

**Evolutionary Analyses and Genomic Characterization of Emerging Viruses
from Animal Reservoirs Before and After the Host Switch**

Dissertation

to obtain the academic degree

Doctor rerum naturalium

(Dr. rer. nat.)

submitted to the Faculty of Life Sciences of Humboldt-Universität zu Berlin

by

M. Sc. Anna-Lena Sander

Institute of Virology; Charité - Universitätsmedizin Berlin

President of Humboldt-Universität zu Berlin

Prof. Dr. Julia von Blumenthal

Dean of the Faculty of Life Sciences of Humboldt-Universität zu Berlin

Prof. Dr. Dr. Christian Ulrichs

1st reviewer: Prof. Dr. Kai Matuschewski

2nd reviewer: Prof. Dr. Jan Felix Drexler

3rd reviewer: Prof. Dr. Alex Greenwood

Date of defense: 27 March 2023

Parts of this doctoral thesis are published in peer-reviewed journals.

Sander AL, Corman VM, Lukashev AN, Drexler JF. Evolutionary Origins of Enteric Hepatitis Viruses. *Cold Spring Harb Perspect Med.* 2018;8(12).

de Oliveira Carneiro I, **Sander AL**, Silva N, Moreira-Soto A, Normann A, Flehmig B, et al. A Novel Marsupial Hepatitis A Virus Corroborates Complex Evolutionary Patterns Shaping the Genus *Hepatovirus*. *J Virol.* 2018;92(13).

Feng H, **Sander AL**, Moreira-Soto A, Yamane D, Drexler JF, Lemon SM. Hepatovirus 3ABC proteases and evolution of mitochondrial antiviral signaling protein (MAVS). *J Hepatol.* 2019 Jul;71(1):25-34.

Sander AL*, Yadouleton A*, Moreira-Soto A*, Tchibozo C, Hounkanrin G, Badou Y, et al. An Observational Laboratory-Based Assessment of SARS-CoV-2 Molecular Diagnostics in Benin, Western Africa. *mSphere.* 2021 Jan 13;6(1).

Sander AL*, Yadouleton A*, de Oliveira Filho EF, Tchibozo C, Hounkanrin G, Badou Y, et al. Mutations Associated with SARS-CoV-2 Variants of Concern, Benin, Early 2021. *Emerg Infect Dis.* 2021 Nov;27(11):2889-903.

Yadouleton A*, **Sander AL***, Adewumi P, de Oliveira Filho EF, Tchibozo C, Hounkanrin G, et al. Emergence of SARS-CoV-2 Delta Variant, Benin, May-July 2021. *Emerg Infect Dis.* 2021 Nov 22;28(1).

Sander AL, Moreira-Soto A, Yordanov S, Toplak I, Balboni A, Ameneiros RS, et al. Genomic determinants of Furin cleavage in diverse European SARS-related bat coronaviruses. *Commun Biol.* 2022;5(1):491.

* These authors are shared first authors.

Declaration

I hereby declare that I completed the doctoral thesis independently based on the stated resources and aids.

I have not applied for a doctoral degree elsewhere and do not have a corresponding doctoral degree.

I have not submitted the doctoral thesis or parts of it, to another academic institution and the thesis has not been accepted or rejected.

I declare that I have acknowledged the Doctoral Degree Regulations which underlie the procedure of the Faculty of Life Sciences of Humboldt-Universität zu Berlin, as amended on 5th March 2015.

Furthermore, I declare that no collaboration with commercial doctoral degree supervisors took place, and that the principles of Humboldt-Universität zu Berlin for ensuring good academic practice were abided by.

Berlin, 15 December 2022

Anna-Lena Sander

Erklärung

Hiermit erkläre ich, die Dissertation selbstständig und nur unter Verwendung der angegebenen Hilfen und Hilfsmittel angefertigt zu haben.

Ich habe mich anderwärts nicht um einen Doktorgrad beworben und besitze keinen entsprechenden Doktorgrad.

Ich erkläre, dass ich die Dissertation oder Teile davon nicht bereits bei einer anderen wissenschaftlichen Einrichtung eingereicht habe und dass sie dort weder angenommen noch abgelehnt wurde.

Ich erkläre die Kenntnisnahme der dem Verfahren zugrunde liegenden Promotionsordnung der Lebenswissenschaftlichen Fakultät der Humboldt Universität zu Berlin vom 5. März 2015.

Weiterhin erkläre ich, dass keine Zusammenarbeit mit gewerblichen Promotionsbearbeiterinnen/Promotionsberatern stattgefunden hat und, dass die Grundsätze der Humboldt-Universität zu Berlin zur Sicherung guter wissenschaftlicher Praxis eingehalten wurden.

Berlin, den 15. Dezember 2022

.Anna-Lena Sander

In conclusion, it appears that nothing can be more improving to a young naturalist, than a journey in distant countries.

Charles Darwin

TABLE OF CONTENTS

LIST OF ABBREVIATIONS	i
-----------------------------	---

SUMMARY	v
---------------	---

ZUSAMMENFASSUNG	vii
-----------------------	-----

CHAPTER 1

GENERAL INTRODUCTION

1.1 Virus and disease emergence	1
1.2 Barriers to host switching.....	2
1.3 Past and present viral host switches into humans:HAV and SARS-CoV-2.....	4
1.4 Open scientific questions	11

CHAPTER 2

EVOLUTIONARY CONSERVATION OF IMMUNE ESCAPE MECHANISMS IN HEPATOVIRUSES

2.1 Introduction.....	13
2.2 Outline and aims of this chapter.....	16
2.3 Methods.....	17
2.4 Results.....	20
2.5 Discussion.....	28

CHAPTER 3

SARS-CoV-2 EVOLUTION IN BENIN, WEST AFRICA

3.1 Introduction.....	33
3.2 Outline and aims of this chapter.....	35
3.3 Methods.....	36
3.4 Results.....	39
3.5 Discussion.....	51

CHAPTER 4

GENOMIC DETERMINANTS OF FURIN CLEAVAGE IN EUROPEAN SARS-RELATED BAT CORONAVIRUSES

4.1 Introduction.....	55
4.2 Outline and aim of this chapter	58
4.3 Methods.....	59
4.4 Results.....	63
4.5 Discussion.....	74

CHAPTER 5

FINAL DISCUSSION AND OUTLOOK

5.1 Looking forward: Prospective actions for pandemic prevention	79
5.2 Concluding remarks	84

SUPPLEMENTARY INFORMATION	85
---------------------------------	----

REFERENCES.....	127
-----------------	-----

LIST OF ABBREVIATIONS

aa	Amino acid
ACE2	Angiotensin-converting enzyme 2
Acc. No.	Accession number
AI	Arginine index
AIV	Avian Influenza A virus
CoV	Coronavirus
COVID-19	Coronavirus disease 2019
Ct	Cycle threshold
CUB	Codon usage bias
DNA	Deoxyribonucleic acid
dsRNA	Double-stranded RNA
E	Envelope
EBOV	Ebola virus
eHAV	Enveloped HAV
EID	Emerging infectious disease
ENC	Effective number of codons
ESCRT	Endosomal sorting complexes required for transport
FCS	Furin cleavage site
FEL	Fixed effect likelihood
Gt	Genotype
GTR	General time reversible
HA	Hemagglutinin
HAV	Hepatitis A virus
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HCoV	Human coronavirus
HIV	Human immunodeficiency virus
H5N1	High-pathogenic avian influenza virus
ICTV	International Committee on the Taxonomy of Viruses
IFN	Interferon

