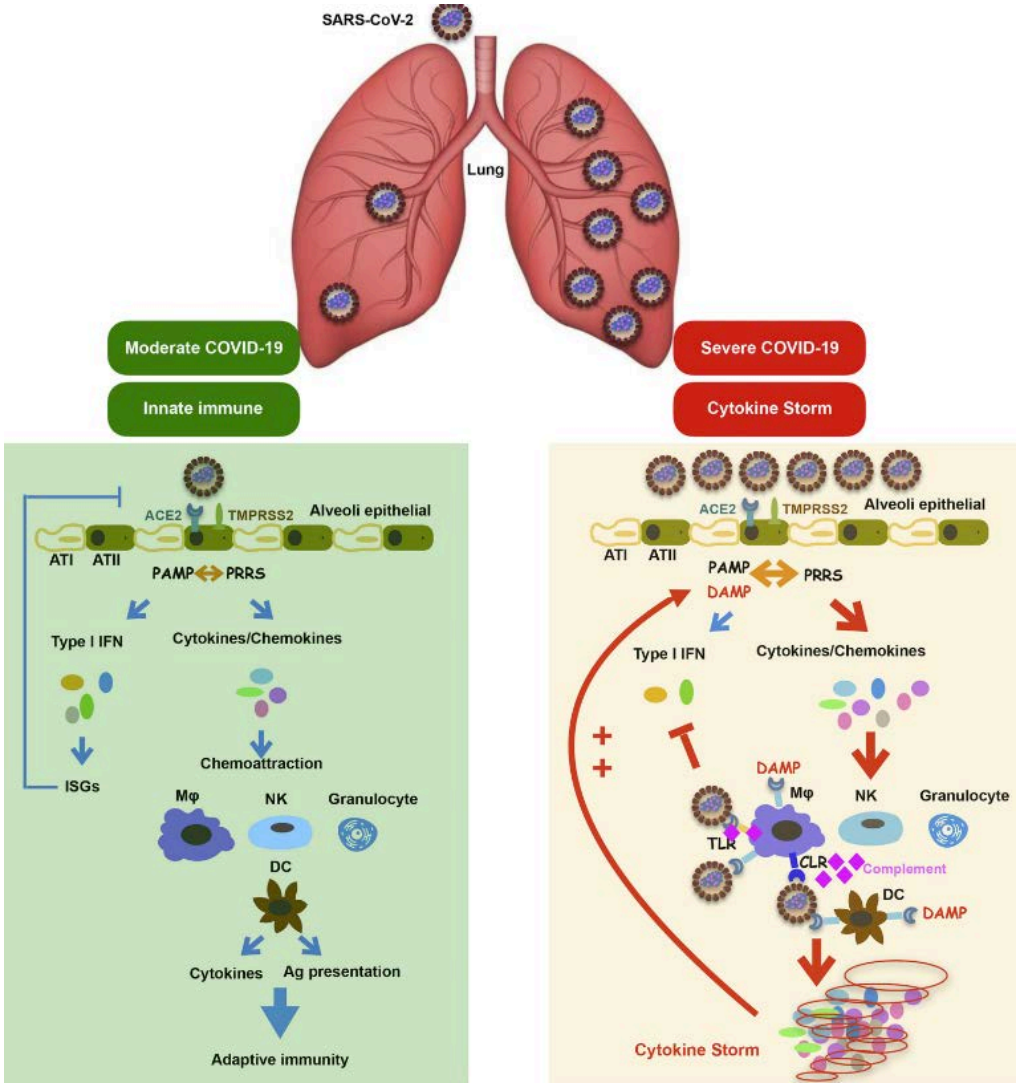


Fig. 2.3.2: Host immune response of SARS-CoV-2 in humans (Copyright Molecular Therapy)

Sophisticated strategies have been evolved by coronaviruses that immune-surveillance the hosts to escape, which allows for the persistence and dispersion of mutant strains. According to **(Zheng et al., 2020)**, the role of viral proteins in the counteraction of host immune responses cooperating with the inhibition of interferon signaling pathways and alteration of innate immune senses has been clarified. The virus evades the host's immune system by inhibiting the type I interferon response, as demonstrated in mouse models of SARS-CoV-1 and MERS. SARS-CoV-2 shows greater sensitivity to type I IFN treatment compared to SARS and MERS coronaviruses. Furthermore, they noted that mutations in ORF3b and the shortening of ORF6 have increased the susceptibility of SARS-CoV-2 to type I IFN therapy **(Lokugamage et al., 2020).** Nevertheless, several immune-evasion mechanisms have to be recognized to develop immune-therapy treatments and coronaviral vaccines.

3.6 Pathogenesis

Interaction between viral replication kinetics and host immune advances has been reported in coronavirus infection pathogenesis. According to research by **(Zhou et al., 2020)**, COVID-19 revealed immunopathological traits that included dysregulated cytokine responses and immune-mediated tissue damage. Coronavirus-affected patients often exhibit a decrease in the numbers of CD4+ and CD8+ T cells in peripheral blood leading to secondary bacterial infections and exacerbating disease severity. MERS-CoV can directly infect T lymphocytes, promoting their apoptosis and causing lymphopenia. SARS-CoV-2 can also directly infect T lymphocytes via its S glycoprotein through membrane fusion. Additionally, a deficiency in T cells may impair the activation of antibody-producing B lymphocytes, negatively affecting the production of immunoglobulins in patients **(Wang et al., 2020)**.

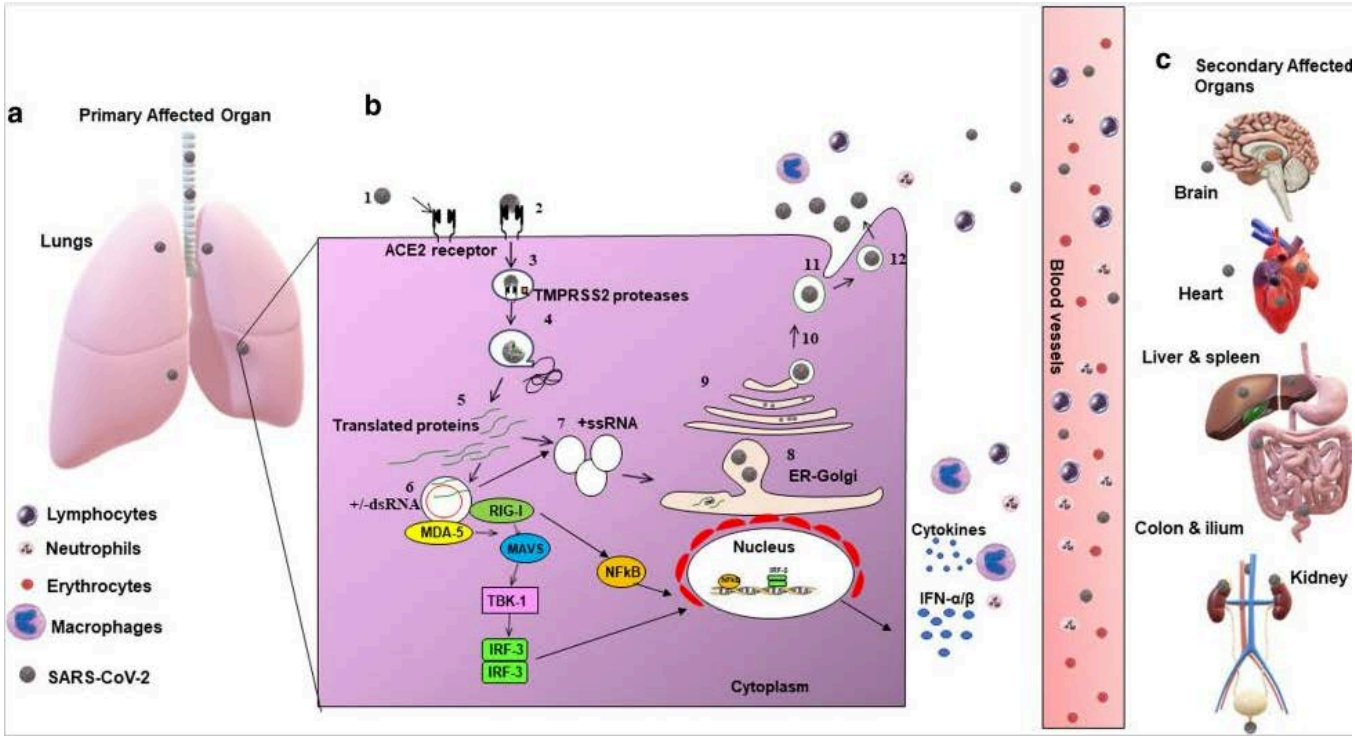


Fig. 2.3.3: Pathogenesis of coronavirus (Copyright Springer Nature, Respiratory research, 2020)

4. Zoonosis

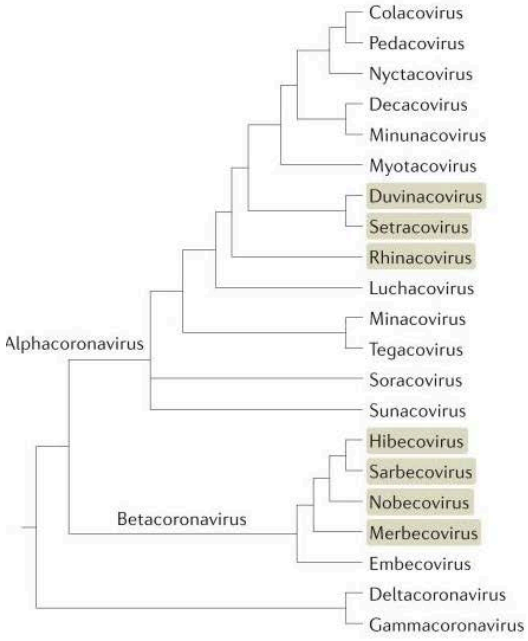
Coronaviruses (CoVs) are commonly found among birds and mammals, with bats serving as the primary evolutionary reservoir and key contributors to CoV diversity. Notable diseases caused by CoVs in these species include transmissible gastroenteritis virus, porcine epidemic diarrhea virus, porcine hemagglutinating encephalomyelitis virus, and murine hepatitis virus. In humans, alpha and beta CoVs can lead to a range of illnesses, from mild, self-limiting respiratory infections (like HCoV-229E, HCoV-NL63, HCoV-OC43, and HCoV-HKU1) to severe acute respiratory distress syndrome (ARDS) **(Umakanthan, S., et al., 2020)**. In an effort to identify potential virus reservoirs, a thorough genetic sequence analysis was conducted across various animal species. The findings indicated that SARS-CoV-2 is a recombinant virus, derived from both a bat CoV and a currently unidentified source. Additionally, a study utilizing relative synonymous codon usage (RSCU) across different animal species pointed to bats as the most likely wildlife reservoir for SARS-CoV-2 **(Umakanthan, S., et al., 2020)**.

4.1 Implications for Public Health and Disease Prevention

The discovery that bats serve as reservoir hosts for coronaviruses has several important implications for public health and disease prevention. It is critical to develop an understanding of the ecological and behavioral influences that lead to coronavirus spillovers from bats to people to design focused surveillance and early-warning programs. It will be critical to strengthen bat monitoring, notably in regions with a substantial biodiversity of bats and a high intensity of human-wildlife encounters to recognize and respond to, with early, possibly embarked upon bat-mediated virus spillover . It will also be critical to employ conservation and biosecurity efforts to limit future human-bat contacts.

4.2 Exploring the Genetic Diversity of Coronaviruses in Bat Populations

Bats are well-known as the natural reservoir hosts of a wide range of coronaviruses from which coronaviruses may have high genetic diversity. Bat virology is important for our understanding of the evolutionary trajectory of these viruses, their potential zoonotic capacity, and the threat of spillover to human and animal populations. Here we review the parametric studies of bat virology conducted to date on genetic diversity and variation studies infected with coronavirus. This review also identifies how the identification of genetic diversity was undertaken and practices to address the reasons for the high diversity.



Genus/subgenus	Notable virus species	Hosts species
Alphacoronavirus		
Colacovirus	Myl-CoV	Bats (Vespertilionidae)
Pedacovirus	PEDV	Bats (Vespertilionidae), pigs
Nyctacovirus		Bats (Vespertilionidae)
Decacovirus		Bats (Hipposideridae, Rhinolophidae)
Minunacovirus		Bats (Miniopteridae)
Myotacovirus		Bats (Vespertilionidae)
Duvinacovirus	HCoV-229E	Bats (Hipposideridae), dromedary camels, alpacas, humans
Setracovirus	HCoV-NL63	Bats (Rhinonycteridae), humans
Rhinacovirus	SADS-CoV	Bats (Rhinolophidae), pigs
Luchacovirus		Rodents (Muridae, Cricetidae)
Minacovirus		Ferrets, minks
Tegacovirus	CCoV, FCoV, TGEV	Cats, dogs, pigs
Soracovirus		Shrews (*Suncus murinus*)
Sunacovirus		Shrews (*Sorex araneus*)
Betacoronavirus		
Hibecovirus		Bats (Hipposideridae)
Sarbecovirus	SARS-CoV, SARS-CoV-2	Bats (Rhinolophidae), Malayan pangolins, carnivores (Canidae, Felidae, Mustelidae, Viverridae), humans
Nobecovirus		Bats (Pteropodidae)
Merbecovirus	MERS-CoV	Bats (Vespertilionidae), dromedary camels, humans
Embecovirus	BCoV, CRCoV, HCoV-OC43, HCoV-HKU1, MCoV	Rodents (Muridae, Cricetidae), dogs, rabbits, cattle, horses, pigs, sable antelopes, dromedary camels, giraffes, humans
Deltacoronavirus	PorCoV-HKU15	Birds, pigs
Gammacoronavirus	IBV	Birds, cetaceans

Fig. 2.4.1: Representation of genetic diversity of bats as coronavirus host. The suggested phylogeny is based on the analysis of complete genomes and/or specific gene segments. The lengths of the branches do not indicate the evolutionary distance between taxa; they are designed solely to clearly depict the relationships among and within genera. Additionally, the distribution of bat species that contain the highlighted subgenera is provided. (Copyright Nat Rev Microbiol., 2021)

4.3 Bats as a carrier of Nipah Virus and its impact on Coronavirus pathogenesis

The Nipah virus (NiV) is a viral pathogen capable of transmitting between animals and humans. It was initially identified in Malaysia in 1998 during an outbreak that resulted in brain inflammation affecting both pigs and humans. NiV is classified under the Paramyxoviridae family and is recognized as a Biosafety Level 4 (BSL-4) pathogen due to its high mortality rate and potential for human-to-human transmission **(Lee, 2012)**. NiV is characterized by an envelope that surrounds its nucleocapsid, which has a helical configuration. During the process of budding from an infected host cell, the virus acquires a portion of the cell's membrane,

forming its viral envelope. This envelope is embedded with two glycoproteins: the fusion (F) protein and the attachment (G) protein. The F protein aids in the fusion of the viral membrane with the host cell membrane, allowing virus entry, while the G protein assists in the attachment of virus to the host cell. The nucleocapsid comprises the viral RNA genome paired with two proteins, the nucleoprotein (N) and phosphoprotein (P), which together form the helical structure of the virus **(McKee, 2021)**.

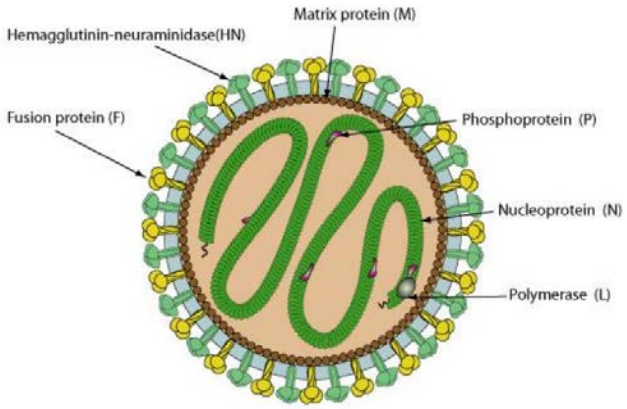


Figure 2.4.2: Nipah virus and its structural components ()

The bat immune system is very unique which allows them to host viruses such as Nipah without susceptibility. Their immune response is characterized by a reduced inflammatory reaction alongside a high metabolic rate, facilitating their coexistence with various pathogenic viruses **(O'Shea et al., 2014)**. These characteristics help minimize the harm caused by viral replication and inflammation, enabling bats to serve as reservoirs for numerous pathogens, including coronaviruses and Henipaviruses. Investigations into the ecology of the Nipah virus reveal how human activities, including deforestation and the expansion of agriculture, influence the rate of zoonotic spillovers. These ecological factors disrupt bat habitats, causing bats to seek food in human habitations, consequently increasing the risk of human exposure to the viruses they host **(Plowright et al., 2015)**. The similarities between Nipah virus spillovers and the emergence of coronaviruses highlight the necessity of understanding the dynamics between bats and viruses. The viruses responsible for the outbreaks of SARS, MERS, and COVID-19—namely SARS-CoV, MERS-CoV, and SARS-CoV-2—have also been linked to bats as their primary

reservoirs. Just like the Nipah virus, coronaviruses are believed to have made the jump to humans through intermediate hosts such as civets or camels, subsequently adapting for human transmission **(Wang et al., 2006; Zhou et al., 2020)**. Research into the mechanisms underlying coronavirus pathogenesis can gain insights from studies focused on Nipah virus. Both types of viruses attach to specific cellular receptors for entry; Nipah virus interacts with ephrin-B2 and ephrin-B3, while coronaviruses use either ACE2 (angiotensin-converting enzyme 2) or DPP4 (dipeptidyl peptidase 4), depending on the strain. These findings emphasize the importance of exploring receptor-ligand interactions in bats to anticipate and mitigate future zoonotic outbreaks. The rise of zoonotic diseases such as Nipah virus and coronaviruses underscores the pressing necessity for comprehensive global surveillance and preventative strategies. Nipah virus, known for having a case fatality rate of as high as 90%, poses a significant danger, especially in South Asia, where outbreaks frequently occur **(CDC, 2019)**. Conversely, the widespread impact of SARS-CoV-2, characterized by its extensive transmission yet lower case fatality rate, illustrates the potential of zoonotic diseases to severely disrupt social structures and economies.

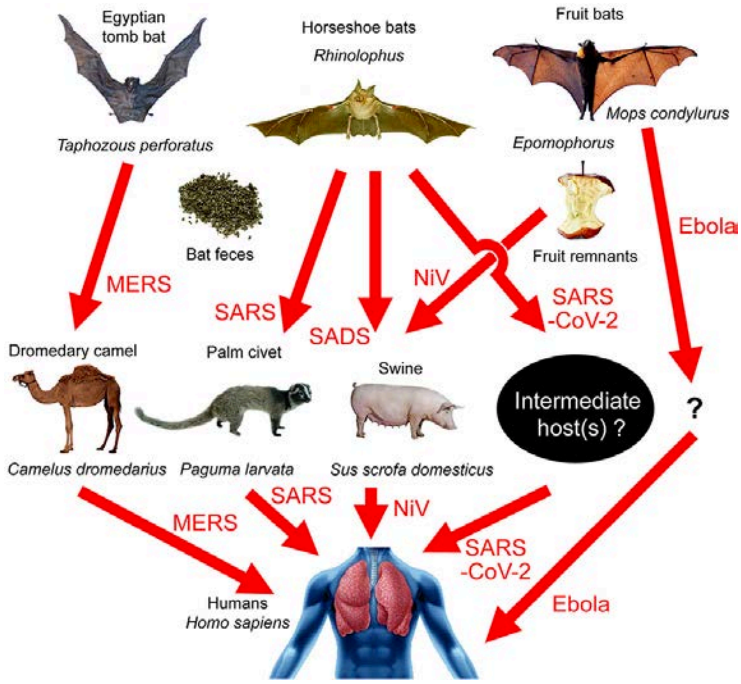


Figure 2.4.3: Coexistence of SARS, MERS, SARS-CoV-2 and NiV in host bat (Copyright Front. Vet. Sci., 09 June 2020)

4.4 Review of Studies Investigating Genetic Diversity of Coronaviruses in Bat Populations Worldwide

Several studies have reported the wide distribution of coronaviruses in the bat populations of different regions and habitats worldwide. These analyses are performed through numerous means such as viral surveillance, molecular characterization, and phylogenetic analysis to evaluate bat coronaviruses genetic variations. Some of the examples are explained above; for instance, based on the study by **(Anthony et al., 2017)** conducted viral surveillance in bats in Madagascar, new isolates were discovered from both the Alphacoronavirus and Betacoronavirus genera. Similarly, **(Drexler et al., 2014)** conducted a surveillance study in the European region and discovered a wide range of novel coronaviruses that possess distant genetic lineages.

5. Methodologies Used to Assess Genetic Diversity

In coronaviruses, the evaluation of genetic differentiation is based on molecular biology methods and bioinformatics. Virus identification and characterization usually consist of bat sampling, viral RNA extraction, performing a semi nested polymerase chain reaction (PCR), and sequence targeted genomic fragments, such as the spike (S) or polymerase (RdRp) genes. A phylogenetic method can be applied early to evaluate bat coronavirus genetic sequence homology with well-known reference lines, which determined evolutionary correlations and diversity.

The development of sequencing technologies, including sequencing methods like sanger sequencing, has greatly advanced research into viral diversity by allowing deep, intensive sequencing of complete viral genomes from random samples. Such metagenomic analyses permit the identification of unknown coronaviruses using established approaches that do not require prior information about viral sequences. Bioinformatics have incorporated computational algorithms to mine useful information from high-throughput sequenced datasets by discovering genetic mutations and reconstructing viral phylogenies.

5.1 Analysis of Factors Contributing to Genetic Variability in Coronaviruses

Host ecology, geographic distribution, host-virus interactions, and viral evolutionary mechanisms play a role in the observed genetic variability of coronaviruses among bat populations. Bats exhibit a large variation in ecological niches with some species such as the cave dwellers and the migratory species having a large range of immigration and resident species that have thus created different reservoirs of the virus due to the different species and environment that is occupied. The wide variation facilitates a high rate of transmission and adaptation to different species and environments. Geographic distribution has affected not only the host but has also shaped the viral lineages of bat coronaviruses. The genetic diversity of the host region has impacted the variation of the viral species of the bat. Some studies have shown unique coronaviruses to have effects in bat populations that are observed in Asia, Africa, Europe and the Americas.

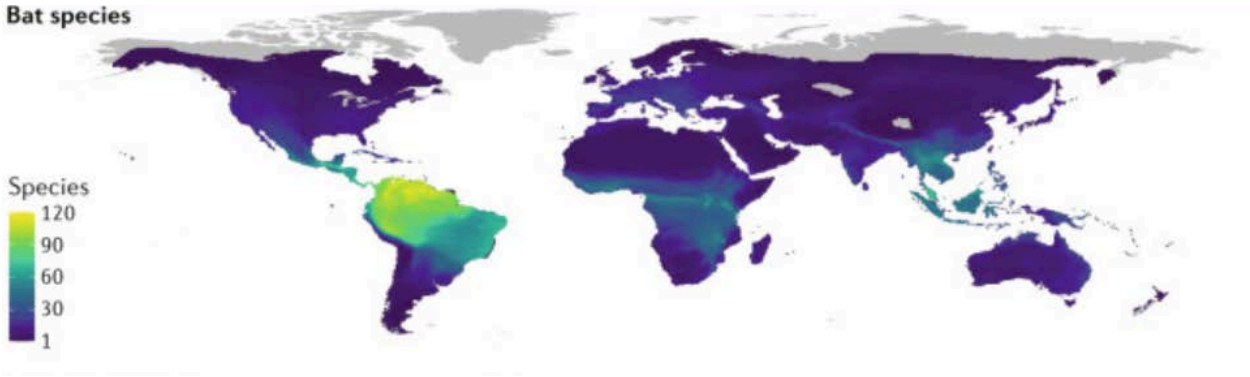


Fig. 2.5.1: Distribution of bat species around the globe (Copyright Nat Rev Microbiol., 2021)

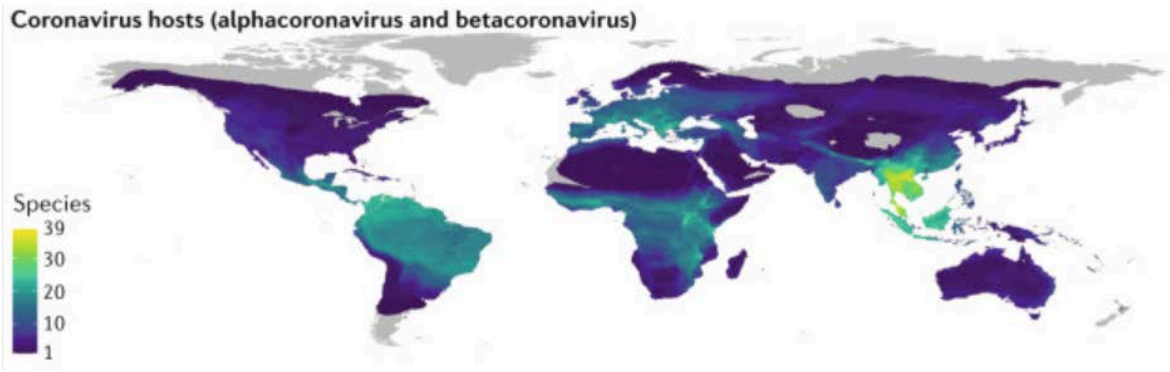


Fig. 2.5.2: Prevalence of alpha and beta coronavirus across the globe (Copyright Nat Rev Microbiol., 2021)

Genetic recombination associated with host-virus interactions plays a crucial role in determining the genetic evolution of coronaviruses. Indeed, these processes enable host switching, mutation and recombination, and the occurrence of new viral strains. Bats have a vast variety of coronaviruses and are potential sources of multiple zoonotic events and diseases. The recombination of coronaviruses from different viral species may alter virulence, transmissibility and host range. Similarly, the recombination within coronaviruses of the same species enables the emergence of new strains. The viral genetic diversity in bats depends on the interplay of several ecological and evolutionary aspects, such as proximity between hosts, environment sharing, and genetic background. Surveillance studies conducted in several continents have generally revealed that bats house vast repertoires of different coronaviruses, demanding constant monitoring activities for pandemic preparedness and intervention.

5.2 Cross-Species Transmission Events Involving Coronaviruses due to host

Coronaviruses demonstrate a high capacity to infect a broad spectrum of host species, including mammals and birds. Furthermore, a number of prominent cross-species transmission events have been observed, resulting in the emergence of novel coronaviruses with zoonotic potential. SARS-CoV-2 emerged in humans in 2002-2003, likely owing to zoonotic spillover from bats through other mammals, such as civet cats, which were sold at a live animal market **(Li et al., 2005)**. Moreover, the Middle East Respiratory Syndrome Coronavirus was first identified in humans in 2012 and transmitted from infected dromedary camels **(Memish et al., 2020)**.

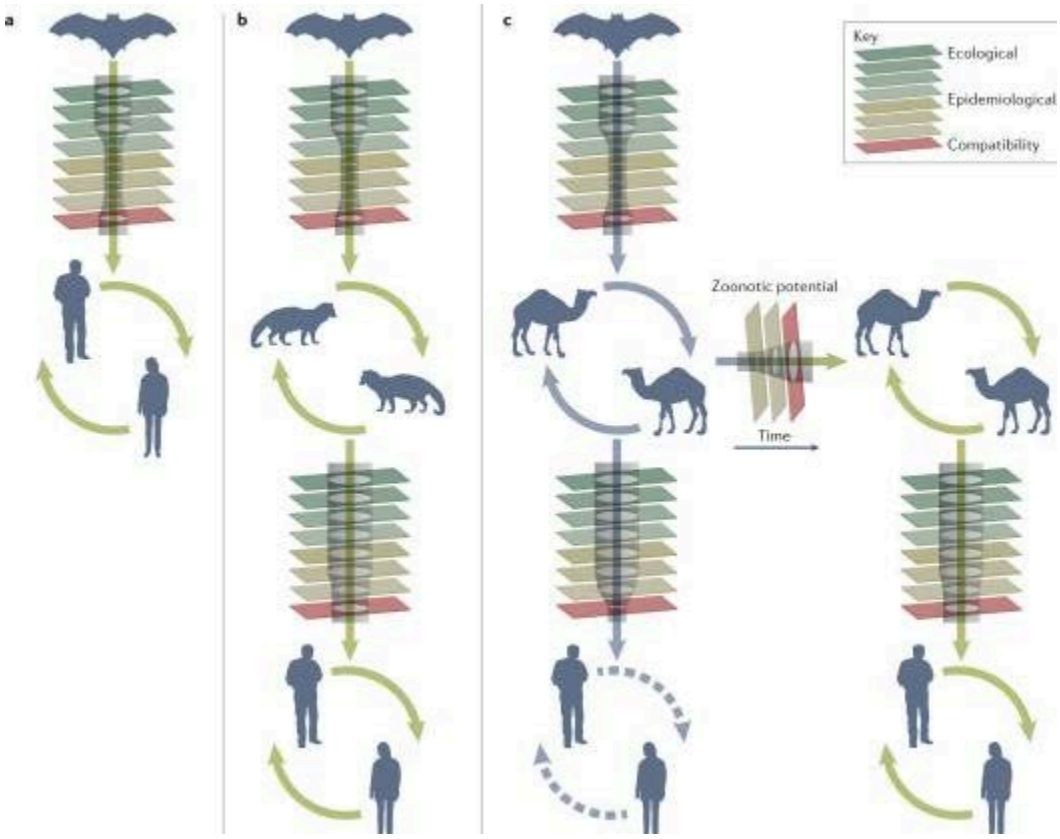


Fig. 2.5.3: Zoonotic potential of bat coronavirus in cross species transmission. In all three scenarios a,b and c, the bat acts as the primary host and the final affected is human(Copyright Nat Rev Microbiol., 2021)

COVID-19, an ongoing pandemic resulting from a zoonotic transmission event of the novel coronavirus virus SARS-CoV-2, continues to emphasize the potential threat of cross-species transmission events applied by the host-virus coevolution term. Bats are believed to be the zoonotic source, with possible intermediate pangolins in the SARS-CoV-2 zoonotic transmission chain, the animal host **(Zhang et al., 2020)**. Cross-species transmission has been facilitated by the genetic adaptation in the mink-adapted variant V656F that gains increased fitness in human cells of SARS-CoV-2.

5.3 Review of Genetic Adaptations Facilitating Transmission to New Hosts

Cross-species transmission events often require genetic adaptations that enable the virus to overcome species-specific barriers and establish productive infections in new hosts. Genetic

changes may occur in various regions of the viral genome, including the spike (S) protein and the polymerase (RdRp) gene, which is essential for viral replication and transcription.

One of the key genetic adaptations associated with host switching in coronaviruses is the receptor-binding domain (RBD) of the S protein, which interacts with host cell receptors to initiate infection. Changes in the receptor-binding domain (RBD) can modify how effectively the virus interacts with host receptors, promoting the possibility of cross-species transmission. Research has shown that certain mutations in the S protein of SARS-CoV and MERS-CoV improve their capacity to bind to human receptors, which allows for more effective replication and spread among human populations **(Li et al., 2005; Lu et al., 2015)**. Additionally, alterations in other viral genes may play a role in the adaptation of coronaviruses to infect new hosts. For example, changes in the RdRp gene can affect viral replication fidelity, RNA synthesis, and the ability to evade host immune responses. Furthermore, genetic recombination between different coronaviruses or host cellular genes can generate novel viral strains with altered virulence, host range, and transmission dynamics.

5.4 Implications for Understanding the Potential for Future Outbreaks

Potential future outbreak predictions and informed strategy planning for prevention and control can be done through studies of cross-species transmission events and genetic adaptation in coronaviruses. Studying the genetic basis for host specificity and transmission potential will improve the identification of high-risk viral strains and surveillance targets, inform resource allocation for monitoring and response, and vaccines and therapeutics targeting conserved viral epitopes.

In addition, identifying the ecological, environmental and anthropogenic drivers of cross-species transmission events is needed to predict and prevent potential future outbreaks. Particularly in the context of wildlife trade, human encroachment into natural ecosystems leads to more interactions between humans and domesticated animals or wild flora and fauna over time–which increases opportunities for the exchange of pathogens from animal to humans.

Further, it requires integrated approaches incorporating virology, epidemiology, ecology and social sciences to tackle the multifaceted drivers of zoonotic disease emergence and inform sustainable interventions based on One Health principles.

Interspecies metastasis and genetic adaptation are the major risk factors for viruses to evolve into novel coronaviruses which can cause worldwide spread. This line of research not only informs mechanisms of viral evolution, but also through knowledge of the genetic determinants of host specificity and transmission efficiency helps predict the potential for future outbreaks. This is important information with which to inform disease surveillance, prevention and control strategies that adopt One Health approaches acknowledging the interrelationship of human, animal and environmental health.

6. Surveillance and Monitoring of Bat Coronaviruses: Current Approaches and Challenges

These challenges are intensified by likely emerging coronaviruses that pose an even more significant threat to public health, so the role of surveillance and monitoring in better understanding viral ecology as well as minimizing zoonotic spillover is essential. It describes monitoring of coronaviruses circulating in bat populations, assesses current surveillance methodologies and technologies, and highlights challenges and limitations to surveillance.

6.1 Overview of Surveillance Efforts

Surveillance of coronaviruses in bats involves activities like field sampling, lab testing and data analysis. Such initiatives frequently include collaborations among virologists with ecological and wildlife biologists, as well as public health professionals. Field sampling — Collection of fecal samples, oral swabs, or tissue from a wild bat population captured by mist netting, harp trapping or acoustic monitoring.

After they are collected, samples are sent to laboratories for the detection and characterization of viruses. Molecular techniques to identify coronaviruses in samples from bats, like detection of

viral RNA using PCR (polymerase chain reaction). The use of metagenomic sequencing approaches, such as Sanger sequencing, allows for the characterization of the diversity of viruses in bats and the discovery of new coronaviruses.

6.2 Challenges and Limitations in Surveillance Efforts

Despite the progress made in surveillance of bat coronaviruses, several challenges and limitations persist. One of the main challenges is the vast diversity of bat species and habitats, which makes comprehensive surveillance efforts logistically challenging and resource-intensive. Bats exhibit diverse ecological niches, ranging from cave-dwelling species to migratory bats that traverse vast distances, complicating efforts to sample representative populations.

Furthermore, the emergence of novel coronaviruses requires ongoing monitoring of bat populations over time to detect changes in viral diversity and identify potential spillover events. Longitudinal surveillance studies are essential for understanding the dynamics of viral transmission within bat populations and predicting the risk of zoonotic transmission to humans and other animals.

Another challenge is the limited understanding of the ecological and environmental factors driving viral transmission and spillover events. Factors such as habitat destruction, wildlife trade, and human encroachment into natural ecosystems can increase the frequency and intensity of interactions between humans, domestic animals, and wildlife, facilitating the spillover of pathogens from animals to humans.

Surveillance and monitoring of bat coronaviruses are essential for detecting emerging threats, understanding viral ecology, and mitigating the risk of zoonotic transmission. Current methodologies and technologies, including molecular techniques and metagenomic sequencing, enable researchers to detect and characterize viral diversity with greater sensitivity and accuracy.

However, challenges such as logistical constraints, limited resources, and gaps in knowledge about viral ecology and transmission dynamics remain. Addressing these challenges requires interdisciplinary collaborations, innovative surveillance strategies, and sustained investment in research and capacity-building efforts.

By enhancing our understanding of the ecological, environmental, and anthropogenic factors driving viral transmission, we can improve surveillance efforts and develop evidence-based strategies for disease prevention and control, ultimately safeguarding human and animal health.

7. Vaccine Development for Bat-Origin Coronaviruses

Vaccine development for bat-origin coronaviruses represents a crucial area of research for preventing future outbreaks and mitigating the impact of zoonotic transmission events. While vaccines against human coronaviruses such as SARS-CoV-2 have been developed and deployed with unprecedented speed, efforts to develop vaccines targeting bat-origin coronaviruses are still in their infancy.