IRES	Internal ribosome entry site
IRF3	Interferon regulatory factor 3
LFHB	Laboratoire des Fièvres Hémorragiques Virales du Benin
LPAI	Low-pathogenic avian influenza virus
LRT	Likelihood ratio tests
MAVS	Mitochondrial antiviral-signalling protein
MDA-5	Melanoma differentiation-associated protein-5
MERS-CoV	Middle East respiratory syndrome coronavirus
MHAV	Marsupial HAV
ML	Maximum-likelihood
N	Nucleoprotein
NF	Nuclear factor
NGS	Next-generation sequencing
NHP	Non-human primate
NSP	Non-structural protein
Nt	Nucleotide
ORF	Open reading frame
PAML	Phylogenetic analysis by maximum likelihood
PAMP	Pathogen-associated molecular patterns
PCR	Polymerase chain reaction
PRR	Pattern recognition receptor
qPCR	Quantitative PCR
RACE	Rapid amplification of cDNA ends
RBD	Receptor binding domain
REL	Random effect likelihood
RIG-I	Retinoic acid activated gene I
RLR	Retinoic acid activated gene I (RIG-I)- like RNA helicases
RNA	Ribonucleic acid
RT-PCR	Reverse transcription-PCR
S	Spike protein
SARS-CoV	Severe acute respiratory syndrome coronavirus
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SLAC	Single likelihood ancestor counting

SrC	SARS-related coronavirus
ssRNA	Single-stranded virus RNA
TIM-1	T cell immunoglobulin and mucin-containing domain protein 1
TLR	Toll-like receptors
TLR3	Toll-like receptor 3
TMPRSS2	Transmembrane Serine Protease 2
tRNA	Transfer RNA
UTR	Untranslated region
VSV	Vesicular stomatitis virus
VOC	Variant of concern
VP	Viral protein
ZAP	Zinc-finger antiviral protein

SUMMARY

Emerging viral diseases of zoonotic origin pose an increasing threat to global health. Despite intense research, we do not understand where and when animal-associated pathogens emerge, or by what evolutionary processes these transmissions become possible; best illustrated by the unexpected emergence of the Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) at the end of 2019. This thesis is concerned with the topic of viral genetic adaptation before and after cross-species transmissions, contributing to a better understanding of host switching processes and their underlying mechanisms.

The 1st chapter of this thesis introduces newly emerging viruses in humans and backgrounds on host switching. In addition, this chapter provides a broad overview of the viruses discussed in this thesis, hepatitis A virus (HAV) and SARS-CoV-2. The following chapters 2-4 present the performed studies and their results.

Chapter 2 examines the adaptation of hepatoviruses to their reservoir hosts. In the first part of the chapter, I address the question of whether HAV-specific immune escape mechanisms are conserved in hepatoviruses from small mammals. To this end, I use genomic sequences of highly diverse hepatoviruses from different hosts and analyse their codon usage. The second part of the chapter examines the extent to which MAVS protein cleavage, a strategy used by hepatoviruses to disrupt the host immune response, has influenced the evolution of hepatoviruses and their hosts. Therefore, I am investigating MAVS coding sequences from bats, rodents, and primates using selection pressure analyses to identify genomic regions that have been genetically altered as a result of evolutionary pressure. I can show with these analyses that genetic adaptations have occurred at hepatovirus cleavage sites in MAVS genes, suggesting an evolutionary arms race between hepatoviruses and their reservoir hosts. In sum, the results of this chapter deepen the basic understanding of which mechanisms have influenced the evolution and host switching processes of hepatoviruses.

The 3rd chapter of this thesis focuses on the genetic adaptation of SARS-CoV-2 in the human host. In the context of a genomic surveillance strategy, I determine the viral variants in several hundred SARS-CoV-2 positive nasopharyngeal swab samples from patients sampled during the first 1.5 years of the pandemic in Benin; a West African country with high transmission rates. I use phylogenetic analysis to show the parallel introduction of different SARS-CoV-2 lineages into Benin at the onset of the pandemic. Detailed sequence comparisons reveal the evolution of

mutations of concern over the course of the pandemic, as well as their convergent adaptation and accumulation within different SARS-CoV-2 lineages. The results of this chapter contribute to a better understanding of the adaptive evolution of SARS-CoV-2 in humans while contributing significantly to much-needed genomic surveillance in a resource-limited country.

Chapter 4 examines the genetic diversity and evolution of SARS-related coronaviruses in the bat reservoir. I determine and analyse sequences of the genes encoding the spike protein of SARS-related coronaviruses to provide evidence for which mechanisms or genetic predispositions enable the acquisition of a furin cleavage site. *In silico* modelling reveals that nucleotide exchanges, insertions, or recombination likely underly the emergence of furin cleavage sites in the spike protein of some bat-associated coronaviruses. In this context, I discover two reservoir-bound SARS-related coronaviruses in which a single nucleotide exchange would predictably suffice to enable furin cleavage. In addition, using RNA structural modelling, I show that these viruses possess genetic features such as purine-richness or RNA secondary structures, that likely facilitate genetic modifications. The results of this chapter make a significant contribution to the biological risk assessment of the zoonotic potential of reservoir-bound coronaviruses and provide insights into the origin and evolution of SARS-CoV-2.

Finally, in chapter 5, I explain which new insights result from the achieved results of this thesis and address actions, which in my opinion will be of particular importance to prevent zoonotic infectious diseases in the future.

ZUSAMMENFASSUNG

Neu auftretende Viruskrankheiten zoonotischen Ursprungs stellen eine zunehmende Gefahr für die globale Gesundheit dar. Wie das unerwartete Auftreten von dem Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Ende 2019 zeigte, gibt es allerdings trotz jahrelanger intensiver Forschung Verständnislücken, wo und wann mit Tieren assoziierte Krankheitserreger auftauchen, oder durch welche evolutionären Vorgänge diese Übertragungen möglich werden. Diese Arbeit befasst sich mit dem Thema der genetischen Adaptation von Viren bevor und nachdem ein Wirtswechsel stattgefunden hat und leistet damit einen Beitrag zum besseren Verständnis von Wirtswechselprozessen und ihren zugrundeliegenden Mechanismen.