Recent studies have explored the potential for cross-protection against bat-origin coronaviruses using vaccines developed for related human coronaviruses. For example, a study by **(Menachery et al., 2016)** demonstrated that a SARS-CoV vaccine provided protection against heterologous bat coronaviruses in a mouse model, suggesting the potential for broad-spectrum coronavirus vaccines. Similarly, efforts to develop MERS-CoV vaccines have provided insights into vaccine platforms and immunization strategies that could be applied to bat-origin coronaviruses.

Future research should focus on the development of vaccines that target conserved viral epitopes shared among bat-origin coronaviruses, allowing for cross-protection against a broad range of viral strains. Additionally, studies on the immune responses of bats to coronaviruses could inform vaccine design and immunization strategies that induce robust and long-lasting immunity.

8. Suggestions for Future Research Directions

Several key research priorities can help advance our understanding of bat-origin coronaviruses and inform strategies for disease prevention and control:

1. **Ecological and Behavioral Studies:** Investigating the ecological and behavioral factors that influence viral transmission and spillover events in bat populations is essential for predicting and mitigating the risk of zoonotic transmission. Longitudinal studies on bat populations, habitat ecology, and interactions with other animals can provide insights into the dynamics of viral transmission and identify high-risk areas and species.
2. **Viral Evolution and Genetic Diversity:** Understanding the evolutionary dynamics of bat-origin coronaviruses, including genetic diversity, recombination, and mutation rates, is crucial for predicting the emergence of novel viral strains and designing effective surveillance and monitoring strategies. Studies on viral evolution can help identify genetic determinants of host specificity, transmission efficiency, and virulence, informing vaccine development and therapeutic interventions.
3. **One Health Approaches:** Adopting One Health approaches that integrate human, animal, and environmental health perspectives is essential for addressing the complex drivers of zoonotic disease emergence and transmission. Collaborative research efforts involving multidisciplinary teams of virologists, ecologists, epidemiologists, and social scientists can facilitate the development of holistic strategies for disease surveillance, prevention, and control.

Advancing our understanding of bat-origin coronaviruses and mitigating the risk of zoonotic transmission require concerted research efforts, collaborative partnerships, and sustained investment in surveillance, monitoring, and vaccine development. By prioritizing research on viral evolution, ecological dynamics, and One Health approaches, we can develop

evidence-based strategies for disease prevention and control that protect human and animal health and safeguard global health security.

9. Objective of this Study

The study we designed for the detection and characterization of coronaviruses in bat populations of Nipah-prone areas was aimed to establish co-phylogenetic relations, determinants of spillover events, host specificity, viral recombinant features of circulating coronaviruses. The purpose of this study are:

- Molecular detection and characterization of coronavirus in the collected bat samples
- Identification of mutations and mutational analysis of targeted RdRp gene in bat coronavirus
- Establishing correlation between potential cross species transmission, cospeciation and host switching with viral recombination

We plan to use both data collection efforts across various nipah-prone habitats and regions in order to determine areas and species within which there is a high risk of zoonotic transmission, understand hotspots of emerging viral threats, and inform prevention strategies, and metagenomic analysis of the retrieved bat coronavirus in the future.

9.1 Description of Study Design

The bat coronavirus surveillance study that was conducted uses different approaches of sampling collection, viral detection and genetic analysis. The research involves field sampling of populations of bats across a variety of nipah prone habitats and regions, employing mist netting, harp trapping, and acoustic monitoring techniques. Samples of feces, oral swabs and tissues are

collected and transferred to the laboratory for test determination, according to viral detection here.

It includes the use of molecular techniques in the laboratory for detection of viral RNA and coronaviruses using methods such as conventional semi nested polymerase chain reaction (PCR) on bat samples. Further, metagenomic sequencing approaches (cycle sequencing (sanger sequencing)) allow us to study the entire viral diversity within bat populations and can result in the identification of new coronaviruses or sequence variants from previously known coronaviruses.

Chapter – 3

Materials and Methods

1. Sample Collection

The capture and handling of bat species and their samples were performed by trained personnel in accordance with established safe animal capture and sampling protocols. Physical restraint of the bats was achieved using comprehensive personal protective equipment, which included nitrile gloves, N95 masks, goggles, Tyvek-type suits, and leather gloves. Bats were captured using mist nets at five distinct locations: Dhaka, Cox's Bazar, Faridpur, Sylhet, and Rajshahi (nipah prone areas). A total of 1,668 bat samples were collected (mostly *Pteropus medius* along with some small fruit bats), comprising five different types of specimens: oropharyngeal swabs, pooled roost urine and feces, throat swabs, and urogenital swabs. These samples were collected in 3.5 ml cryo-vials containing 1 ml of TRIzol (Zymo Research), stored in liquid nitrogen in the field, and subsequently transported to the laboratory, where they were stored in an -80°C freezer pending further analysis. Following the collection of samples and necessary data, the bats were released back into their natural habitat.

Fig. 3.1.1: Swab sample from bat

Fig. 3.1.2: *Pteropus medius* bat (Indian flying fox)

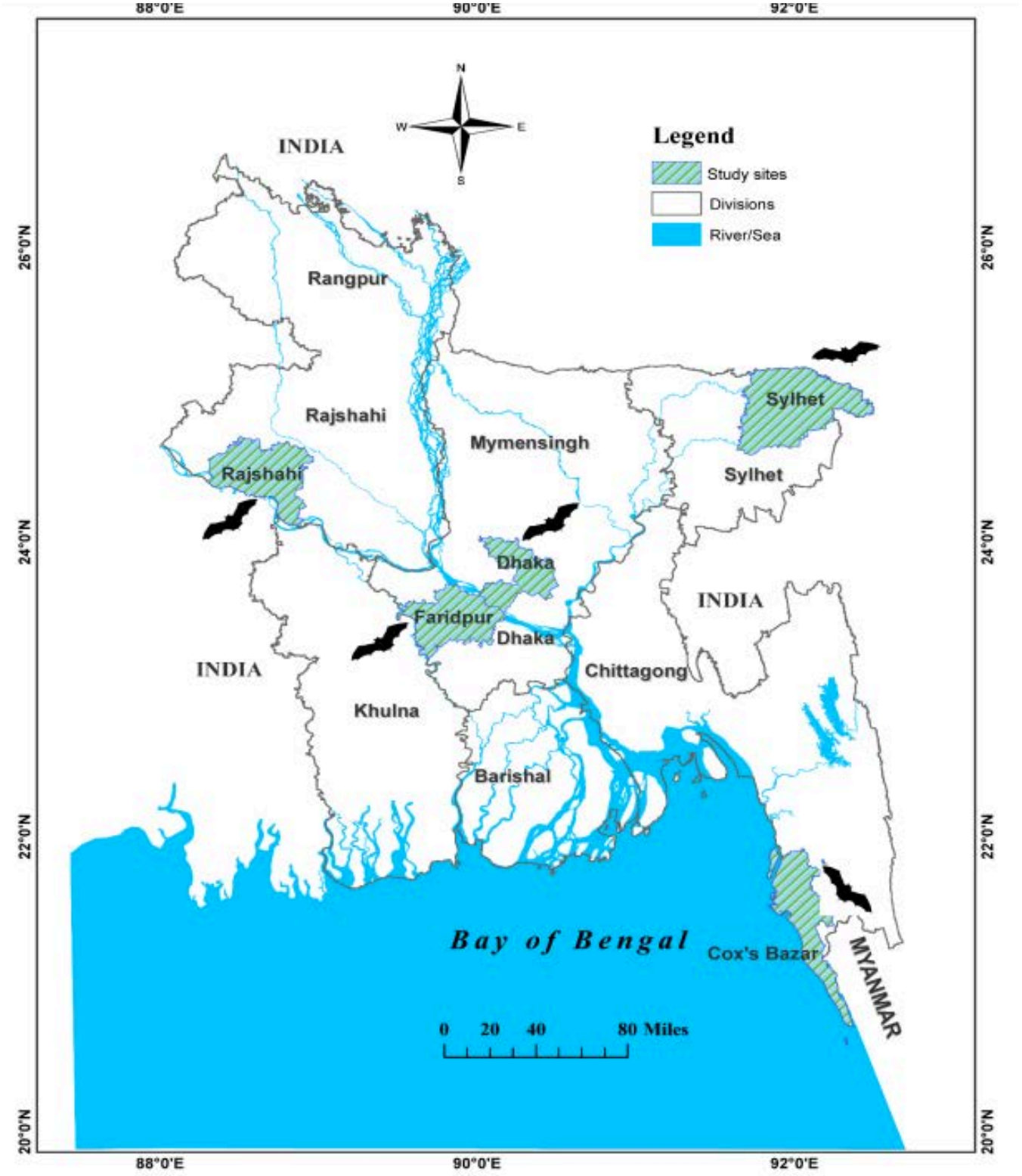


Fig. 3.1.3: Bat sample geographical distribution across Bangladesh (Nipah virus prone areas)

2. Experiment Design

The flowchart in the figure below illustrates the chronological order in which this research was carried out, providing an experimental model for further investigations.

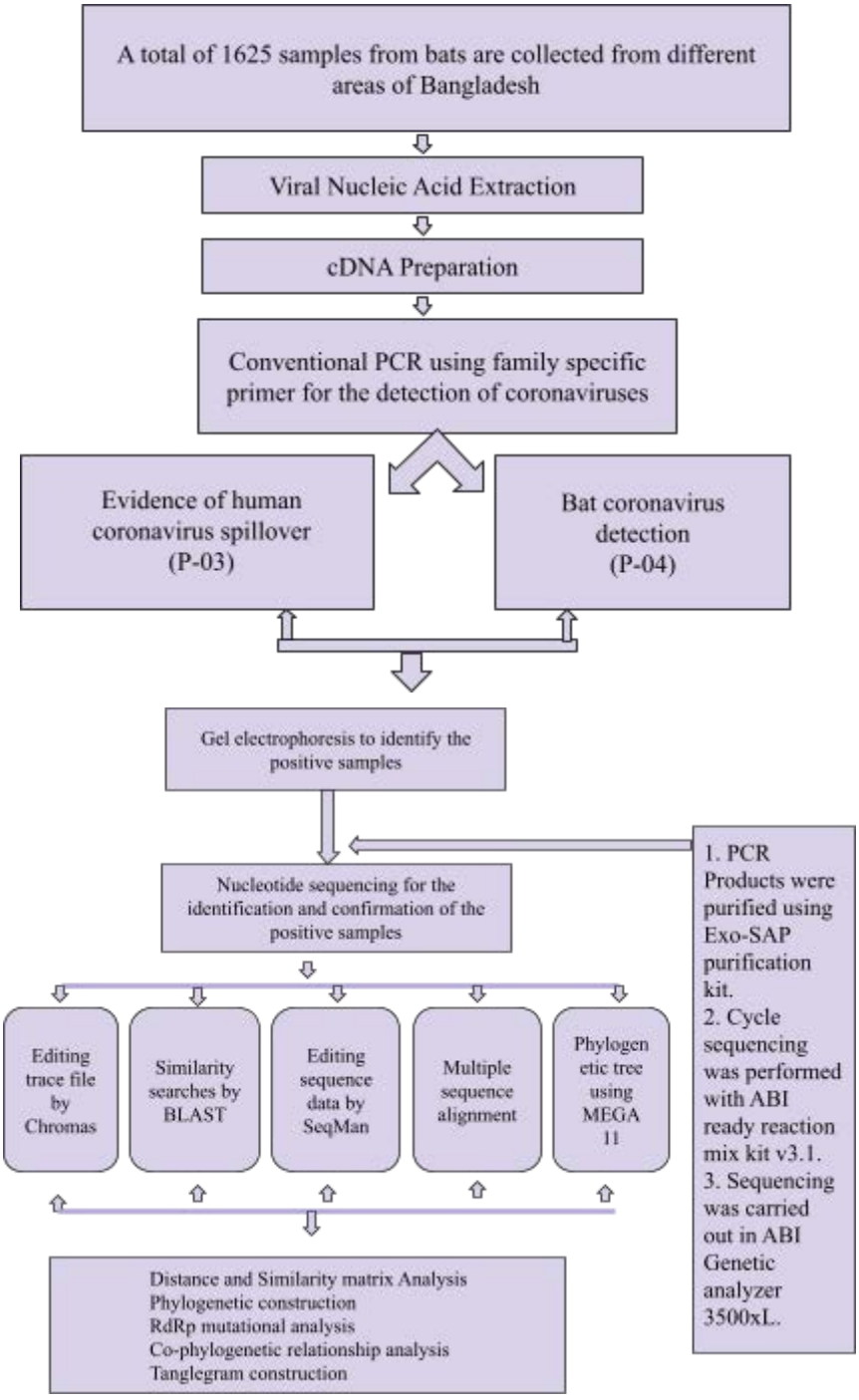


Fig. 3.2.1: Chronological representation of research procedure

3. Sample Preparation

The samples were stored and preserved in trizol and were subjected to vigorous vortexing for a duration of 5 seconds within a biosafety cabinet. Subsequently, centrifugation of the sample was done for 30 seconds at 2000 rpm using a swinging bucket centrifuge.

4. Viral Nucleic Acid Extraction

The Direct-zol™ RNA MiniPrep Plus Kit **(zymoresearch.eu, 2020)** was employed for the extraction of viral nucleic acids following the manufacturer's guidelines. This kit facilitates a spin-column-based purification of total RNA, inclusive of small and microRNAs, directly from TRIzol. The kit components consist of TRI Reagent (TRIzol), RNA PreWash 1, RNA Wash Buffer 2, DNase I, DNA Digestion Buffer, DNase/RNase-Free Water, Zymo-Spin columns, and Collection Tubes. The procedure is divided into four steps: lysis, binding, washing, and elution.

For viral nucleic acid extraction, following steps were performed:

I. Reagent Preparation

- Ethanol (95-100%) was added into 160 ml of concentrated Direct-zol™ RNA PreWash in a volume of 40 ml.
- Subsequently, 100% ethanol was added to 12 ml of concentrated RNA Wash Buffer, yielding a total volume of 48 ml.
- Lyophilized DNase I was reconstituted using DNase/RNase-Free Water, gently mixed by inversion, and subsequently stored in frozen aliquots.

II. RNA Extraction & Purification Procedure

- 200 µL of the liquid sample was transferred into a 1.5 mL RNase-free microcentrifuge tube.
- 600 µL of GENEzol™ Reagent (3 volumes) to the 200 µL sample (3:1 ratio) was added and mixed thoroughly by vortexing. Incubation of sample solution is done at room temperature for 5 minutes.

- The mixture was centrifuged at 12,000–16,000 × g for 1 minute to separate cell debris, then transferring the clear supernatant to a new 1.5 mL RNase-free microcentrifuge tube.
- 200 µL of absolute ethanol was added to the supernatant (1:1 ratio) in GENEzol™ Reagent, mixed well by vortexing, and a 2 mL collection tube was placed into an RB spin Column.
- 400 µL of the sample was transferred to the RB spin Column and centrifuged at 14,000–16,000 × g for 1 minute; the flow-through was discarded, any remaining sample mixture was transferred to the RB Column.
- 400 µL of Wash Buffer (with ethanol) was added to the RB Column and centrifuged at 14,000–16,000 × g for 30 seconds, discarding the flow-through and placing the RB Column back in the collection tube.
- The DNase 1 solution was then gently mixed, and 50 µL was added to the center of the RB column matrix, then incubated the column at room temperature (20-30 °C) for 15 minutes before proceeding with RNA Wash.
- 400 µL of Pre-Wash Buffer (with ethanol) was added to the RB Column, centrifuged at 14,000–16,000 × g for 30 seconds, discarding the flow-through, and returning the RB Column to the collection tube.
- 600 µL of Wash Buffer (with ethanol) was again added to the RB Column, centrifuged at 14,000–16,000 × g for 30 seconds, discarding the flow-through, and RB Column was kept in the collection tube.
- The column matrix was dried by centrifuging at 14,000–16,000 × g for 3 minutes, then returned the RB Column to the collection tube.
- 25–50 µL of RNase-free water was added to the center of the column matrix and incubated for 3 minutes for complete absorption.
- Finally, centrifugation at 14,000–16,000 × g for 1 minute was performed to elute the purified RNA.

5. Complementary DNA (cDNA) preparation

A High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA, USA) was employed to synthesize cDNA from the extracted RNA samples. The components of the cDNA Kit, which include Buffer, dNTPs, RNase inhibitor, reverse transcriptase enzyme, random

primer, and nuclease-free water, were combined to achieve a volume of 10 µL for each reaction tube. An additional 10 µL of RNA was added to each reaction, resulting in a total volume of 20 µL (Table 3.5.1).

Table 3.5.1: Master mixture for cDNA preparation

Reagent	Amount (µL)
Buffer (10×)	2.0
dNTP (25x)	0.8
RNase inhibitor (RNasin)	1.0
Random primer	2.0
RT mix	1.0
Nuclease free water	3.2
Total	**10 µL**

During the initial step of the PCR process, the temperature was maintained at 25°C for 10 minutes to facilitate the extension of oligo(dT) and random primers prior to the commencement of cDNA synthesis. The temperature was subsequently elevated to 37°C for a duration of 2 hours, during which RNA was transcribed into cDNA. Finally, the temperature was increased to 85°C for 5 minutes to inactivate the enzyme, followed by a hold at 4°C indefinitely.

6. Coronaviruses detection using semi nested real time PCR (Conventional PCR)

PCR amplification of coronaviruses was performed using the Hotstartaq DNA polymerase kit (Qiagen, 2016) through a semi nested PCR approach.

6.1 Enzymes

The QIAGEN PCR enzyme mixture comprises HotstarTaq DNA polymerase, which is a modified version of a recombinant 94-kDa DNA polymerase originally derived from *Thermus aquaticus* and expressed in *Escherichia coli*.

6.2 Buffers and Reagents

1. QIAGEN-PCR buffer (10X concentrated)
2. dNTPs Mixture: 10 mM each dATP, dCTP, dTTP, dGTP
3. RNase-free water

6.3 Semi-Nested PCR Assay

The thermal cycling protocol was employed to amplify the RdRp gene of coronaviruses using designated primer sets. The amplification process adhered to the semi-nested conventional PCR method.

6.4 Primer sets for the amplification of target fragments

To amplify the partial sequence (RdRp region) of the ORF1b gene, a sufficient quantity of reagents—including RNase-free water, 10X buffer, dNTPs, and enzyme—was combined with the corresponding primers specific to the ORF1 gene. The primers utilized in this study are detailed in Tables 3.6.1 and 3.6.2.

Protocol-03

This protocol specifically focuses on amplifying the RdRp gene of the Human Coronavirus, targeting the region between 17,480 and 17,820 of the genome for strain 229E. This could identify human coronavirus strains in bat samples

Table 3.6.1: 1st round PCR Primer set for the amplification of RdRp gene

Primer	Strand	Sequence(5′-3′)	Amplicon Size (bp)
CoV-F1	Forward (Plus)	CGTTGGIACWAAYBTVCCWYTICARBTR GG	520 bp
CoV-R1	Reverse (Minus)	GGTCATKATAGCRTCAVMASWWGCNAC ATG	

Table 3.6.2: 2nd round PCR Primer set for the amplification of RdRp gene

Primer	Strand	Sequence(5′-3′)	Amplicon Size (bp)
CoV-F2	Forward (Plus)	GGCWCCWCCHGGNGARCAATT	328 bp
CoV-R2	Reverse (Minus)	GGWAWCCCCAYTGYTGWAYRTC	

Semi nested PCR was employed to amplify the highly conserved region of the RNA dependent RNA Polymerase (RdRp) gene (approximately 328-520 bp) for the detection of viruses belonging to the *Coronaviridae* family.

Protocol-04

This assay also targets coronavirus's polymerase gene (RdRp) in bat populations. However, it targets a different region that is slightly more upstream. The Human Coronavirus genome (Strain 229E) is targeted upto roughly 14,370-14,750 nucleotides. This could be beneficial in our research to identify coronaviruses in bats, as numerous partial sequences of bat coronaviruses published so far have been obtained using these primers. It is worth mentioning that this assay has been updated since its original launch; modifications to the original primer sequences have enhanced its capacity to detect a broader range of coronaviruses.

Table 3.6.3: 1st round PCR Primer set for the amplification of RdRp gene

Primer	Strand	Sequence(5′-3′)	Amplicon Size (bp)
CoV-F3	Forward (Plus)	GGTTGGGAYTAYCCHAARTGTGA	440 bp
CoV-R3	Reverse (Minus)	CCATCATCASWYRAATCATCATA	

Table 3.6.4: 2nd round PCR Primer set for the amplification of RdRp gene

Primer	Strand	Sequence(5′-3′)	Amplicon Size (bp)
CoV-F4	Forward (Plus)	GAYTAYCCHAARTGTGAYAGAGC	434 bp
CoV-R3	Reverse (Minus)	CCATCATCASWYRAATCATCATA	

6.5 PCR 1st Round

Protocol-03

In the initial round of PCR, sequences approximately 520 base pairs in length were amplified. The thermal cycling conditions are detailed in Table 3.6.5.

Table 3.6.5: Round 1 semi nested PCR cycle

Steps	Temperature	Duration
Enzyme activation	95℃	5 minutes
3 steps cycling (15 cycle)		
Denaturation	95℃	30 seconds
Annealing	65℃	30 seconds
Extension	72℃	45 minutes

Protocol-04

In the initial round of PCR for protocol-04, sequences approximately 440 base pairs in length were amplified. The thermal cycling conditions are detailed in Table 3.6.6.

Table 3.6.6: Round 1 semi nested PCR cycle

Steps	Temperature	Duration
Enzyme activation	94℃	2 minutes
3 steps cycling (15 cycle)		
Denaturation	95℃	20 seconds
Annealing	20℃	20 seconds
Extension	72℃	30 seconds

6.6 PCR 2nd Round

Protocol-03

The thermal cycling conditions for this round were adjusted slightly compared to those in round one. The specific thermal cycling conditions for round 2 PCR are detailed in Table 3.6.7

Table 3.6.7: Round 2 semi nested PCR cycle

Steps	Temperature	Duration
Enzyme activation	95℃	5 minutes
3 steps cycling (35 cycle)		
Denaturation	94℃	30 seconds
Annealing	50℃	30 seconds
Extension	72℃	45 seconds
Final extension	72℃	7.0 minutes
Holding	4℃	∞

Protocol-04

In the second round of PCR, distinct forward primers were employed alongside the same reverse primer used in the first round for amplification.

Table 3.6.8: Round 2 semi nested PCR cycle

Steps	Temperature	Duration
Enzyme activation	94℃	2 minutes
3 steps cycling (35 cycle)		
Denaturation	95℃	20 seconds
Annealing	20℃	20 seconds
Extension	72℃	30 seconds
Final extension	72℃	7.0 minutes
Holding	4℃	∞

Fig. 3.6.1: Protocol-03 & protocol-04 on T100 Thermal Cycler

6.7 PCR Control:

To assess the proper PCR run, Universal Control 1 is employed, as it binds effectively to the primers and signifies an appropriate PCR reaction.

6.8 Preparation of Master Mixture

The components of the Qiagen HotStarTaq® PCR kit, including the 10X buffer, dNTPs, and enzyme, were mixed with the corresponding primers at a concentration of 20µM to achieve a final master mixture volume of 19µL for each reaction tube, as detailed in Table 3.6.9.

Table 3.6.9: Reagents of Master Mix (P-03)

Reagent	Quantity
Nuclease free water	10.5
Buffer (10x)	4.0
dNTP	0.4
Forward primer (F1/F2)	0.5
Reverse primer (R1/R2)	0.5
Pol (enzyme) (5u/µl)	0.1
MgCl2 (25mM)	2.0
Total	**18**

Table 3.6.10: Reagents of Master Mix (P-04)

Reagent	Quantity
Nuclease free water	10.5
Buffer	4.0
dNTP	0.4
Forward primer (F1/F2)	0.5
Reverse primer (R1)	0.5
Pol (enzyme)	0.1
MgCl2 (25mM)	2.0
Total	**18**

6.9 Analysis of amplified PCR Product

The amplified products were visualized using agarose gel electrophoresis. A total of 6 µL of each PCR product was loaded into separate wells of a 1.5% agarose gel, which was stained with GelRed. A 100 bp DNA ladder (Invitrogen) served as a molecular marker. Electrophoresis was conducted on an Apogee Horizon 20.25-Horizontal Electrophoresis Apparatus at a voltage of 170 volts for a duration of 50 minutes. Subsequently, the DNA bands were visualized with a X-ray Gel doc (BioDoc-Bio Rad systems).

Fig. 3.6.2: Apogee Gel apparatus and Biorad Gel Doc for electrophoresis and visualization

7. The Sequencing of amplified PCR amplicons

The nucleotide sequences of the cDNAs were verified employing the dideoxynucleotide chain termination method **(Sanger, Nicklen et al., 1977)**. Sequencing of the PCR products was performed utilizing the ABI BigDye Terminator v.3.1 Cycle Sequencing Reaction Kit (Applied Biosystems, USA). The sequencing process was conducted on an automated ABI 3500 XL genetic analyzer (Applied Biosystems, Foster City, USA) using the forward and reverse primers independently.

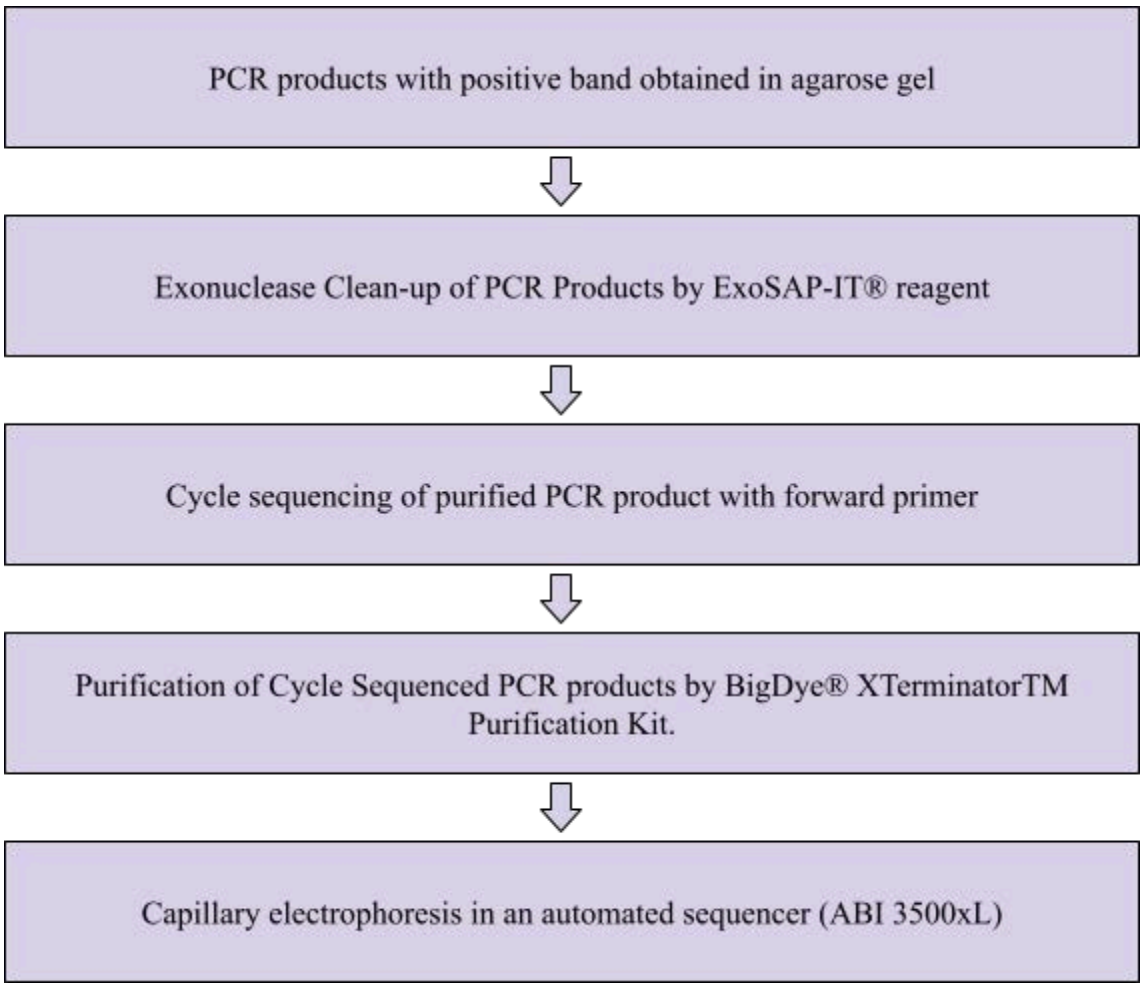


Fig. 3.7.1: Sanger sequencing procedure in chronological order

Fig. 3.7.2: ABI 3500xL genetic analyzer (Applied Biosystem)

7.1 Exonuclease Clean-up of PCR Products

In this study, the ExoSAP-IT® protocol (Affymetrix, California, USA) was utilized to clean-up PCR products prior to sequencing. This methodology employs two hydrolytic enzymes: Exonuclease I and Shrimp Alkaline Phosphatase (Usher, 2022). Exonuclease I serves to remove any remaining primers and single-stranded DNA fragments generated during the PCR process. Conversely, Shrimp Alkaline Phosphatase functions by deactivating unincorporated deoxyribonucleotide triphosphates (dNTPs) present in the PCR mixture, which may interfere with the sequencing procedure and lead to the production of an incomprehensible sequence.

7.2 Procedure

1. A total of 2 µL of ExoSAP-IT® was added directly to 5 µL of PCR products.
2. This mixture was subsequently incubated at 37°C for 15 minutes to activate ExoSAP, followed by a 15-minute incubation at 80°C to inactivate ExoSAP. The resulting treated PCR products are now prepared for further analysis in applications that necessitate the absence of excess primers and nucleotides.

7.3 Cycle Sequencing

Following the purification of the PCR product using ExoSAP-IT®, the resulting mixture was subjected to cycle sequencing employing the BigDye® Terminator v3.1 Cycle Sequencing Kit (California). A volume of 1µL of the purified PCR product was directly added to 9µL of the pre-prepared master mixture in each well of the PCR plate. Details regarding the preparation of the master mixture and the thermal cycling program utilized for the cycle sequencing reactions are provided in Table 3.7.1 and Table 3.7.2.

Table 3.7.1: Constituents of cycle sequencing reaction mixture

Reagent	Amount (µL)
Buffer (5x)	1.0
Terminator ready reaction mixes (Big dye)	0.5
Primer (Forward/Reverse) (5µM)	1.0
Nuclease free water	6.5
Template cDNA	1.0
Total	**10 µL**

Table 3.7.2: Thermal cycling steps used for cycle sequence reaction

Stage	Temperature	Duration
Initial denaturation	96℃	1 minute
3 steps cycling (25 cycle)		
Denaturation	96℃	10 seconds
Annealing	50℃	5 seconds
Extension	60℃	4 minutes
Holding Stage	4℃	∞

7.4 Purification of cycle sequencing reaction products

The products obtained from cycle sequencing were purified using the BigDye® XTerminator™ Purification Kit. This purification kit involves the addition of only two reagents, which can be added sequentially or pre-mixed.

- XTerminatorTM Solution
- SAMTM Solution

The purification steps follows:

1. The reaction products from the cycle sequencing were subjected to a brief centrifugation, followed by the addition of 45 µL of SAMTM solution into each well.
2. Subsequent to the addition of SAMTM solution, 10 µL of XTerminatorTM solution was introduced using a wide bore pipette tip.
3. The plate was then sealed with clear adhesive ABI film.
4. After sealing, the plate was vortexed at 2000 rpm for a duration of 30 minutes. Following the vortexing step, the plate underwent centrifugation at 200 x g for 3 minutes.
5. Finally, the plate was transferred to the ABI PRISM® 3500xl Genetic Analyzer (Figure 3.7.2).

8. Nipah virus screening of CoV positive bat samples using RT-PCR (Reverse Transcription-Polymerase chain reaction)

The nipah virus is a negative sense RNA virus which can be detected using qRT-PCR molecular technique. In this research, to screen the samples from nipah virus, Real-time RT PCR was performed using Applied Biosystems™ AgPath-ID™ One-Step RT-PCR kit.

8.1 Enzymes

AgPath-ID™ One-Step RT-PCR Reagents are designed for sensitive, robust amplification of RNA targets using a single-tube TaqMan™-based real-time RT-PCR strategy. It contains ArrayScript™ Reverse Transcriptase and AmpliTaq Gold™ DNA Polymerase.

8.2 Buffers and Reagents

- 2X RT-PCR Buffer
- 25X RT-PCR non-frost-free freezer Enzyme Mix
- Nuclease-free water

8.3 RT-PCR Assay

In one-step real-time RT-PCR, the reverse transcription (RT) phase is combined with the polymerase chain reaction (PCR). This method involves converting RNA into complementary DNA (cDNA) and amplifying it all in a single reaction. Reverse transcription PCR enables the use of RNA as a template to produce cDNA. The reverse transcriptase enzyme creates a single-stranded cDNA copy, which can then be amplified by DNA polymerase to produce double-stranded cDNA, facilitating the standard PCR amplification process.

8.4 Primer sets for targeted amplification in RT-PCR

Nipah N gene-specific primer-probe is used for the detection of Nipah virus (NiV)

Table 3.8.1: RT-PCR Primer set for the amplification of N gene

Primer	Strand	Sequence(5′-3′)	Amplicon Size (bp)
NiV-F1	Forward (Plus)	CTGGTCTCTGCAGTTATCACCATCGA	200 bp
NiV-R1	Reverse (Minus)	ACGTAYTTAGCCCATCTTCTAGTTTCA	
Niv-probe	NVBN654P	5′-/FAM/-AC AGC TCC M/ZEN/G ACA CTG CCG AGG A-/IBFQ/-3′	

8.5 RT-PCR Mastermix preparation and Steps

Each RT-PCR master mix (AgPath-ID™ One-Step RT-PCR Reagents) should be prepared on ice according to Table 3.8.2. The appropriate volume of RT-PCR master mix should then be transferred to a PCR plate or tubes. Samples or controls should be added to the wells or tubes containing the RT-PCR master mix, ensuring a final volume of 25 µL per reaction.