Das 1. Kapitel dieser Arbeit stellt neuauftretenden Viren im Menschen vor und gibt Hintergründe zu Wirtswechseln. Zudem gibt dieses Kapitel einen groben Überblick über die in dieser Arbeit behandelten Viren, das Hepatitis A Virus (HAV) sowie SARS-CoV-2. Die folgenden Kapitel 2-4 präsentieren die durchgeführten Studien und ihre Ergebnisse

Kapitel 2 untersucht die Adaptation von Hepatoviren an ihre Reservoirwirte. Im ersten Teil des Kapitels wird der Frage nachgegangen, ob HAV-spezifische Immunfluchtsmechanismen in Hepatoviren aus kleinen Säugetieren konserviert sind. Hierzu benutze ich Genomsequenzen hochdiverser Hepatoviren aus verschiedenen Wirtstieren und analysiere deren Codonnutzung. Der zweite Teil des Kapitels untersucht inwieweit die MAVS-Protein Spaltung, eine von Hepatoviren genutzte Strategie die Wirtsimmunantwort herunterzufahren, auf die Evolution von Hepatoviren und deren Wirte gewirkt hat. Ich analysiere MAVS Gensequenzdatensätze von Fledermäusen, Nagetieren und Primaten mittels Selektionsdruckanalysen um genomische Regionen zu identifizieren die sich durch Evolutionsdruck genetisch verändert haben. Ich kann mittels dieser Analysen zeigen, dass es an den Hepatovirus Schnittstellen in den MAVS Genen zu genetischen Anpassungen gekommen ist, was auf ein evolutionäres Wettrüsten zwischen Hepatoviren und ihren Reservoirwirten hindeutet. Alles in allem vertiefen die Ergebnisse dieses Kapitels das grundlegende Verständnis, welche Mechanismen die Evolution und Wirtswechsel von Hepatoviren beeinflusst haben.

Das 3. Kapitel dieser Arbeit befasst sich mit der genetischen Anpassung des SARS-CoV-2 im menschlichen Wirt. Im Kontext der genomischen Überwachung bestimme ich die Virusvarianten in mehreren Hundert SARS-CoV-2 positiv getesteten

Nasenrachenabstrichproben von Patienten, die während der ersten 1,5 Jahre der Pandemie in Benin; einem westafrikanischen Land mit hohen Transmissionsraten, beprobt wurden. Ich zeige anhand einer phylogenetischen Analyse die parallele Eintragung verschiedener SARS-CoV-2 Linien nach Benin zu Beginn der Pandemie. Detaillierte Sequenzvergleiche zeigen die Evolution besorgniserregender Mutationen im Laufe der Pandemie sowie deren konvergente Adaptation und Akkumulation innerhalb verschiedener SARS-CoV-2 Linien. Die Ergebnisse dieses Kapitels tragen zum besseren Verständnis der adaptiven Evolution von SARS-CoV-2 im Menschen bei und leisten gleichzeitig einen Beitrag zur dringend benötigten genomischen Überwachung in einem ressourcenbeschränkten Land.

Kapitel 4 untersucht die genetische Vielfalt und Evolution von SARS-verwandten Coronaviren im Fledermaus Reservoir. Ich bestimme und analysiere die genomischen Sequenzen von spike-codierenden Genen verschiedener SARS-verwandter Coronaviren, um Hinweise zu erhalten durch welche Mechanismen oder genetische Prädispositionen der Erwerb einer Furin-Spaltstelle ermöglicht wird. Durch *in-silico* Modellierungen zeigt sich, dass Nukleotidaustausche, Insertionen oder Rekombination dazu führen können, dass die Spike Proteine einiger Fledermaus-assoziierten Coronaviren eine Furinspaltstelle erwerben können. In diesem Zusammenhang entdeckte ich zwei reservoirgebundene SARS-verwandte Coronaviren in welchen prädiktiv ein einziger Nukleotidaustausch die Spaltung durch Furin ermöglichen könnte. Darüber hinaus zeige ich mittels RNA Strukturmodellierungen, dass diese Viren RNA-Sekundärstrukturen besitzen, wodurch genetische Veränderungen besonders leicht hervorgerufen werden könnten. Die Ergebnisse dieses Kapitels leisten einen bedeutenden Beitrag zur evolutionsbiologischen Risikobewertung des zoonotischen Potentials reservoirgebundener Fledermauscoronaviren und geben Einblicke in den Ursprung und die Evolution von SARS-CoV-2.

Abschließend, in Kapitel 5, erläutere ich welche Erkenntnisse sich aus den erzielten Ergebnissen dieser Arbeit ergeben und gehe auf Handlungsaspekte ein, die meiner Meinung nach von besonderer Wichtigkeit sein werden, um zukünftig präventiv zoonotische Infektionskrankheiten zu erkennen und vorzubeugen.

CHAPTER 1

GENERAL INTRODUCTION

1.1 VIRUS AND DISEASE EMERGENCE

Infectious diseases constitute illnesses caused by microorganisms such as protozoa, bacteria, fungi, parasites, prions or viruses (Jones et al., 2008). An infectious disease is defined as an „emerging” infectious disease (EID) when a previously unknown disease occurs in humans for the first time, or as "re-emerging" when a previously known disease re-occurs in temporally or spatially fulminant cases (reviewed in (Fauci and Morens, 2012)). Pathogen transmission can be facilitated through the environment, humans or via animal contact. The latter are so-called “zoonotic” diseases and account for more than 60% of EID events (Jones et al., 2008; Woolhouse and Gowtage-Sequeria, 2005) .

Zoonotic viruses, especially those that have a reservoir in wild animals (Jones et al., 2008; Woolhouse and Gowtage-Sequeria, 2005) have been the cause of serious EID outbreaks in recent decades (**Figure 1**).

The emergence of Monkeypox virus, Nipah virus or this era’s most recent example the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), perfectly display the potential of viral agents to switch hosts and cause deadly outbreaks in the human population. A commonality of all these emerging or re-emerging viruses is that they were previously established animal viruses that emerged in the human population after a successful cross-species host switch with or without the involvement of intermediate hosts

¹ Parts of this chapter have been previously published in:

Sander AL, Corman VM, Lukashev AN, Drexler JF. Evolutionary Origins of Enteric Hepatitis Viruses. Cold Spring Harb Perspect Med. 2018;8(12).

Copyright permissions are granted to the authors of the publications.

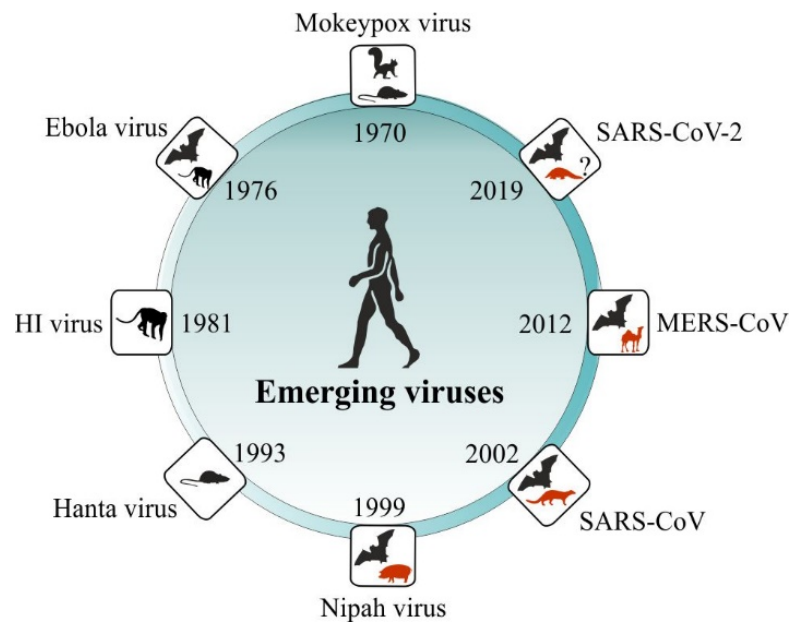


Figure 1. Examples of emerging and re-emerging viruses that were zoonotically transmitted from wild animals to humans in the past decades. Depicted are reservoir hosts (black) and in transmission involved intermediate hosts (red).