Table 3.8.2: Constituents of RT-PCR master mixture

Reagent	Amount (µL)
2X RT-PCR Buffer	12.5
TaqMan™ probes (10 µM)	0.5
Forward Primer (20µM)	1.0
Reverse Primer (20µM)	1.0
Nuclease free water	4
25X RT-PCR Enzyme Mix	1.0
RNA Template	5
Total	**25 µL**

8.6 PCR Thermal cycle protocol

Table 3.8.3: Thermal cycling steps used for RT-PCR

Step	Stage	Cycles	Temp	Time
Reverse transcription	1	1	45°C	10 minutes
RT inactivation/initial denaturation	2	1	95°C	10 minutes
Amplification	3	40	95°C 60°C	15 seconds 1 min

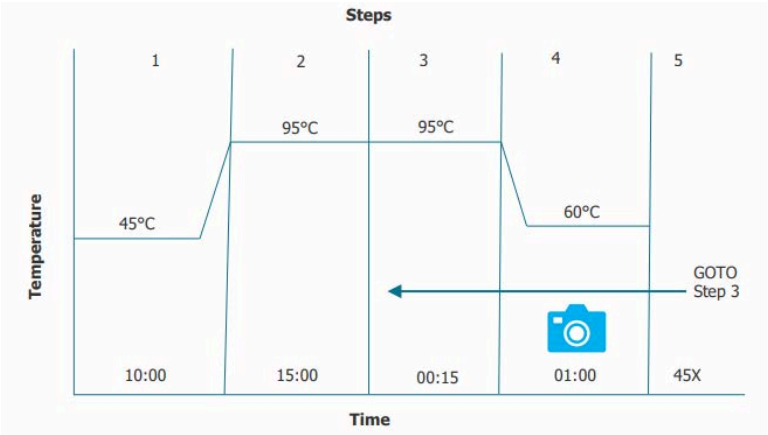


Fig. 3.8.1: Thermal cycling conditions for the detection of Nipah virus by RT-qPCR assay

8.7 RT-PCR Result Analysis

For Ct values ≤37, the samples are considered POSITIVE. As for control, negative control Ct value should be zero; positive control Ct value should be <37. For RNP (Internal control), CT value should be less than 30

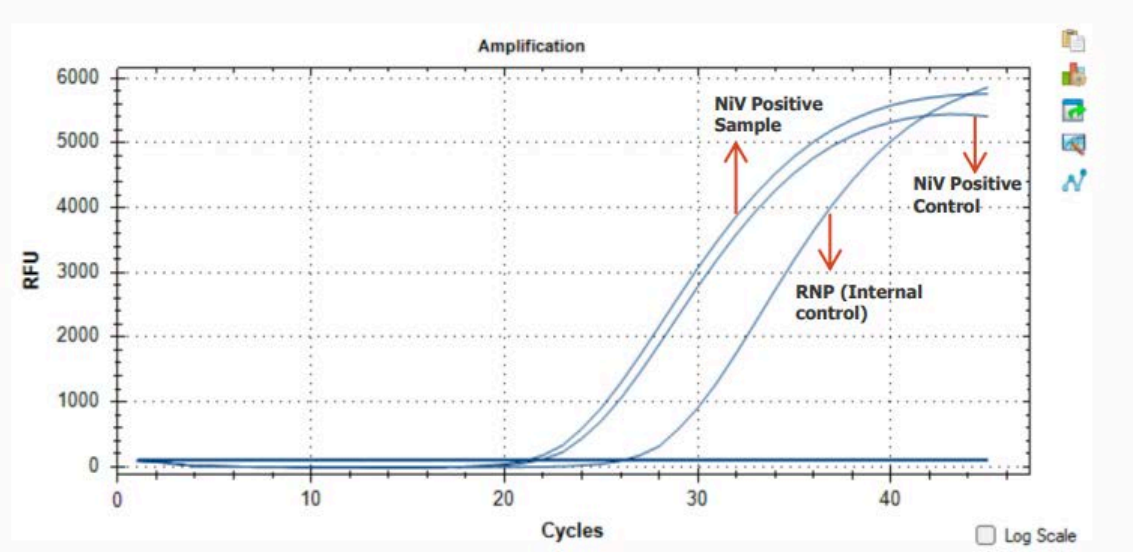


Fig. 3.8.2: Graphical depiction of a Nipah PCR detection

9. Analysis of Nucleotide Sequence

The chromatogram sequences were analyzed using Chromas 2.6.6 (Technelysium, Australia), and contigs were generated with SeqMan II (DNASTAR, Madison, Wis.). For multiple sequence alignment, ClustalW in Bioedit version 7.2.5 **(Hall, Biosciences et al., 2011)** was utilized and manually refined. Phylogenetic and molecular evolutionary assessments were carried out using the MEGA 11 software package **(Tamura, Stecher et al., 2021)**. The dendrograms will be created based on the maximum likelihood method **(Ranneby, B., 1984)**. In addition to that, bat species likelihood tree and bat coronavirus from Nipah virus host bats tree were constructed to represent and visualize the tanglegram. Co-phylogenetic analysis was performed from the nucleotide sequences.

9.1 Nucleotide Sequence Retrieval

In order to construct phylogenetic relationships among other bat coronaviruses in the alpha and beta genera of coronavirus, RdRp sequences of 26 existing bat coronaviruses from the NCBI database were retrieved. Two of the coronavirus strains were identified as human coronavirus in the database but yet were found in Bat samples. Additionally, bat coronavirus strains in Pteropus, Rousettus and Megaderma bats were retrieved from the Genbank in order to identify co-phylogeny in case of both coronavirus and nipah virus. Nipah virus sequence strains from the *Paramyxoviridae* family with similar bat species as host were also retrieved from the database for tanglegram construction.

9.2 Sequence similarity and distance analysis

BLAST (Basic Local Alignment Search Tool) application from the National Center for Biotechnology Information website (available at http://www.ncbi.nlm.gov/BLAST/), was utilized to submit all query sequences in FASTA files to determine their similarity and distance matrix. Sequence similarity was searched using the "Queryanchored with dots for identities" alignment view approach. Sequences were compiled and edited using the Laser gene sequence analysis software package (DNASTAR Inc, Madison, USA)

9.3 Reference sequences retrieval

The integration of sequencing data with genetic and functional information depends on reference sequences. These sequences are created to provide benchmarks for various applications, including the identification of sequence variations in medical datasets and genome annotation. Nucleotide-based reference sequences were obtained from the GenBank database of the National Center for Biotechnology Information (NCBI), part of the National Institutes of Health in Bethesda, MD (http://www.ncbi.nlm.nih.gov/). The NCBI Reference Sequence (RefSeq) database comprises a diverse, non-redundant, and well-annotated collection of sequences that represent naturally occurring DNA, RNA, and protein molecules. This collection includes sequences from plasmids, organelles, viruses, archaea, bacteria, and eukaryotes. Each reference sequence is generated from complete sequence data provided to the International Nucleotide

Sequence Database Collaboration (INSDC). Utilizing these reference sequences from the NCBI GenBank database, Multiple Sequence Alignment and Phylogenetic trees were generated.

9.4 Computational analysis of the retrieved RdRp genes

The mutation in the RdRp gene was examined using MEGA 11 software **(Tamura et al., 2021)**. SNPs were detected by comparing coronavirus genomes and identifying the most significant matches **(Kumar et al., n.d.)**. To compare the SNPs with the Reference Sequence, BioEdit software version 7.2.5 **(Hall et al., 2011)** was employed.

9.5 Co-phylogenetic Analysis to establish relationships between bat species and various Bat-CoV strains

Co-phylogenetic analysis covers a broad spectrum of host-virus dynamics. Cross species transmission and host switching are major potential events responsible for viral recombination resulting in RdRp mutation. Two phylogenetic trees using MEGA 11 software package **(Tamura, Stecher et al., 2021)** were constructed using the bat cytochrome b gene and bat coronavirus RdRp gene respectively. Tanglegram was created using (https://colab.google/) to analyze the evolutionary relationship between bats as host and coronavirus as the parasite. 18 bat species and 31 bat coronavirus were analyzed to construct the tanglegram.

Chapter – 4

Results

1. Molecular detection and sequence analysis of bat coronaviruses

1.1 Molecular detection of nipah virus in the bat samples using RT-PCR (Reverse transcriptase-Polymerase chain reaction)

In order to rule out viral co-infection and as the samples were collected from the nipah virus prone areas of Bangladesh, RT-PCR was done using Bio Rad CFX Opus Real-Time PCR machine for detection of nipah virus. All 1,668 samples from Dhaka, Cox's Bazar, Faridpur, Sylhet, and Rajshahi tested negative for nipah virus (Paramyxoviridae family).

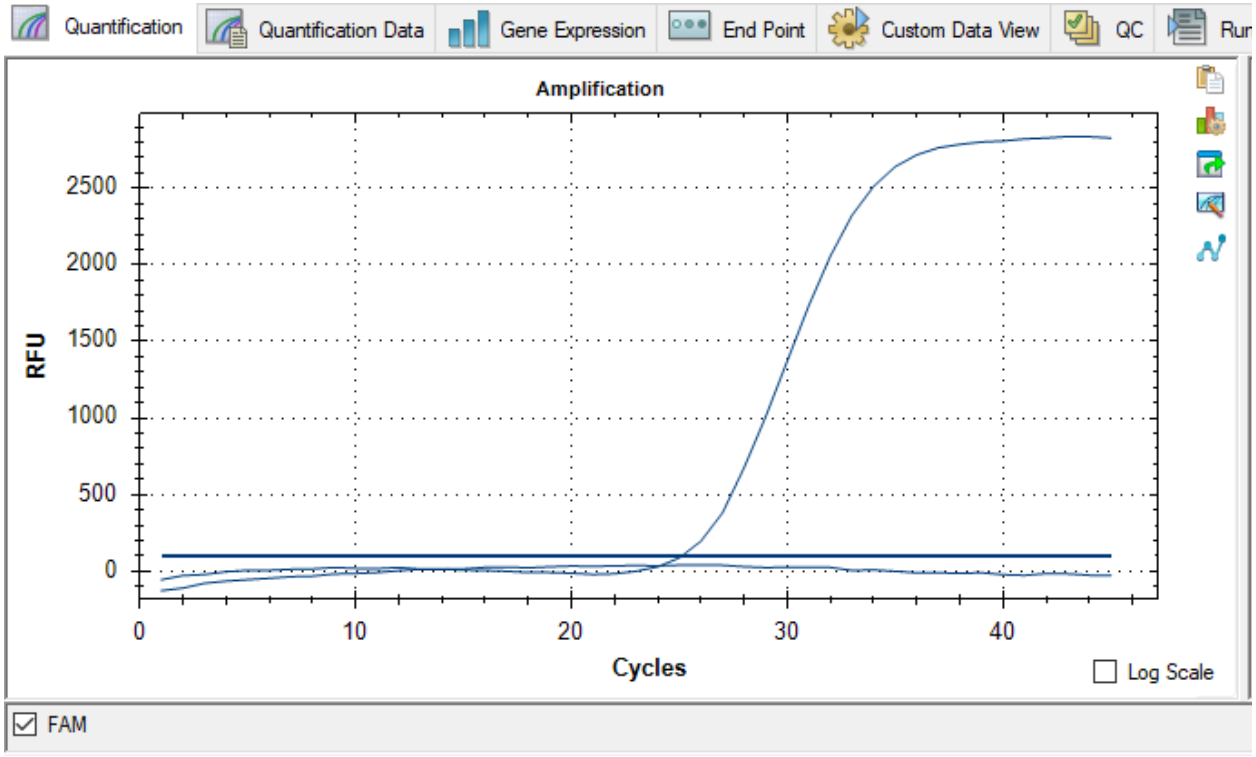
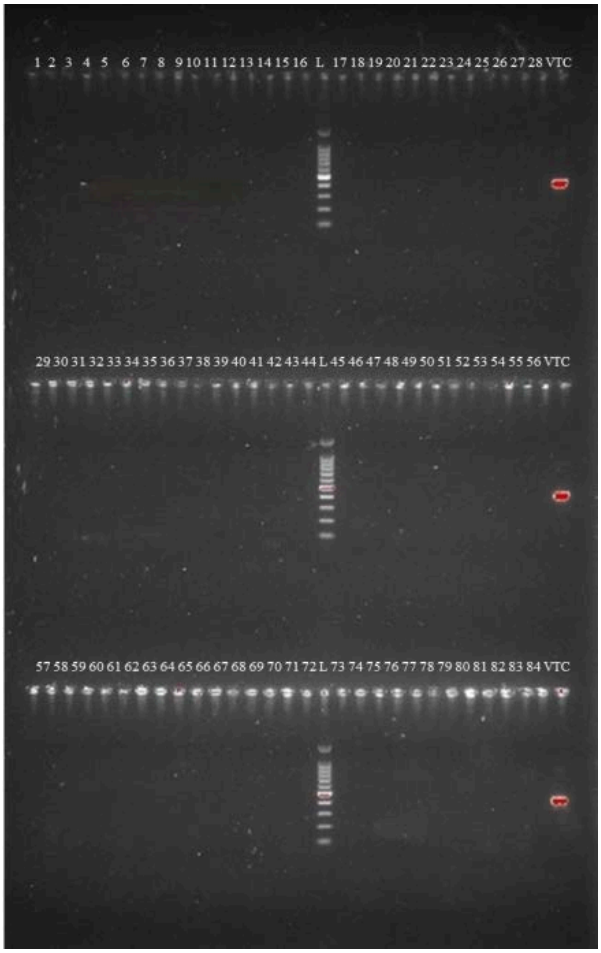


Fig. 4.1.1: Detection of Nipah-virus using RT-PCR (negative results)

1.2 Molecular detection of bat coronavirus by Conventional semi nested PCR

This research aimed to analyze 1,625 samples from two distinct bat species to identify viruses from the Coronaviridae family. The method utilized for detection was a semi-nested polymerase chain reaction (PCR), after which gel electrophoresis was done. Out of all samples, only 16

tested positive, as indicated by the presence of specific ~434 base pair (bp) amplicons (following Protocol-004) and ~328 bp amplicons (according to Protocol-003). These amplicons were validated through gel electrophoresis, as illustrated in the accompanying figure. Following this, the PCR products from the positive samples underwent Sanger sequencing based on PCR.



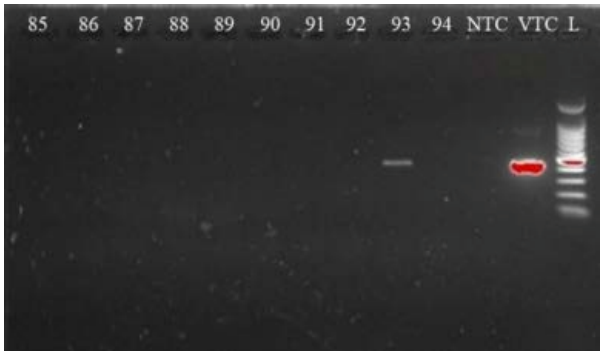


Fig. 4.1.2: The Bat-CoV RdRp gene was visualized using a 1.5% agarose gel stained with SYBR Safe DNA Gel Stain during electrophoresis (specific bands visualized). In well 93, a segment of the CoV RdRp (~434bp) from coronavirus positive samples can be seen. Well 97 carries the DNA marker (100bp Ladder) whereas well 96 includes the positive control (VTC), and well 95 contains the negative control (NTC).

1.3 Nucleotide sequencing

After sequencing and purifying the targeted RdRp regions using Sanger sequencing (ABI 3500xL automatic genetic analyzer), 10 out of 16 positive samples yielded recoverable sequences (434 & 328 bp amplicons). The analysis of electropherograms was conducted using Chromas software. The peaks displayed in four distinct colors represented the four nucleotides:

green for A (Adenine), red for T (Thymine), black for G (Guanine), and blue for C (Cytosine). Using Chromas v2.6, the sequence data were manually cross-verified, addressing any gaps or insertions between nucleotides as well as any unreadable sections at the beginning and end of the sequences.

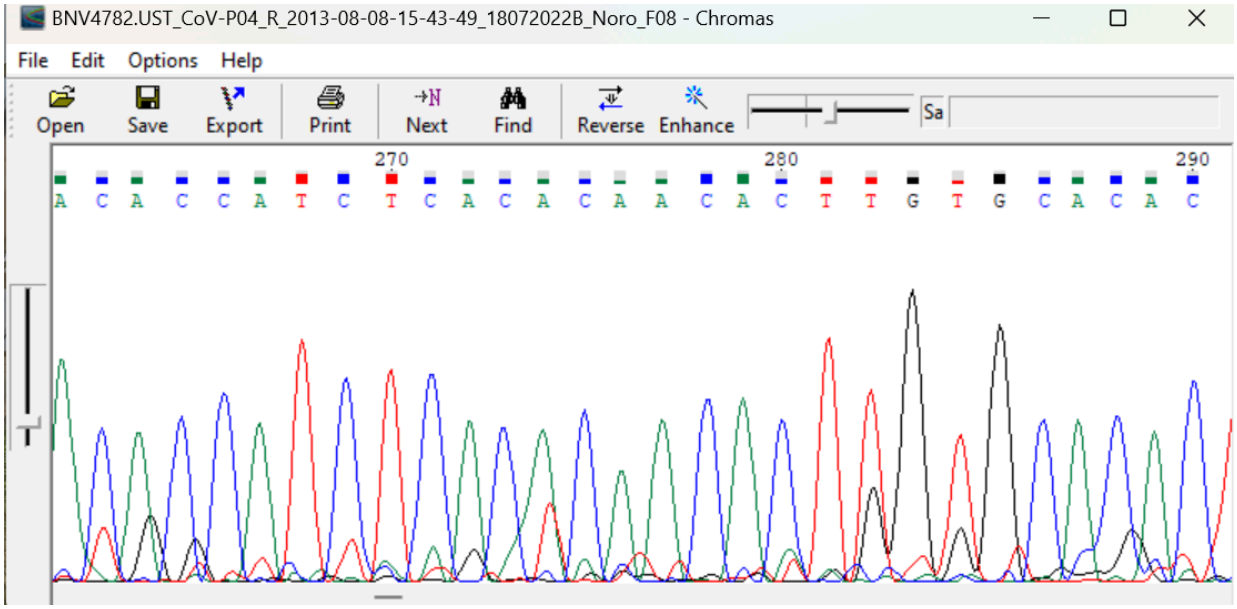
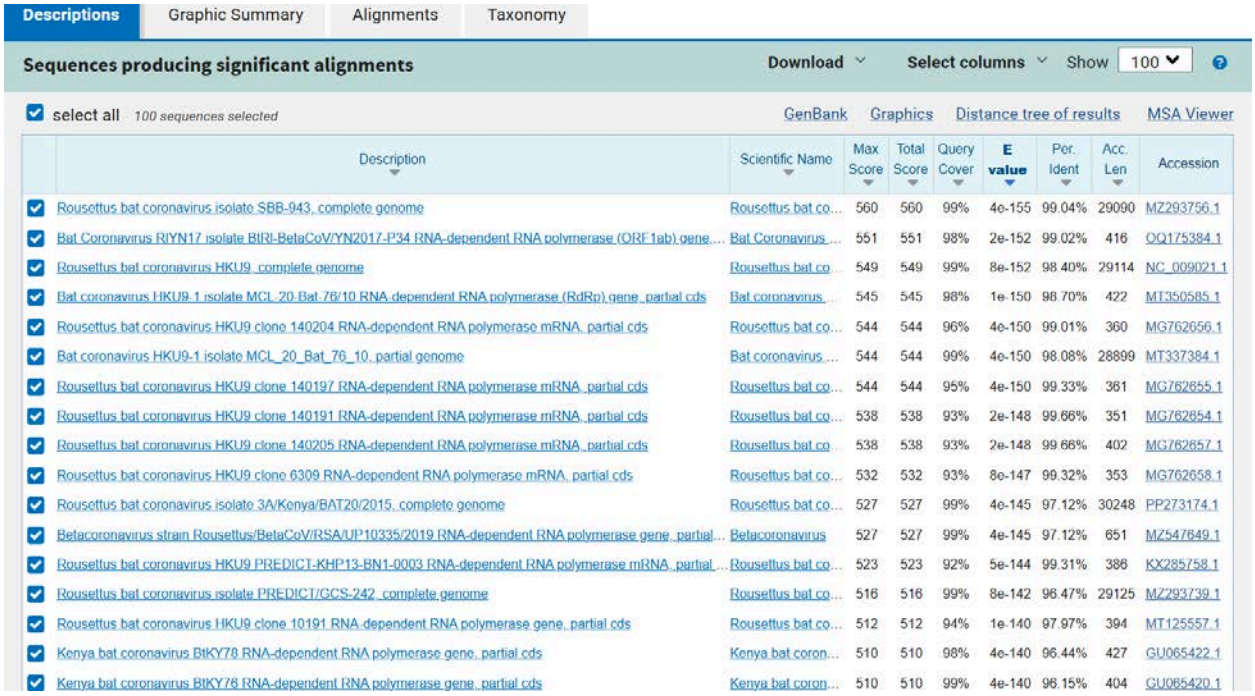


Fig. 4.1.3: Partial nucleotide sequence of a positive sample after sanger sequencing against the RdRp conserved region of ORF1b using Chromas

1.4 Homology report using NCBI Blast

Sequence homology searches were conducted using the BLAST (Basic Local Alignment Search Tool) server from the National Center for Biotechnology Information (NCBI, National Institutes of Health, Bethesda, MD) on the GenBank database. A nblast search was performed to identify highly similar sequences, revealing that 8 out of 10 sequences had about 98% or more similarity with RNA-dependent RNA polymerase (RdRp) genes from the database. This finding implies that the evolutionary pathways of coronaviruses in Bangladesh might suggest spillover events and transmissions since similar viral sequences were found in the database but in other bat species. The other two sequences showed no resemblance with RdRp genes from coronaviruses in the database. This might lead to the possibility of extensive mutations in the RdRp regions of

these bat coronaviruses resulting in novel viral sequencing. Furthermore, the BLAST search utilizing somewhat similar sequences (blastn) indicated that the RdRp gene sequences exhibited varying similarities with different coronavirus genera, which may help in identifying the specific virus species.



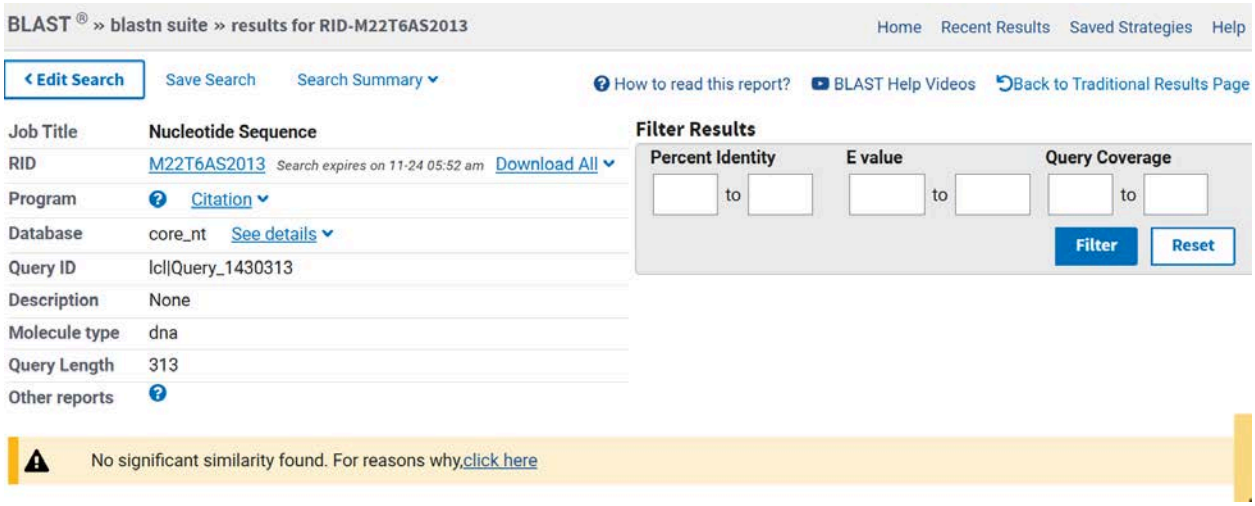
Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments — Download | Select columns | Show 100

select all *100 sequences selected* — GenBank | Graphics | Distance tree of results | MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Rousettus bat coronavirus isolate SBB-943, complete genome	Rousettus bat co...	560	560	99%	4e-155	99.04%	29090	MZ293756.1
Bat Coronavirus RIYN17 isolate BIRI-BetaCoV/YN2017-P34 RNA-dependent RNA polymerase (ORF1ab) gene,...	Bat Coronavirus ...	551	551	98%	2e-152	99.02%	416	OQ175384.1
Rousettus bat coronavirus HKU9, complete genome	Rousettus bat co...	549	549	99%	8e-152	98.40%	29114	NC_009021.1
Bat coronavirus HKU9-1 isolate MCL_20-Bat-76/10 RNA-dependent RNA polymerase (RdRp) gene, partial cds	Bat coronavirus ...	545	545	98%	1e-150	98.70%	422	MT350585.1
Rousettus bat coronavirus HKU9 clone 140204 RNA-dependent RNA polymerase mRNA, partial cds	Rousettus bat co...	544	544	96%	4e-150	99.01%	360	MG762656.1
Bat coronavirus HKU9-1 isolate MCL_20_Bat_76_10, partial genome	Bat coronavirus ...	544	544	99%	4e-150	98.08%	28899	MT337384.1
Rousettus bat coronavirus HKU9 clone 140197 RNA-dependent RNA polymerase mRNA, partial cds	Rousettus bat co...	544	544	95%	4e-150	99.33%	361	MG762655.1
Rousettus bat coronavirus HKU9 clone 140191 RNA-dependent RNA polymerase mRNA, partial cds	Rousettus bat co...	538	538	93%	2e-148	99.66%	351	MG762654.1
Rousettus bat coronavirus HKU9 clone 140205 RNA-dependent RNA polymerase mRNA, partial cds	Rousettus bat co...	538	538	93%	2e-148	99.66%	402	MG762657.1
Rousettus bat coronavirus HKU9 clone 6309 RNA-dependent RNA polymerase mRNA, partial cds	Rousettus bat co...	532	532	93%	8e-147	99.32%	353	MG762658.1
Rousettus bat coronavirus isolate 3A/Kenya/BAT20/2015, complete genome	Rousettus bat co...	527	527	99%	4e-145	97.12%	30248	PP273174.1
Betacoronavirus strain Rousettus/BetaCoV/RSA/UP10335/2019 RNA-dependent RNA polymerase gene, partial...	Betacoronavirus	527	527	99%	4e-145	97.12%	651	MZ547649.1
Rousettus bat coronavirus HKU9 PREDICT-KHP13-BN1-0003 RNA-dependent RNA polymerase mRNA, partial ...	Rousettus bat co...	523	523	92%	5e-144	99.31%	386	KX285758.1
Rousettus bat coronavirus isolate PREDICT/GCS-242, complete genome	Rousettus bat co...	516	516	99%	8e-142	96.47%	29125	MZ293739.1
Rousettus bat coronavirus HKU9 clone 10191 RNA-dependent RNA polymerase gene, partial cds	Rousettus bat co...	512	512	94%	1e-140	97.97%	394	MT125557.1
Kenya bat coronavirus BtKY78 RNA-dependent RNA polymerase gene, partial cds	Kenya bat coron...	510	510	98%	4e-140	96.44%	427	GU065422.1
Kenya bat coronavirus BtKY76 RNA-dependent RNA polymerase gene, partial cds	Kenya bat coron...	510	510	99%	4e-140	96.15%	404	GU065420.1

Fig. 4.1.4: BLAST search showing homological comparison among similar Bat-CoV RdRp gene sequences (for 8 sequences out of 10)



BLAST® » blastn suite » results for RID-M22T6AS2013 — Home | Recent Results | Saved Strategies | Help

‹ Edit Search | Save Search | Search Summary | How to read this report? | BLAST Help Videos | Back to Traditional Results Page

Job Title	Nucleotide Sequence
RID	M22T6AS2013 *Search expires on 11-24 05:52 am* Download All
Program	Citation
Database	core_nt See details
Query ID	lcl|Query_1430313
Description	None
Molecule type	dna
Query Length	313
Other reports	

Filter Results

Percent Identity [ ] to [ ] | E value [ ] to [ ] | Query Coverage [ ] to [ ]

Filter | Reset

⚠ No significant similarity found. For reasons why, click here

Fig. 4.1.5: BLAST search showing no resemblance found for 2 sample Bat-CoV RdRp gene sequences in the database

Table 4.1.1: Sequence homology search through NCBI BLAST-Tool for all the sequences

Serial	Sequence ID	Highly Similar Sequences	Identity	Coverage	Similar sequence accession number
1	01-Bat-CoV	No Significant Similarity			
2	02-Bat-CoV	Coronavirus PREDICT CoV-17 isolate MCL-19-BAT-665/2 RNA-dependent RNA polymerase (RdRp) gene	98.70%	98%	MT350588.1
3	03-Bat-CoV	Coronavirus PREDICT_CoV-17 clone 9154 RNA-dependent RNA polymerase gene, partial cds	100.00%	92%	MT064842.1
4	04-Bat-CoV	No Significant Similarity			
5	05-Bat-CoV	Rousettus bat coronavirus isolate SBB-943, complete genome	99.04%	99%	MZ293756.1
6	06-Bat-CoV	Coronavirus PREDICT CoV-17 isolate MCL-19-BAT-665/2 RNA-dependent RNA polymerase (RdRp) gene, partial cds	100.00%	100%	MT350588.1
7	07-Bat-CoV	Coronavirus PREDICT CoV-17 isolate MCL-19-BAT-665/2 RNA-dependent RNA polymerase (RdRp) gene, partial cds	98.05%	98%	MT350588.1
8	08-Bat-CoV	Coronavirus PREDICT CoV-17 isolate MCL-19-Bat-748/2 RNA-dependent RNA polymerase (RdRp) gene, partial cds	98.05%	98%	MT350591.1
9	09-Bat-CoV	Coronavirus PREDICT_CoV-17 clone 4052 RNA-dependent RNA polymerase gene, partial cds	99.67%	97%	MT221706.1
10	10-Bat-CoV	Coronavirus PREDICT CoV-17 isolate MCL-19-Bat-746/2 RNA-dependent RNA polymerase (RdRp) gene, partial cds	98.72%	100%	MT350587.1

1.5 Identification of Bat Coronavirus in different Bat species

After homology analysis using BLAST search was done to the 10 sequences analyzed through Sanger Sequencing, the majority of the viral sequences were matched to Bat-CoV of different species. Therefore further analysis was done to identify the Bat-CoV and find potential transmission across species. The majority of the positive samples were sourced from the Urogenital Swab. Analysis using the Coronavirus typing tool (https://www.genomedetective.com/app/typingtool/cov/) revealed that seven of the positive samples were similar to the Rousettus bat coronavirus HKU9. This indicates that Bat-CoV HKU9 can be found in both Rousettus and Pteropus bats along with small fruit bats suggesting chances of transmission and viral coinfection. One sample showed similarities with a Bat coronavirus, while the remaining two did not match any known coronavirus sequences, suggesting the potential emergence of a novel coronavirus.

Table 4.1.2: Identification of Bat Coronaviruses by Coronavirus Typing Tool

Serial	Sample ID	Location	Species	Sample Type	Coronavirus Typing Tool
1	01-Bat-CoV	Sylhet	Bat	Throat swab	Possible Corona (Not Identified)
2	02-Bat-CoV	Dhaka	*Pteropus Medius*	Oropharyngeal Swab	*Rousettus* bat coronavirus HKU9
3	03-Bat-CoV	Dhaka	*Pteropus Medius*	Oropharyngeal Swab	Bat Coronavirus
4	04-Bat-CoV	Faridpur	*Pteropus Medius*	Oropharyngeal Swab	Possible Corona (Not Identified)
5	05-Bat-CoV	Sylhet	Small Fruit Bat	Pooled Roost Urine and Feces	*Rousettus* bat coronavirus HKU9

6	06-Bat-CoV	Rajshahi	*Pteropus Medius*	Urogenital Swab	*Rousettus* bat coronavirus HKU9
7	07-Bat-CoV	Cox's Bazar	*Pteropus Medius*	Urogenital Swab	*Rousettus* bat coronavirus HKU9
8	08-Bat-CoV	Cox's Bazar	*Pteropus Medius*	Urogenital Swab	*Rousettus* bat coronavirus HKU9
9	09-Bat-CoV	Faridpur	*Pteropus Medius*	Urogenital Swab	*Rousettus* bat coronavirus HKU9
10	10-Bat-CoV	Faridpur	*Pteropus Medius*	Urogenital Swab	*Rousettus* bat coronavirus HKU9

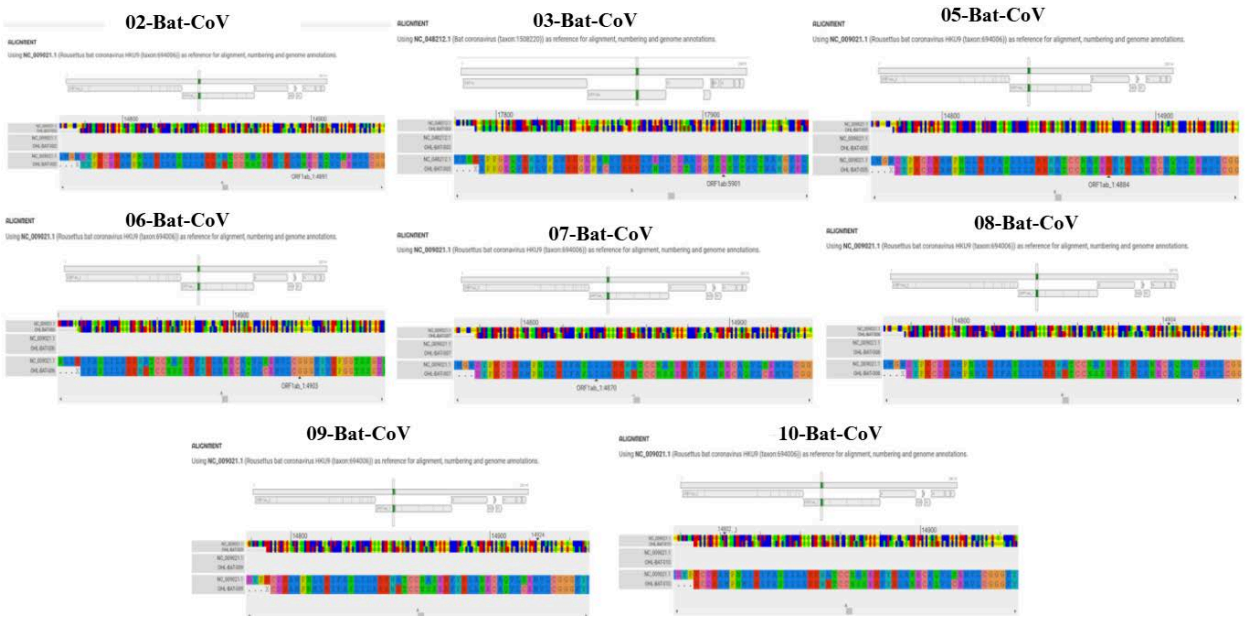


Fig. 4.1.6: The reference sequences of Bat Coronaviruses were identified using the Coronavirus Typing Tool. For the samples 02-Bat-Cov, 05-Bat-Cov, 06-Bat-Cov, 07-Bat-Cov, 08-Bat-Cov, 09-Bat-Cov and 10-Bat-Cov, the reference sequence was NC_009021.1. In contrast, the reference sequence for sample 03-Bat-Cov was NC_048212.1. Unfortunately, no reference sequences were available for samples 01-Bat-Cov and 04-Bat-Cov.

1.6 Nucleotide Sequence Alignment

Aligning all 36 sequences of RdRp gene in bat coronavirus from Bangladesh (nipah prone areas) and across the globe, we assess the similarities and differences of the highly conserved RdRp region. Some specific SNPs can also be observed through the alignment. The ClustalW alignment was done by Bioedit version 7.2.5 **(Hall, Biosciences et al., 2011)** .

Fig. 4.1.7: Nucleotide sequence alignment of RdRp region of existing Bat-CoVs with 10 positive sequences

1.7 Phylogenetic Analysis of the Coronaviruses with other reference strains based on RdRp Gene

Different bat coronavirus sequences and reference sequences were obtained from the NCBI database to compare retrieved sequences with bat coronavirus sequences across the globe. A phylogenetic cladogram of 36 strains targeting RdRp gene is constructed using the principle of maximum likelihood method with 1000 bootstrap replications. The strain 03-Bat-CoV showed similarities to bat coronaviruses found in Thailand and China, as indicated by the phylogenetic study (a) conducted under Protocol-003. It was clustered alongside beta bat coronaviruses, which have connections to human coronaviruses and may lead to human spillover events. In the

phylogenetic analysis of bat coronaviruses (b) based on the P-004 protocol, our sequences formed a cluster with beta bat coronaviruses. In contrast, samples 01-Bat-CoV and 04-Bat-CoV created a distinct cluster separate from the others which identifies them as unclassified Bat-CoV. Other strains were positioned apart from strain 05-Bat-CoV. 8 of the retrieved bat coronavirus from the nipah-prone areas of Bangladesh can also be characterized as beta coronavirus by clustering of distinct clades.