1.2 BARRIERS TO HOST SWITCHING

Viral agents are usually restricted to their host species due to several ecological and molecular factors (Parrish et al., 2008). Nevertheless, in rare cases viruses can be transmitted from one host to another. While the mechanisms that allow a virus to cross the species barrier are diverse but not entirely understood, the new host must first be exposed to the virus, then the virus must successfully infect the new host, spread within the new host population, and subsequently adapt to maintain ongoing transmission cycles (reviewed in (Parrish et al., 2008)). Each of these steps is influenced by several evolutionary and ecological variables that can be decisive for a virus to switch hosts and

a) immediately becomes extinct in the new host species without establishing ongoing transmission (spillover event)

b) maintains limited transmission chains but nevertheless becomes extinct in the new host species (outbreak)

or c) adapts successfully to establish ongoing transmission chains in the new host population (epidemic) (Antia et al., 2003).

The most important prerequisite for cross-species transmissions to occur is the contact of the virus with an alternative host. Limited contact due to environmental or demographic barriers can therefore likely prevent many cross-species transmissions already at the first stage (Morse, 1995). However, through the destruction of natural habitats, the close cohabitation of humans and animals, livestock breeding and the consumption of (bush)meat, humans have overcome nearly all geographic barriers that have ever separated us from animal territories. As a result, we are highly exposed to (wildlife) animals and their associated pathogens (Gibb et al., 2020; Morse, 1995) which dramatically increases the risk of direct or indirect zoonotic transmission from reservoir or intermediate hosts. Deforestation in Asia, for example, has led to closer contact between humans and bats and has resulted in the deadly Nipah virus being transmitted to humans through fruits contaminated with bat excreta. (Hahn et al., 2014). The implications of climate change are yet contributing factors driving host switches as extended periods of rain or drought for example lead to loss of biodiversity or food scarcity, resulting in the displacement of animal population to human habitats or vice versa (Carlson et al., 2022). At the same time, global warming contributes primarily to the spread of zoonotically transmitted EIDs in previously spared, often colder regions of the world, exemplified by the spread of *Aedes aegypti* and *Aedes albopictus*, key mosquito vectors of several Arboviruses, in Central Europe (Kraemer et al., 2019; Liu-Helmersson et al., 2019).

At the next stage, host barriers may hinder the virus to infect the new host. The successful infection of the appropriate target cell of the new host can be restricted at many different levels. After a receptor-related barrier, which directly restricts the initial viral interaction with the target cell, the virus needs to evade the antiviral host defence such as the interferon system, to at least be permitted for genome replication. For example, human Hepatitis A virus (HAV) disrupts innate immune response in its host by cleavage of the mitochondrial antiviral signalling (MAVS) protein, an important mediator for signal transduction within the cascade of type-I interferon immune response (Yang et al., 2007). Studies revealed that this mechanism also defines the host species range of HAV (Hirai-Yuki et al., 2016), suggesting that this host-pathogen interaction may play an important role in cross-species transmission of hepatoviruses in general. While the evolutionary relatedness of hosts may facilitate transmission processes due to higher homology of, for example entry receptors, it does not seem to be the most critical factor (Parrish et al., 2008) as recent research revealed that host switches of RNA viruses between different vertebrate classes are not infrequently happening during virus evolution (Shi et al., 2018). Moreover, for example Avian influenza viruses can be successfully transmitted from their primary natural reservoir host (vertebrate class Aves) to different hosts within the

vertebrate class Mammalia. In case of successful cell infection of a new host, the virus has to adapt to potentially different intracellular environmental conditions in order to successfully replicate its genome in the new host. Finally, as one of the last intracellular host restrictions, the release of viral particles can be prevented by the host cell.

Following the successful infection of a new host, ongoing transmission in the new host population is obligatory to keep the virus from extinction. For example, increased transmission potential is given in the ecology of rodents, where offspring are produced in large and frequent litters (Han et al., 2015; Olival et al., 2017) or within bats that live in densely populated caves (Calisher et al., 2006; Luis et al., 2013). Likewise, enormous human population densities and global connections likely promote ongoing transmission. The current SARS-CoV-2 pandemic is the best example of how unlimited geographic mobility contributes to the global spread of a newly emerging virus within a short time.

Upon successfully overcoming all aforementioned barriers, a co-evolutionary arms race of host and virus originates, the dynamics of which drive the evolution of the other. According to the Red Queen hypothesis, competing species (e.g. host and virus) are constantly adapting and counter-adapting their genomes in order to stay in the same place and avoid extinction (Van Valen, 1973). Genetic mutations enable viruses to change their genetic material directly, which may facilitate to escape the hosts immune system or enhance their virulence or transmission efficacy. With around 10^{-4} - 10^{-6} substitutions per genomic site per year (Duffy et al., 2008), viral mutation rates are much higher than those of their hosts (Duffy, 2018). On that scale however, RNA viruses mutate faster than DNA viruses due to different polymerase fidelity (Sanjuan et al., 2010) which may explain why RNA viruses shift hosts more frequently than other viruses (Geoghegan et al., 2017). Nevertheless, too many mutations can come at the expense of other aspects of the virus' fitness, which may then result in a single spillover event without ongoing transmission. Reassortment and recombination are further mechanisms by which viruses create new variants or strains which alter the pool of genetic variation (Vijaykrishna et al., 2015). Population bottlenecks (genetic drift) or natural selection then determine which of the strains persist in the new host population and become stabilized (Geoghegan et al., 2016).

1.3 PAST AND PRESENT VIRAL HOST SWITCHES INTO HUMANS: HAV AND SARS-COV-2

Theoretically, genetic adaptation may allow all viruses to switch hosts, however, some viruses tend to do so more frequently than others. This is for example the case for RNA viruses in

general (Kitchen et al., 2011; Woolhouse and Gowtage-Sequeria, 2005) and viruses of the family *Picornaviridae* in particular (Geoghegan et al., 2017). As one of the largest virus families, they possess one of the smallest RNA genomes, with only about 7-9 kb. One prototypic agent of the human picornaviruses is HAV. HAV is known since the 1970s and is one of the major causes of acute viral hepatitis in humans, accounting for millions of infections and more than 11,000 deaths annually (WHO, 2017). Another example of an RNA virus, and one which has probably only recently successfully switched hosts, is SARS-CoV-2. SARS-CoV-2 belongs to the species *SARS-related coronavirus* within the genus *Betacoronavirus* of the family *Coronaviridae*. It newly emerged at the end of 2019 and causes severe respiratory illness, accounting for more than 300 million infections and 5.5 million deaths until January 2022 (WHO, 2022a).

Just as these two viruses belong to different viral families, they also differ in several aspects such as their genomic properties, clinical outcome, and infection patterns. However, as both viruses are zoonoses and both originated in small mammals, they also share common features.

The following chapter compares the characteristics and evolutionary biology of both viruses.

1.3.1 Differences between HAV and SARS-CoV-2

Transmission route

While transmission of HAV is predominantly associated with contaminated water and food and thus constitutes a so-called fecal-oral infection (Jacobsen, 2018), SARS-CoV-2 is transmitted via infectious respiratory fluids that are either directly inhaled or absorbed by exposed mucous membranes, which represents a so-called droplet or aerosol infection (Richard et al., 2020).

Disease outcome

HAV causes acute and usually self-limiting liver infections. A chronic form of the disease does not exist. Most infections in children below 5 years are asymptomatic or mild whereas the majority of adults develop symptoms (Armstrong and Bell, 2002). Only about 30% or less of young children develop symptoms, but in most cases the infection is asymptomatic. Symptomatic infections last about 2 months on average but can prolong for more than half a year. HAV infection causes lifelong immunity. SARS-CoV-2 causes an acute respiratory disease named coronavirus disease 2019 (COVID-19). Asymptomatic infections are common and account for about one-fourth to one-third of infections (Sah et al., 2021). Reinfection with

SARS-CoV-2 is possible, as protective immunity after natural infection or vaccination seems to wane rapidly after three to seven months (Rennert et al., 2022). To date, the course of infection seems to be acute but non-chronic. However, cases of long-term symptoms or consequences affecting pulmonary as well as numerous extrapulmonary organs have been reported, a syndrome that has been named Long COVID (Al-Aly et al., 2021; O'Donnell and Chappell, 2021).

Virus properties

With a genome size of about 7,5 kb, HAV has one of the smallest RNA genomes and is about four times smaller than the genome of SARS-CoV-2, which has an approximate size of 30 kb. HAV is shed in faeces as naked, non-enveloped virus particles, whereas in blood the virus circulates in an enveloped form (eHAV), enclosed in host cell membranes (Feng et al., 2013). In contrast, the SARS-CoV-2 virus particles are always enveloped by a lipid bilayer in which the envelope, membrane and spike proteins are attached.

The HAV genome consists of one single open reading frame (ORF) encoding one polyprotein which is subsequently processed into 10 mature proteins, namely viral protein (VP) 4, VP2, VP3, VP1-pX, 2B, 2C, 3A, VPg, 3C^{pro} and 3D^{pol} (McKnight and Lemon, 2018). Untranslated regions (UTRs) at the genome termini generally contain significant secondary structures, among them a type III internal ribosome entry site (IRES) within the 5' untranslated region (UTR) which promotes the initiation of protein synthesis in a cap-independent manner (Brown et al., 1991) and a 3' terminating poly(A) tail (**Figure 2A**). Distinct from other picornaviruses, the VP1 capsid protein of HAV has a carboxy-terminal extension termed pX (often referred to as 2A), which is involved in capsid assembly (Cohen et al., 2002), pathogenesis (Emerson et al., 2002; Harmon et al., 1995) and possibly quasi-envelopment (Feng et al., 2014).

The SARS-CoV-2 genome on the other hand, contains approximately 12 ORFs encoding 27 proteins (**Figure 2B**). ORF1ab represents about two-thirds of the genome and encodes the polyproteins pp1a and pp1ab, which are processed by a -1 frameshift due to a hairpin structure (Bhatt et al., 2021). ORF1ab encodes a total of 16 non-structural proteins (NSP) that constitute the replicase gene and are mainly involved in the replication process (reviewed in (Rohaim et al., 2021)). According to the current state of knowledge, the other one-third of the genome encodes at least seven (3a, 6, 7a, 7b, 8b, 9b, and ORF10) accessory proteins (Finkel et al., 2020) associated with pathogenesis and/or evasion of host antiviral defences (reviewed in (Redondo et al., 2021)), and another four structural proteins forming the virus particle. The latter are

conserved among coronaviruses (CoVs) and include the spike (S) protein mediating attachment to cellular receptors and subsequent entry of the virus particle into the host cell, the nucleocapsid protein (N) encapsulating the viral RNA thereby forming the ribonucleocapsid, and the membrane (M) and envelope (E) proteins localized in the bilayer membrane surrounding the ribonucleocapsid. The SARS-CoV-2 genome is flanked by 5' and 3' UTRs which contain cis-acting secondary RNA structures crucial for RNA replication and transcription.

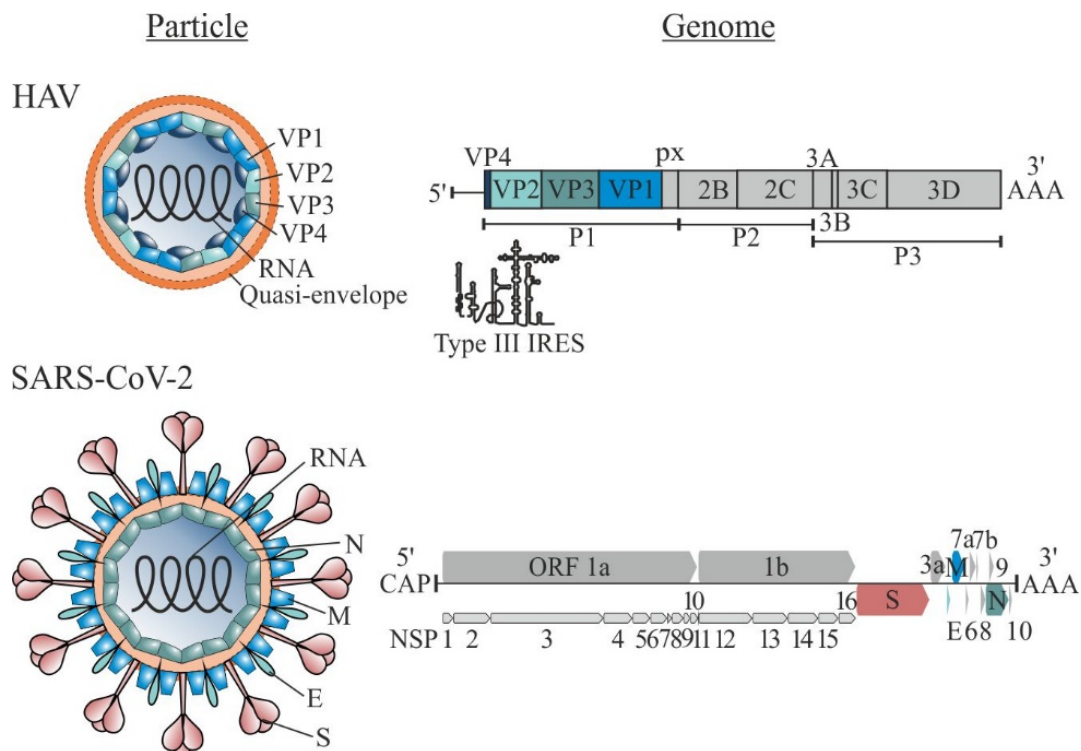


Figure 2. Virus particle and genome characteristics of HAV and SARS-CoV-2. E, envelope protein; IRES, internal ribosomal entry site; M, membrane protein; N, nucleocapsid protein; NSP, nonstructural protein; ORF, open reading frame; S, spike protein; VP, viral protein. Type III IRES of HAV is depicted below the genome. Adapted from (Rasche et al., 2019b).