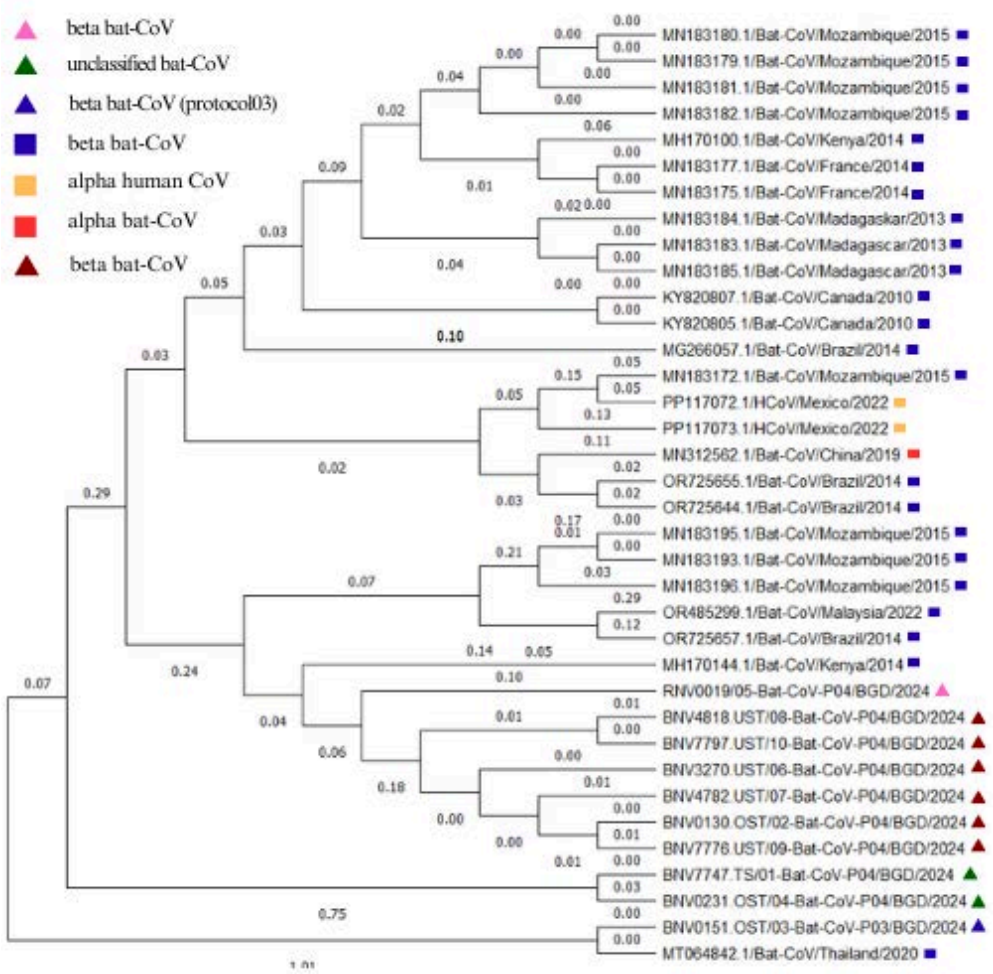


Fig. 4.1.8: Phylogenetic relationships between Bat-CoVs of nipah-prone areas with global CoV strains in the form of a cladogram. The coloured marks also distinguish the genus of Bat-CoVs. In the figure, the square represents sequences from the database whereas the triangle represents the data obtained from nipah-prone areas of Bangladesh

1.8 Similarity and Distance Matrix Analysis

The Similarity and Distance matrix analysis was conducted to determine which sequences were most similar to the reference sequence and which displayed greater divergence. It's important to mention that the sequences 01-Bat-CoV and 04-Bat-CoV were excluded from this analysis due to their inclusion of an unacceptable negative value. The strain 03-Bat-CoV (protocol-003) was assessed against a bat coronavirus reference sequence, revealing a similarity of 65% (refer to Table 4.1.4). Among the strains, 06-Bat-CoV demonstrated the greatest variation from the reference, while 05-Bat-CoV exhibited the least variation according to the distance matrix. Specifically, 05-Bat-CoV showed approximately 98.05% similarity with the reference sequence, reinforcing its identification as a Rousettus bat coronavirus, as suggested by the similarity matrix. On the other hand, 06-Bat-CoV had a similarity of 69.03% with the reference sequence, while the other strains ranged from 71% to 75%.

Table 4.1.3: Distance Matrix analysis of Bat coronaviruses (for P-003)

		NC_048212.1	**03-Bat-CoV**
Ref Seq	NC_048212.1		
Distance matrix	03-Bat-CoV	0.348	0.00

Table 4.1.4: Similarity Matrix analysis of Bat coronaviruses (for P-003)

		NC_048212.1	**03-Bat-CoV**
Ref Seq	NC_048212.1	100	100
Similarity Matrix	03-Bat-CoV	65.18	100

Table 4.1.5: Distance Matrix analysis of Bat coronaviruses at Nucleotide level (for P-004)

		NC_009021.1	02-Bat-CoV	05-Bat-CoV	06-Bat-CoV	07-Bat-CoV	08-Bat-CoV	09-Bat-CoV	10-Bat-CoV
Ref Seq	**NC_009021.1**								
Distance matrix	02-Bat-CoV	0.27							
	05-Bat-CoV	0.02	0.28						
	06-Bat-CoV	0.31	0.003	0.29					
	07-Bat-CoV	0.28	0.02	0.29	0.015				
	08-Bat-CoV	0.28	0.03	0.29	0.015	0.026			
	09-Bat-CoV	0.28	0.01	0.28	0.014	0.024	0.03		
	10-Bat-CoV	0.29	0.013	0.28	0.007	0.020	0.007	0.023	0.00

Table 4.1.6: Similarity Matrix analysis of Bat coronaviruses at Nucleotide level (for P-004)

		NC_009021.1	02-Bat-CoV	05-Bat-CoV	06-Bat-CoV	07-Bat-CoV	08-Bat-CoV	09-Bat-CoV	10-Bat-CoV
Ref Seq	**NC_009021.1**	100	100	100	100	100	100	100	100
Similarity Matrix	02-Bat-CoV	73.20	100	100	100	100	100	100	100
	05-Bat-CoV	98.05	71.73	100	100	100	100	100	100
	06-Bat-CoV	69.04	99.63	71.48	100	100	100	100	100
	07-Bat-CoV	72.27	98.38	72.12	98.52	100	100	100	100
	08-Bat-CoV	71.73	97.39	72.14	98.51	97.38	100	100	100
	09-Bat-CoV	71.81	98.9	72.4	98.5	97.6	96.98	100	10

			9	3	7	5			0
	10-Bat-CoV	71.32	98.66	72.03	99.29	97.99	99.34	97.70	100

1.9 Mutation of RdRp Gene

The reference sequence chosen for our mutation analysis was NC_009021.1/BetaCoV/China/Bat/2006. During the sequence analysis, strains **01-Bat-CoV** and **04-Bat-CoV** demonstrated the most significant levels of mutation, which is atypical for conserved regions. Conversely, the 05-Bat-CoV strain exhibited about a 2.23% change in nucleotides, while other strains showed mutations ranging from 22% to 25%. The 03-Bat-CoV sample revealed a 27.80% mutation rate in nucleotides compared to the reference sequence (NC_048212.1).

Table 4.1.7: Mutation Analysis of RdRp gene (both at the nucleotide and protein level)

After Comparing with Reference Sequence NC_009021.1 and NC_048212.1 (for 03-Bat-CoV)					
Sample ID	**Length (nt/aa)**	**Number of Mutated nt**	**Rate of mutation**	**Number of Mutated Amino Acids**	**Rate of mutation**
01-Bat-CoV	313	167	53.35%		
02-Bat-CoV	313/132	61	19.49%	21	15.90%
03-Bat-CoV	313/74	87	27.80%	13	17.56%
04-Bat-CoV	313	173	55.27%		
05-Bat-CoV	313/133	7	2.23%	2	1.5%
06-Bat-CoV	313/97	77	24.60%	21	21.64%
07-Bat-CoV	313/118	69	22.04%	28	23.72%
08-Bat-CoV	313/136	70	22.36%	26	19.12%
09-Bat-CoV	313/119	70	22.36%	19	15.97%
10-Bat-CoV	313/114	71	22.68%	19	16.67%

1.10 Co-Phylogenetic Analysis of the Coronaviruses with Bat species as Host

Since mutation was observed at RdRp region of these bat coronaviruses, further analysis was conducted to investigate the causation of mutations. Since the samples were collected from the nipah-prone areas in Bangladesh, potential transmission, spillover, viral coinfection and host switching of the virus could lead to substantial mutagenesis. Co-phylogenetic analysis of coronavirus was done with their respective hosts (bat species).

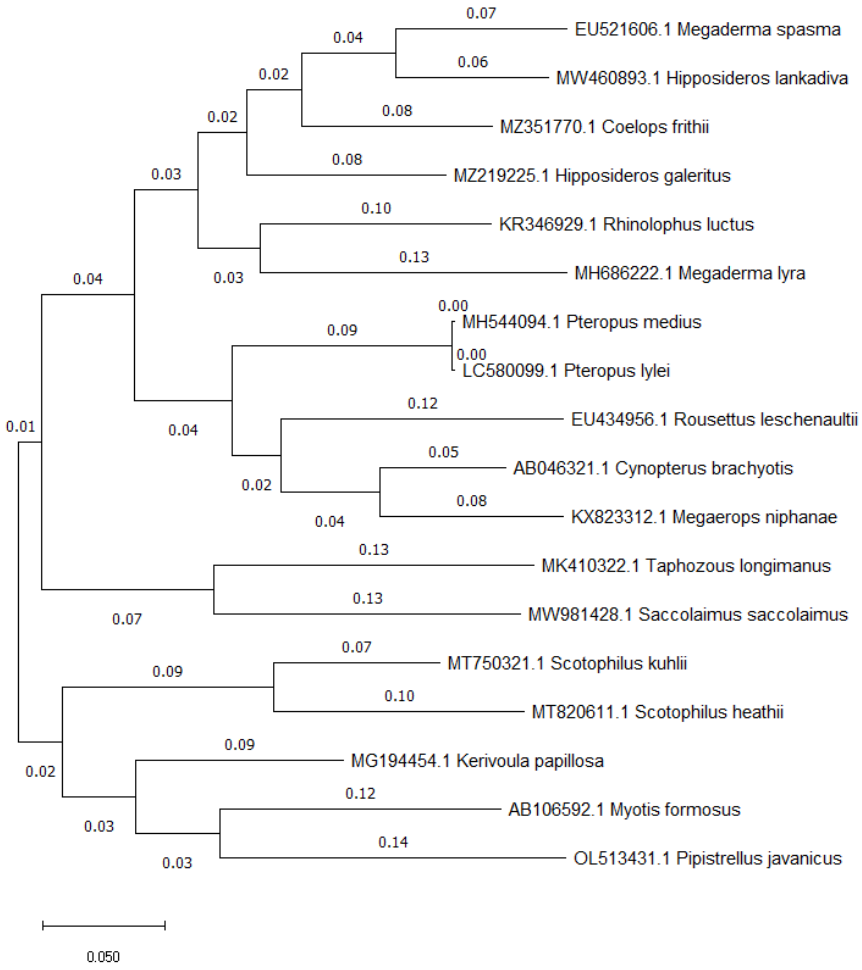


Fig. 4.1.9: Bat species tree (host tree)

The following maximum likelihood tree represents a cladogram of 18 species of bats that are found in Bangladesh. These bat species host both nipah and coronavirus which might lead to spillover events such as host switching, viral co-infections and altered pathogenicity. The tree was constructed using the cytochrome b sequence of the bat species. It has been observed that from the 18 various species, most positive Bat-CoV sequences from this study were found in *Pteropus medius* while other Bat-CoV sequences of Bangladesh retrieved from NCBI database were hosted by *Rousettus leschenaultia* and *Megaderma lyra*.

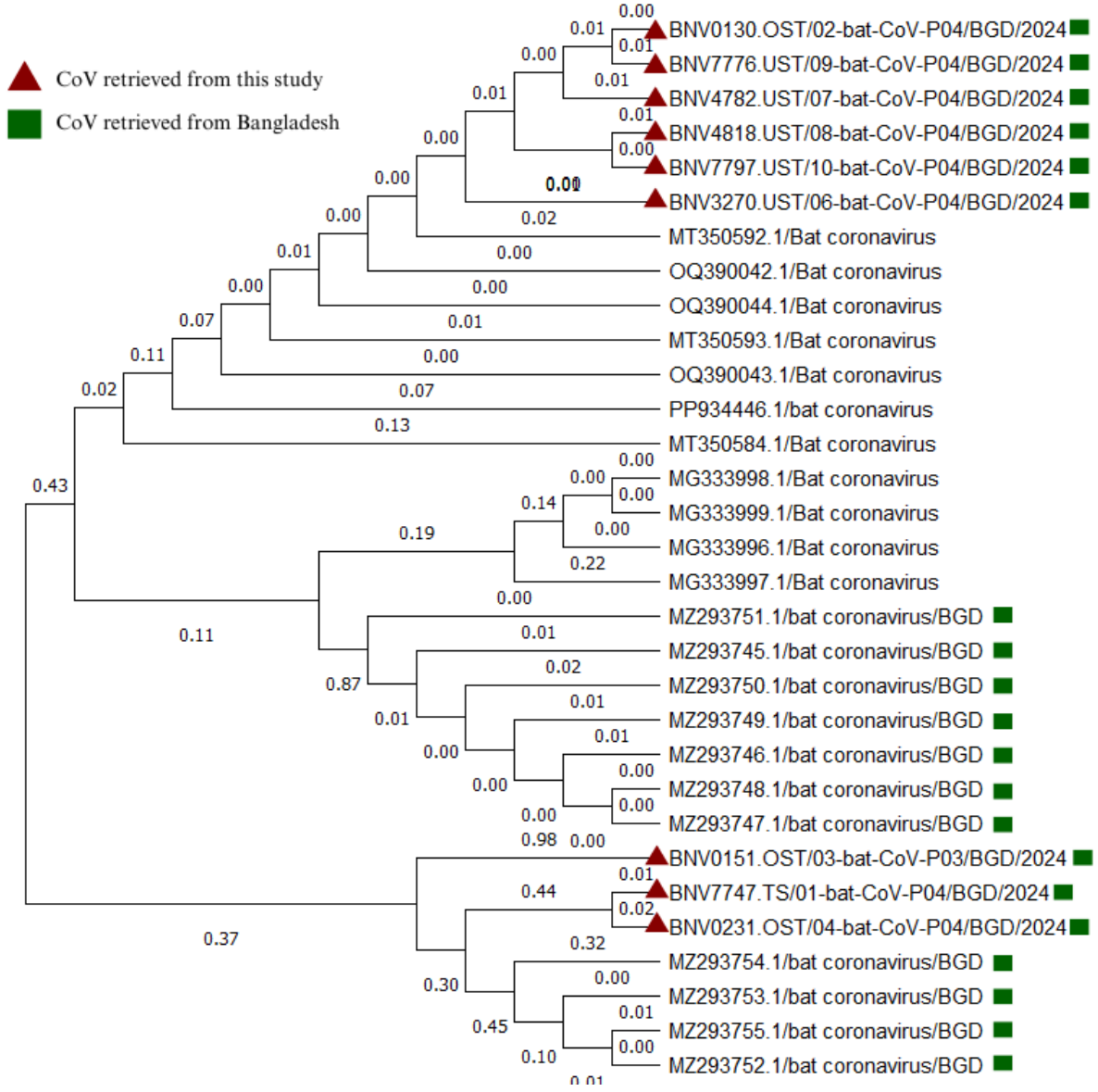


Fig. 4.1.10: Bat coronavirus tree (parasite tree)

This particular phylogeny shows the bat coronaviruses of the three major bat species discussed above along with the sequenced 10 bat coronaviruses from our study. Here, 11 previously sequenced bat coronaviruses were taken from the NCBI database in order to construct the cladogram. The rest of the Bat-CoV sequences from NCBI were retrieved from various locations of Asia (India, Sri Lanka, Thailand). It can be seen from the phylogeny that although 02-Bat-CoV, 06-Bat-CoV, 07-Bat-CoV, 08-Bat-CoV, 09-Bat-CoV and 10-Bat-CoV form clade with the global bat coronavirus sequences, 01-Bat-CoV, 04-Bat-CoV and 03-Bat-CoV show clade formation with the sequences identified within Bangladesh.