Cell entry receptor

Viral entry into host cells is initiated by protein-protein- interactions between the viral cell surface glycoprotein and the host cellular receptor. T cell immunoglobulin and mucin-containing domain protein 1 (TIM-1) were found as the cell entry receptor of HAV for decades (Kaplan et al., 1996). However, with the discovery of quasi-enveloped HAV particles, it was shown that TIM1 facilitates cell surface binding of eHAV but not naked HAV, and that it is not essential for cell entry for either viral form (Das et al., 2017). It thus remains to be elucidated which receptor is utilized by HAV naked and quasi-enveloped particles to enter hepatocytes and epithelial cells (Rasche et al., 2019b). The host receptor protein of SARS-CoV-2 was

identified as angiotensin-converting enzyme 2 (ACE2), the same receptor that is utilized by SARS-CoV and other bat-associated SARS-related coronaviruses (SrC) (Wells et al., 2021).

1.3.2 Similarities between HAV and SARS-CoV-2

Genetic diversity in small mammals

For a long time, it was thought that HAV was restricted to humans and non-human primates (NHPs) only, comprising six different genotypes with genotypes I-III infecting humans, and genotypes IV-VI found in different species of Old World monkeys (Cristina and Costa-Mattioli, 2007; Robertson, 2001). These virus strains show little variation in nucleotide sequence over time or geography. The recent discovery of a great diversity of distantly related hepatoviruses in various small mammals including rodents, a tree shrew, bats, hedgehogs, shrews, and seals (Anthony et al., 2015; Drexler et al., 2015; Yu et al., 2016) expanded the genus *Hepatovirus* dramatically (**Figure 3A**). Hepatoviruses infecting humans thus only represent a small proportion of the known virus diversity of today. The genus now comprises 16 putative virus species, of which nine are approved by the International Committee on the Taxonomy of Viruses (ICTV) as *Hepatovirus A* (HAV) to *Hepatovirus I* (Zell et al., 2017). These 16 putative virus species were recovered from seven different orders of mammalian hosts, including Primates, Rodentia, Scandentia, Chiroptera, Eulipotyphla, Marsupialia and Carnivora.

Soon after the emergence of SARS-CoV-2, genetically related viruses have been identified in different Asian rhinolophid bats (Wacharapluesadee et al., 2021; Zhou et al., 2021; Zhou et al., 2020c) and Malayan pangolins (Lam et al., 2020; Xiao et al., 2020). In the case of the conspecific coronavirus termed SARS-CoV, which caused an outbreak in 2003-2004 mainly in China, genetically closely related viruses were found in masked palm civets and raccoon dogs. Additionally, a large diversity of genetically closely related viruses was found in rhinolophid bats across Asia, Africa and Europe (Drexler et al., 2010; Hu et al., 2017; Li et al., 2005)) (**Figure 3B**). However, SrC from European horseshoe bats are phylogenetically distinct from SrC detected in Asia. Based on genomic sequence distances, all these viruses can be classified within the *SARS-related coronavirus* species and are thus, in contrast to the hepatovirus diversity, conspecific. The virus species *SARS-related coronavirus* thus infects hosts belonging to four different mammalian orders to date, including Chiroptera, Carnivora, Pholidota and Primates.

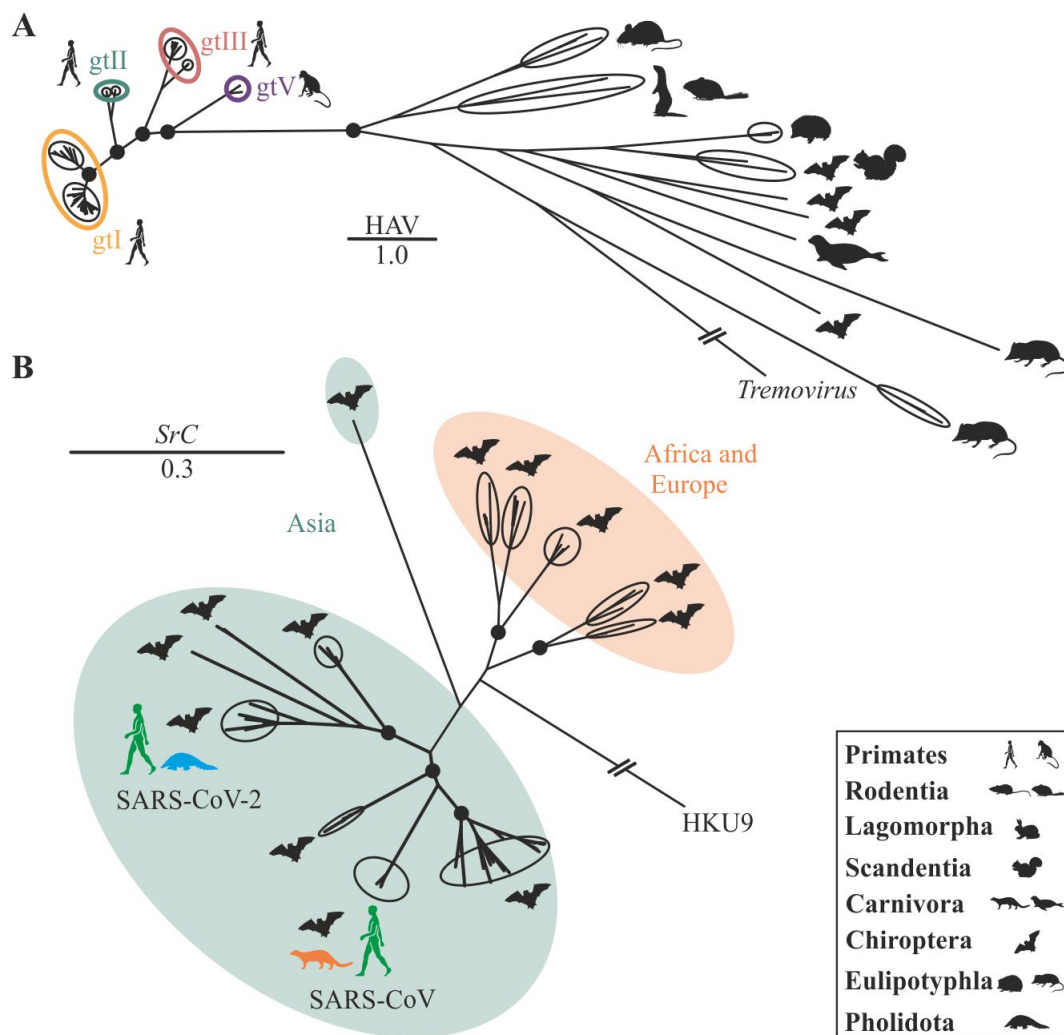


Figure 3. Phylogenies of the complete genetic diversity of (A) hepatoviruses and (B) SARS-related coronaviruses. **A**) Bayesian phylogeny of the full genome of hepatoviruses. Genotype (gt)IV and gtVI are defined only by partial sequence information and were, thus, not included (Nainan et al., 1991; Robertson et al., 1992). Bayesian posterior probability support above 0.9 at nodes is highlighted by filled circles. The scale bar indicates genetic distance. HAV genotypes are indicated by roman numerals. Adapted from (Sander et al., 2018). **B**) Approximately-maximum-likelihood phylogeny showing the complete diversity of members of the species *SARS-related coronavirus*. Circles at nodes indicate support of grouping in $\geq 90\%$ of 1,000 bootstrap replicates. Scale bar represents nucleotide substitutions per site. Adapted from (Sander et al., 2022).