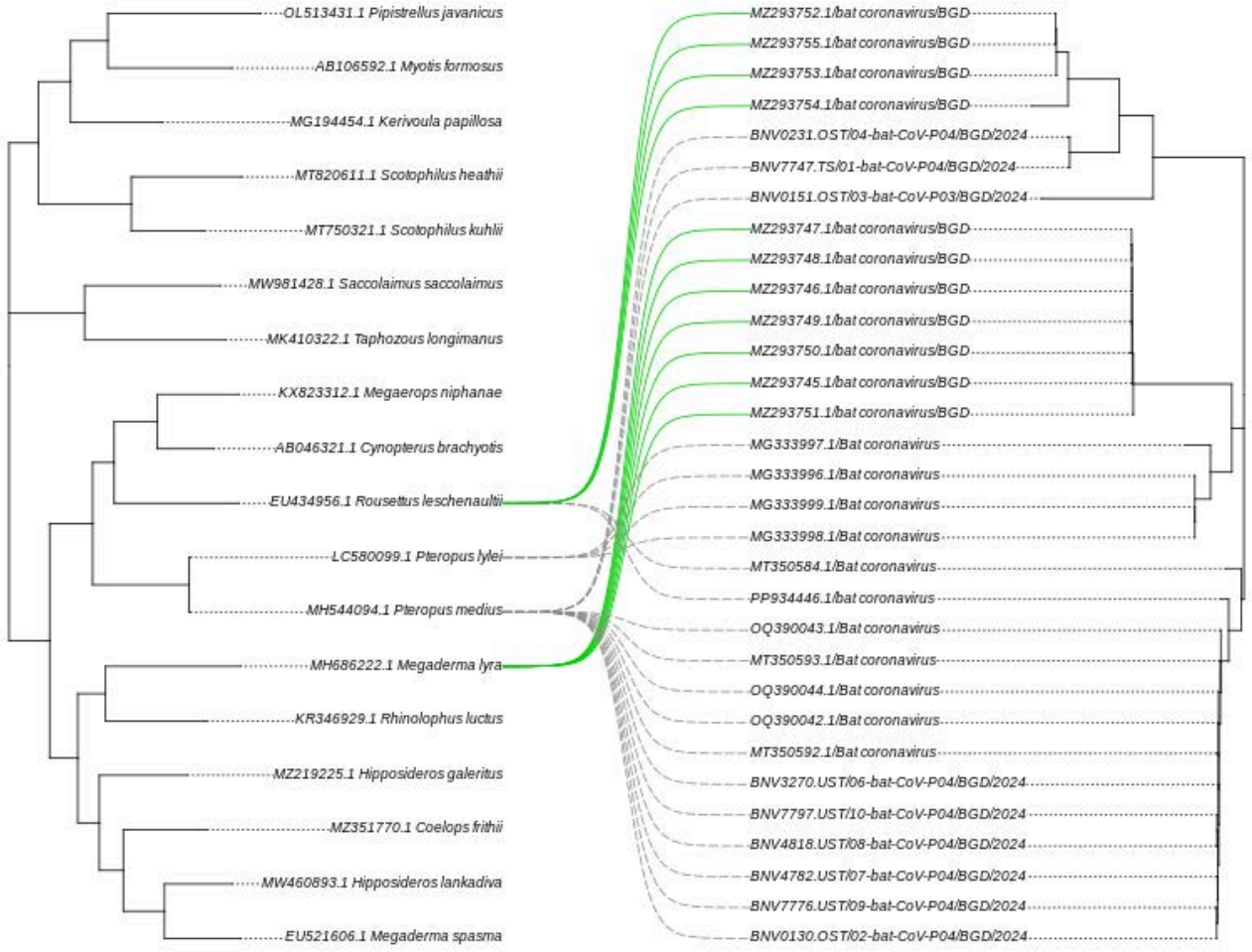


Fig. 4.1.11: Tanglegram (Co-phylogenetic relationship between host-bat and parasite-coronavirus)

The co-phylogenetic analysis was done using two different evolutionary trees, the host tree and the parasite tree. Here, in the tanglegram, out of 18 species of bats found in Bangladesh, four of the bat species (*Pteropus lylei*, *Pteropus medius*, *Megaderma lyra* and *Rousettus leschenaultia*) are seen to have dynamic host-parasite relationship from the current set of data. The green lines seem to represent the linear relationships of *Pteropus lylei* and *Megaderma lyra* whereas the grey dotted lines represent the more dynamic phylogenetic relationships of *Pteropus medius* and *Rousettus leschenaultia* bats.

Chapter – 5

Discussion

The purpose of this study was to determine if there were any impacts in the detection and genetic characterization of coronavirus in bat samples collected from Nipah-prone areas of Bangladesh. After the primary molecular analysis of the bat coronavirus through conventional PCR, it was immediately identified that one of the samples 03-Bat-CoV was a carrier of human strain coronavirus that supports the theory of ecological spillover, cross-species transmission. As coronavirus is a viral character with higher mutation frequency and transmission rate, events like cross-species transmission, the coexistence of multiple coronavirus strains, and host switching influence viral recombination and the emergence of novel mutations.

RT-PCR was performed to detect any presence of Nipah virus in the 1625 sample pool. The result of the RT-PCR was found negative, which indicates that although the bat samples were collected from the Nipah-prone areas of Bangladesh, they were not infected by Nipah virus and tested positive for *Coronaviridae* only. Phylogenetic analysis, similarity and distance matrix analysis, and mutational analysis of the RdRp gene, were performed to investigate any alteration in our sample pool. After obtaining sequences through the Sanger sequencing method, reference sequences for each strain were determined using the coronavirus typing tool. According to the CTT, the majority of strains exhibited similarities with the Rousettus bat coronavirus, while one sample (03-Bat-CoV) demonstrated a notable similarity specifically with the bat coronavirus. However, for the two strains (01-Bat-CoV and 04-Bat-CoV), the coronavirus typing tool could not identify any matching sequences in reference databases, indicating that these may represent novel coronaviruses. After identification of the sample Bat-CoVs, phylogenetic analysis was conducted to cluster them with similar bat coronavirus sequences revealing their genus and clades in the tree constructed using maximum likelihood method. The Cladogram of 36 bat coronavirus with major sub genera under alpha and beta coronavirus were analyzed only to conclude that 08 sequences out of 10 fell under the genus beta coronavirus whereas 01-Bat-CoV and 04-Bat-CoV were still unclassified due to mutations in the RdRp region.

This led us to the distance and similarity matrix using a reference sequence which showed The strain 03-Bat-CoV (protocol-003) was assessed against a bat coronavirus reference sequence, revealing a similarity of 65%. Among the strains, 06-Bat-CoV demonstrated the greatest variation from the reference, while 05-Bat-CoV exhibited the least variation according to the distance matrix. Specifically, 05-Bat-CoV showed approximately 98.05% similarity with the reference sequence, reinforcing its identification as a Rousettus bat coronavirus, as suggested by the similarity matrix. On the other hand, 06-Bat-CoV had a similarity of 69.03% with the reference sequence, while the other strains ranged from 71% to 75%.

The mutational analysis of RdRp revealed that the rate of mutation was highest in the 01-Bat-CoV and 04-Bat-CoV samples which again reasserts the theory of viral recombination leading to such mutations. In order to then establish a correlation between cross-species transmission and viral recombination leading to the mutations, co-phylogenetic analysis between the host species (bat) and coronavirus (parasite) was performed. In this host-parasite evolutionary relationship, different genera or variations of coronavirus strains can be seen in both *Pteropus medius* and *Rousettus leschenaultia* bats. Also identified Bat-CoV strains of Rousettus bats have been found in Pteropus bats which clearly indicates host switching and co-speciation of Bat-CoV strains in Pteropus and Rousettus bats. The coexistence of multiple Bat-CoV strains in a single host leads to viral recombination which can explain the RdRP mutation of our study samples, especially 01-Bat-Cov and 04-Bat-Cov. Since the 03-Bat-Cov resembles a human coronavirus strain found in bats, it confirms the hypotheses of cross-species transmission and spillover events which might lead to mutation and emergence of novel coronavirus strain.

Chapter – 6

Conclusion

In conclusion, studies of bat-origin coronaviruses like "Molecular Detection & Genetic Characterization of Coronavirus from Bat sample in Nipah Prone Areas of Bangladesh" play a crucial role in pandemic preparedness and response efforts. The findings of this research indicate that different types of coronaviruses, some of which could potentially lead to zoonotic diseases, exist in bats in Bangladesh. This study enhances our understanding of the potential risks associated with coronaviruses by utilizing phylogenetic analysis, sequencing of the RdRp gene, and examining mutations in the RdRp gene. Additionally, the findings demonstrated that the coronaviruses found in bats in Bangladesh formed a unique cluster, separate from other coronaviruses and can be affected by other viruses such as Nipah. By systematically monitoring bat populations for the presence of coronaviruses, we can detect emerging threats, identify high-risk areas and species for zoonotic transmission, and inform strategies for disease prevention and control. Understanding the genetic diversity of bat-origin coronaviruses is essential for predicting viral evolution, assessing the risk of spillover events, and developing effective countermeasures to mitigate the impact of emerging infectious diseases on human and animal health. Through collaborative research efforts, interdisciplinary partnerships, and sustained investment in surveillance and monitoring efforts, we can advance our understanding of bat-origin coronaviruses and enhance global health security.

1. Future Prospects

This study shows strong potential for metagenomic analysis of the susceptible bat species that show higher rates of cross-species transmission and viral recombination. In order to classify emergence of novel coronavirus, whole genome sequencing using NGS should be done. 01-Bat-CoV and 04-Bat-CoV samples should be sequenced using NGS molecular techniques and SNPs are to be detected. Secondary and Tertiary protein prediction from the mutant RdRp will allow us to study the changes in the replication complex and mechanism hence enhancing our knowledge of novel coronavirus strains. Lastly, more data collection and surveillance on the bat

population of Bangladesh should be done to determine the ecological factors of bat-borne virus transmissions.

Chapter – 7

Reference

Abdelraheem, E., Osman, M. M., Hospital, S. G., Arabia, S., Elgemaabi, O., Abdelhamid, A., & Alla, F. (2016). Computational Analysis of Deleterious Single Nucleotide Polymorphisms (SNPs) in Human MutS Homolog6 (MSH6) (Vol. 6, Issue July). https://doi.org/10.5923/j.bioinformatics.20160602.03

Anthony, S. J., Ojeda-Flores, R., Rico-Chávez, O., Navarrete-Macias, I., Zambrana-Torrelio, C. M., Rostal, M. K., ... & Goldstein, T. (2017). Coronaviruses in bats from Mexico. Journal of General Virology, 98(8), 1899-1906.

Brian, D. A., & Baric, R. S. (2005). Coronavirus genome structure and replication. Current topics in microbiology and immunology, 287, 1–30. https://doi.org/10.1007/3-540-26765-4_1

Chand, G. B., Banerjee, A., & Azad, G. K. (2020). Identification of novel mutations in RNA-dependent RNA polymerases of SARS-CoV-2 and their implications on its protein structure. PeerJ, 8, e9492. https://doi.org/10.7717/peerj.9492

DOI: 10.1016/j.cell.2020.02.058

Drexler, J. F., Corman, V. M., Müller, M. A., Maganga, G. D., Vallo, P., Binger, T., ... & Drosten, C. (2014). Bats host major mammalian paramyxoviruses. Nature Communications, 3, 796.

Dhama, K., Sharun, K., Tiwari, R., Sircar, S., Bhat, S., Malik, Y. S., Singh, K. P., Chaicumpa, W., Bonilla-Aldana, D. K., & Rodriguez-Morales, A. J. (2020). Coronavirus disease 2019–COVID-19. *Clinical Microbiology Reviews, 33*(4), e00028-20. https://doi.org/10.1128/CMR.00028-20

Du, Y., Wang, C., & Zhang, Y. (2022). Viral Coinfections. Viruses, 14(12), 2645. https://doi.org/10.3390/v14122645

Google Colab

Guillaume, V., Lefeuvre, A., Faure, C., Marianneau, P., Buckland, R., Lam, S. K., Wild, T. F., & Deubel, V. (2004). Specific detection of Nipah virus using real-time RT-PCR (TaqMan). Journal of virological methods, 120(2), 229–237. https://doi.org/10.1016/j.jviromet.2004.05.018

Hall, T., et al. (2011). "BioEdit: an important software for molecular biology." GERF Bull Biosci 2(1): 60-61

Halpin, K., Young, P. L., Field, H. E., & Mackenzie, J. S. (2011). Ecology of Hendra and Nipah viruses. Current Topics in Microbiology and Immunology, 359, 135-153. https://doi.org/10.1007/82_2011_190

https://github.com/Wasi0007/PhyloSphere_backend/blob/main/Tanglegram.r

Jenuth, J. P. (1999). "The NCBI: publicly available tools and resources on the web." Bioinformatics methods and protocols: 301-312.

Jones, K. E., et al. (2008). "Global trends in emerging infectious diseases." Nature 451(7181): 990-993.

Klein, S. L., & Yildirim, M. I. (2021). "SARS-CoV-2 Spike Protein and its Variants: Immunological Implications and Therapeutic Targets." Viral Therapy, 35(5), 233-241.

Kumar, R., Verma, H., Singhvi, N., Sood, U., Gupta, V., Singh, M., & Kumari, R. (n.d.). crossm Comparative Genomic Analysis of Rapidly Evolving SARS-CoV- 2 Reveals Mosaic Pattern of Phylogeographical Distribution. 8.

Larsen, B. B., et al. (2022). "Evolution and diversity of bat and rodent Paramyxoviruses from North America." Journal of virology 96(3): e01098-01021.

Lam, T. T., Jia, N., Zhang, Y. W., Shum, M. H., Jiang, J. F., Zhu, H. C., ... & Zhu, G. (2020). Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. Nature, 583(7815), 282-285.

Lee, B. (2012). Henipavirus: Ecology, molecular virology, and pathogenesis. In Advances in Virus Research (Vol. 80, pp. 85–115). Academic Press.

Li, H. (2018). "Minimap2: pairwise alignment for nucleotide sequences." Bioinformatics 34(18): 3094-3100.

Li, W., Wong, S. K., Li, F., Kuhn, J. H., Huang, I. C., Choe, H., & Farzan, M. (2005). Animal origins of the severe acute respiratory syndrome coronavirus: Insight from ACE2-S-protein interactions. Journal of Virology, 80(9), 4211-4219.

López-Blanco, J. R., et al. (2014). "iMODS: internal coordinates normal mode analysis server." Nucleic acids research 42(W1): W271-W276.

Lu, G., Wang, Q., & Gao, G. F. (2015). Bat-to-human: Spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond. Trends in Microbiology, 23(8), 468-478.

Luby, S. P., Gurley, E. S., & Hossain, M. J. (2006). Transmission of human infection with Nipah virus. Clinical Infectious Diseases, 49(11), 1743-1748. https://doi.org/10.1086/647952

Malik Y. A. (2020). Properties of Coronavirus and SARS-CoV-2. The Malaysian journal of pathology, 42(1), 3–11.

Más, R. (2013). [Details regarding the specific paper or context are unclear; please clarify the exact citation or subject for a precise reference].

McKee, C. D., Islam, A., Luby, S. P., Salje, H., Hudson, P. J., Plowright, R. K., & Gurley, E. S. (2021). The ecology of Nipah virus in Bangladesh: A nexus of land-use change and opportunistic feeding behavior in bats. Viruses, 13(2), 169. https://doi.org/10.3390/v13020169

Memish, Z. A., Perlman, S., & Van Kerkhove, M. D. (2020). Middle East respiratory syndrome. The Lancet, 395(10229), 1063-1077.

O'Shea, T. J., Cryan, P. M., Cunningham, A. A., Fooks, A. R., Hayman, D. T., Luis, A. D., ... & Wood, J. L. (2014). Bat flight and zoonotic viruses. Science, 344(6185), 1234-1235. https://doi.org/10.1126/science.1257890

Plowright, R. K., et al. (2015). Ecological dynamics of emerging bat virus spillover. Proceedings of the Royal Society B: Biological Sciences, 282(1798), 20142124. https://doi.org/10.1098/rspb.2014.2124

Rahman, M. Z., Islam, M. M., Hossain, M. E., Rahman, M. M., Islam, A., Siddika, A., Hossain, M. S. S., Sultana, S., Islam, A., Rahman, M., Rahman, M., Klena, J. D., Flora, M. S., Daszak, P., Epstein, J. H., Luby, S. P., & Gurley, E. S. (2021). Genetic diversity of Nipah virus in Bangladesh. International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases, 102, 144–151. https://doi.org/10.1016/j.ijid.2020.10.041

Ranneby, B. (1984). "The maximum spacing method. An estimation method related to the maximum likelihood method." Scandinavian Journal of Statistics: 93-112.

Rastogi, M., Pandey, N., Shukla, A., & Singh, S. K. (2020). SARS coronavirus 2: from genome to infectome. Respiratory research, 21(1), 318. https://doi.org/10.1186/s12931-020-01581-z

Ruiz-Aravena, M., McKee, C., Gamble, A., Lunn, T., Morris, A., Snedden, C. E., Yinda, C. K., Port, J. R., Buchholz, D. W., Yeo, Y. Y., Faust, C., Jax, E., Dee, L., Jones, D. N., Kessler, M. K., Falvo, C., Crowley, D., Bharti, N., Brook, C. E., Aguilar, H. C., … Plowright, R. K. (2022). Ecology, evolution and spillover of coronaviruses from bats. Nature reviews. Microbiology, 20(5), 299–314. https://doi.org/10.1038/s41579-021-00652-2

Sanger, F., et al. (1977). "DNA sequencing with chain-terminating inhibitors." Proceedings of the National Academy of Sciences 74(12): 5463-5467.

Su, S., Wong, G., Shi, W., et al. (2016). "Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses." Trends in Microbiology, 24(6), 490-502.

Tamura, K., et al. (2021). "MEGA11: molecular evolutionary genetics analysis version 11." Molecular biology and evolution 38(7): 3022-3027.

Umakanthan, S., Sahu, P., Ranade, A. V., Bukelo, M. M., Rao, J. S., Abrahao-Machado, L. F., Dahal, S., Kumar, H., & Kv, D. (2020). Origin, transmission, diagnosis and management of coronavirus disease 2019 (COVID-19). Postgraduate medical journal, 96(1142), 753–758. https://doi.org/10.1136/postgradmedj-2020-138234

V'kovski, P., Kratzel, A., Steiner, S. et al. Coronavirus biology and replication: implications for SARS-CoV-2. Nat Rev Microbiol 19, 155–170 (2021). https://doi.org/10.1038/s41579-020-00468-6

Walls, A. C., Park, Y. J., Tortorici, M. A., et al. (2020). "Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein." Cell, 181(2), 281-292.e6.

Wang, L. F., & Eaton, B. T. (2007). Bats, civets and the emergence of SARS. Current Topics in Microbiology and Immunology, 315, 325-344. https://doi.org/10.1007/978-3-540-70962-6_13

Woo, P. C. Y., de Groot, R. J., Haagmans, B., Lau, S. K. P., Neuman, B. W., Perlman, S., Sola, I., van der Hoek, L., Wong, A. C. P., & Yeh, S. H. (2023). ICTV Virus Taxonomy Profile: Coronaviridae 2023. The Journal of general virology, 104(4), 10.1099/jgv.0.001843. https://doi.org/10.1099/jgv.0.001843

Woo, P. C., Lau, S. K., Lam, C. S., Lau, C. C., Tsang, A. K., Lau, J. H., Bai, R., Teng, J. L., Tsang, C. C., Wang, M., Zheng, B. J., Chan, K. H., & Yuen, K. Y. (2012). Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. Journal of virology, 86(7), 3995–4008. https://doi.org/10.1128/JVI.06540-11

Zhang, T., Wu, Q., & Zhang, Z. (2020). Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. Current Biology, 30(7), 1346-1351.

Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., ... & Shi, Z. L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature, 579(7798), 270-273.

"<_r2070t_r2070_r2071_r2072_r2073_direct-zol_rna_miniprep_plus_kit.pdf>."

"seqtk, Toolkit for processing sequences in FASTA/Q formats. Available from: https://github.com/lh3/seqtk