Zoonotic origin

Despite the diversity of animal hepatoviruses, so far, no ongoing zoonotic transmission of HAV has been reported. However, the expanded genetic diversity of HAV and their hosts suggests that the emergence of HAV in humans seems likely to have had a zoonotic origin. Ancestors of human HAV have been projected to rodents (Drexler et al., 2015), similar to Hantaviruses (Jonsson et al., 2010), Poxviruses (Hughes et al., 2010) or some human coronaviruses (HCoV-OC43 and HKU1) (Corman et al., 2018). When in time exactly the host switch of HAV from rodents into humans took place remains unclear but it is assumed that this event occurred between 10-12 thousand years ago (Kulkarni et al., 2009).

While the host species shift of HAV was likely non-recent, the emergence SARS-CoV-2 in the human population is an example of a zoonotic virus that switched into humans only recently. In fact, the emergence of SARS-CoV-2 in humans can likely be attributed to multiple, but at least two separate cross-species transmission events (Pekar et al., 2022). As nearly all of the first SARS-CoV-2 cases have been linked to the Huanan seafood and wildlife market in Wuhan (Li et al., 2020; Worobey et al., 2022) probable zoonotic transmission within the market seems likely. However, even though several susceptible animals were sold at this market (Freuling et al., 2020; Xiao et al., 2021), the animal host from which transmission to humans has emanated was not yet identified. As for other SrC, the precursor of SARS-CoV-2 most likely originated in rhinolophid bats (Latinne et al., 2020). Direct virus transmission from bats to humans may be facilitated by different traditions including bat hunting, consumption of bat meat (Anti et al., 2015) or their use in traditional medicine (Mildenstein et al., 2016) but no bats were sold at the Wuhan market. Till date, the most closely related animal-associated SARS-CoV-2 strains were found in Asian rhinolophid bats and Malayan pangolins (Lam et al., 2020; Zhou et al., 2020c). However, the genetic relatedness with SARS-CoV-2 of only about 96 and 85-92% nucleotide sequence identity, respectively, suggests that zoonotic infection of humans with these bat- and pangolin-associated SrC seems rather unlikely (Jo et al., 2020), if compared with the more than 99% nucleotide identity that likely facilitated the zoonotic transmission of SARS-CoV between humans and civets (Shi and Hu, 2008). Altogether, these findings indicate that SARS-CoV-2 has been transmitted by an animal intermediate host. Although there is clear evidence of transmission from animals such as minks (Lu et al., 2021; Oude Munnink et al., 2021), white-tailed deer (Pickering et al., 2022), pet hamsters (Yen et al., 2022) or a cat (Sila et al., 2022) to humans, they do not appear to represent the primary zoonotic introduction of SARS-CoV-2 into humans. Efforts to identify the animal source of the SARS-CoV-2 progenitor remain unsuccessful until today (Lytras et al., 2021).

1.4 OPEN SCIENTIFIC QUESTIONS

The awareness that zoonoses emerging from wildlife are a growing global health threat has made the prevention of their emergence a public health priority. Driven by the focus on identifying new zoonotic pathogens together with the routine use of next-generation sequencing, mass screening of animal samples has led to an explosive expansion of a previously unrecognized viral diversity in animals. However, most of these sequences have rarely been exploited in detail at the molecular level, even though their diverse genomes may offer unprecedented possibilities to unravel fundamental biological properties defining these unique pathogens. Analysing and comparing their diverse genomic sequences could enable revisiting the evolutionary history of human viruses, help to identify genomic determinants that enabled interspecies transmission, and allow a detailed risk assessment of their zoonotic potential (Box 1).

For example, little is known to date about the extent to which HAV-specific mechanisms that the virus uses to evade the human immune system are evolutionarily conserved in animal hepatoviruses. This information could provide fundamental evidence as to whether and to what extent these mechanisms may have enabled past host shifts and thus influenced the evolution of hepatoviruses and their hosts.

In the case of SARS-CoV-2, much-needed evidence about its origin or zoonotic source is still lacking, but information seems crucial to obtain a better understanding of its sudden emergence. The process by which this virus emerged is complex and poorly understood but the genomic sequences of a large diversity of SrC may help elucidate what genetic adaptations or genomic determinants helped the virus to switch hosts into humans. Likewise important for elucidating the evolution of SARS-CoV-2 is the close monitoring of the in-host adaptation after host switching. Analysis of the emergence of new strains with potentially altered characteristics and identification of mutations that change the viral phenotype are important key elements of virus surveillance.

Box 1. Open scientific questions addressed in this thesis**HAV:**

- Which genomic determinants contributed to hepatovirus host switching?
- Are biological mechanisms that are used by HAV to adapt to and evade from the host immune response evolutionary conserved in animal hepatoviruses?
- Which interactions between hepatoviruses and their hosts have shaped their evolution and how?

SARS-CoV-2:

- Which genetic mutations does SARS-CoV-2 acquire in a geographic region with high transmission rates and which lineages circulate predominately?
- What is the origin of SARS-CoV-2 and which molecular features facilitated the host switch into humans?
- Do reservoir-bound SARS-related coronaviruses harbour furin cleavage sites or can they acquire them via naturally conserved mechanisms?

CHAPTER 2

EVOLUTIONARY CONSERVATION OF IMMUNE ESCAPE MECHANISMS IN HEPATOVIRUSES¹

2.1 INTRODUCTION

HAV remains a common cause of acute viral hepatitis globally despite circulating in the human population for over 10,000 years (Drexler et al., 2015) and causing lifelong immunity (Normann et al., 2004). The continued presence of this hepatotropic virus suggests that HAV has evolved potent adaptation strategies to resist elimination by their host cell immune systems.

To protect and fight against pathogens, vertebrates have evolved highly sophisticated and effective immune systems which are classically divided into innate (non-specific) and acquired (antigen-specific) immunity (reviewed in (Boehm, 2012)).

The adaptive immune response is highly specific and defined, and uses a mechanism known as “immunological memory”. It is mainly driven by cytotoxic T- and antibody-producing B-lymphocytes whose processing and maturation requires an initial encounter with the antigen in order to induce antigen-specific antibody-mediated neutralization.

¹ Parts of this chapter have been previously published in:

de Oliveira Carneiro I, Sander AL, Silva N, Moreira-Soto A, Normann A, Flehmig B, et al. A Novel Marsupial Hepatitis A Virus Corroborates Complex Evolutionary Patterns Shaping the Genus *Hepatitis A Virus*. *J Virol*. 2018;92(13).

Sander AL, Corman VM, Lukashev AN, Drexler JF. Evolutionary Origins of Enteric Hepatitis Viruses. *Cold Spring Harb Perspect Med*. 2018;8(12).

Feng H, Sander AL, Moreira-Soto A, Yamane D, Drexler JF, Lemon SM. Hepatitis A Virus 3ABC proteases and evolution of mitochondrial antiviral signaling protein (MAVS). *Journal of hepatology*. 2019;71(1):25-34.

Copyright permissions are granted to the authors of the publications.

Therefore, this immune response only begins acting in the late phase of an infection or after reinfection with the same pathogen.

In contrast to the adaptive immune response, the innate immune system is rather unspecific but can immediately interfere when a pathogen invades. Next to anatomical or chemical barriers (skin, gastrointestinal tract, saliva, etc), it includes a cellular (leucocytes) and a humoral component (complement system, interleukins) aiming at identifying, attacking, and eliminating invading pathogens at the first stage. Its processes are initiated by the recognition of “non-self” molecules, so-called pathogen-associated molecular patterns (PAMPs) by pattern recognition receptors (PRRs), the latter being present either on the cell surface or within different intracellular compartments of the immune cells (reviewed in (Iwasaki and Medzhitov, 2015)). The recognition of PAMPs, which are predominately nucleic acids, then initiates antiviral signal cascades resulting in the activation of different transcription factors leading to interferon (IFN) expression (reviewed in (Bowie and Unterholzner, 2008)). IFN secretion by the cell induces hundreds of IFN-stimulated genes which in turn directly interfere with viral replication and prohibit virus spread.

During the infection with HAV, the innate immune response is primarily initiated within hepatocytes of the liver, which express different types of PRRs, including retinoic acid activated gene I (RIG-I)- like RNA helicases (RLRs) and Toll-like receptors (TLRs) (Li et al., 2005a) (reviewed in (Feng and Lemon, 2019)). Toll-like receptor 3 (TLR3) for instance detects viral double-stranded RNA (dsRNA), an intermediate product during negative-strand synthesis, in the endosome and activates downstream signalling cascades of nuclear factor (NF)- κ B and interferon regulatory factor 3 (IRF3) leading to IFN type I and III secretion by the cell. Additionally, the two RLRs, RIG-I and melanoma differentiation-associated protein-5 (MDA-5), which are both located in the cytosol, sense dsRNA as well as single-stranded virus RNA (ssRNA) (Pandey et al., 2014). Upon binding, both receptors interact with the adaptor MAVS protein, which activates induction of the transcription factors NF- κ B and IRF3 and subsequent type I and III IFN production.

HAV infection causes acute and short-termed injury in the liver of humans and chimpanzees but induces only minimal IFN response (reviewed in (Lanford et al., 2011)). Several genomic attributes of HAV, such as linkage of RNA 5' ends by a small protein named VPg or poly(A)-tailing of HAV genomic 3' ends to protect newly synthesized RNA from PRR recognition, have been identified to contribute to this stealthy nature (reviewed in (Feng and Lemon, 2019)).

Yet another, and one of the most distinctive features of HAV is a marked codon usage bias (CUB) (Pinto et al., 2007), reflected as one of the lowest effective number of codons (ENC) known of RNA viruses (Belalov and Lukashev, 2013; Pinto et al., 2018). The causes for a biased codon usage are diverse, but the influence of translational- as well as selection pressure are likely major factors (Bulmer, 1987). The former means preferences to use rare transfer RNAs to speed up the translation rate. In HAV, this may be particularly the case because of the virus's inability to shut off host protein synthesis compared to other picornaviruses (Ali et al., 2001; Aragonés et al., 2010). However, it is suggested that the underlying reasons for a biased codon usage are to a larger extent a product of mutation pressure on the genomic nucleotide and dinucleotide content (Belalov and Lukashev, 2013; Jenkins and Holmes, 2003). For example, the HAV genome shows very low CpG and UpA dinucleotide contents (Karlin et al., 1994; Rima and McFerran, 1997; Sander et al., 2018) which are underrepresented in many mammalian RNA viruses, but to a variable extent (Belalov and Lukashev, 2013; Jenkins and Holmes, 2003). This decreased frequency of these dinucleotides may be associated with the fact that these dinucleotides are suspected to be sensed by PAMP receptors (Atkinson et al., 2014; Lester and Li, 2014). To avoid triggering host cell innate immune responses, HAV thus may have decreased the frequency of these dinucleotides, which indirectly influences codon usage (Pinto et al., 2007). Indeed, depletion of CpG was attributed to pressure from a recently discovered zinc-finger antiviral protein (ZAP) (Takata et al., 2017), while UpA may be a target of anti-viral RNase L (Player and Torrence, 1998).

In addition to these mechanisms that are responsible for restricted immune recognition, HAV has developed strategies to actively disrupt the innate immune response of its host (Fensterl et al., 2005; Yang et al., 2007). For example, the virally encoded 3C^{pro} protease intermediate, 3CD, cleaves the TLR3 adaptor protein TRIF in the signalling pathway and thereby interrupts the IFN production in the endosomal compartment (Qu et al., 2011). Furthermore, HAV disrupts innate immune response in its host by cleaving the MAVS protein, an important mediator for signal transduction within the cytosolic cascade of type-I interferon immune response by another precursor of the virally encoded 3C^{pro} protease, named 3ABC protease (Yang et al., 2007).

2.2 OUTLINE AND AIMS OF THIS CHAPTER

The availability of full genome sequences of highly diverse non-primate hepatoviruses offers important opportunities for genomic characterisation as well as studies on evolutionary relationships (Anthony et al., 2015; de Oliveira Carneiro et al., 2018; Drexler et al., 2015; Yu et al., 2016). Phylogenetic analyses suggest multiple past host shifts during the evolution of these hepatoviruses (de Oliveira Carneiro et al., 2018; Drexler et al., 2015). The determinants of such host shifts are not understood in detail, but to overcome the innate immune response, a host defence mechanism against invading pathogens, is likely one of the main obstacles. Like HAV, non-primate hepatoviruses are hepatotropic and share antigenic and genomic properties with HAV (de Oliveira Carneiro et al., 2018; Drexler et al., 2015; Sander et al., 2018). However, while multiple studies have been conducted on how HAV elicits an attenuated immune response, these mechanisms are poorly understood for hepatoviruses from small mammals.

In this chapter, I use available full genome sequences of highly diverse hepatoviruses from different mammalian species to investigate the evolutionary conservation of biological mechanisms that are used by HAV to adapt to and evade its host immune response.

First, I analyse the codon usage of animal hepatoviruses in general and that of a hepatovirus from marsupials in particular (Chapter 2.4.1) to identify if a biased codon usage is a conserved genomic feature among the viral genus *Hepatovirus*. In the following part, I perform pressure analyses on MAVS orthologs from bat, rodent and primate species (Chapter 2.4.2). As the cleavage of MAVS proteins seems to be a conserved immune evasion strategy among different hepatoviruses, I aim at identifying if hepatovirus infections in different mammalian hosts have placed MAVS proteins under evolutionary pressure.

The results of this chapter enable an advanced understanding of how these host-pathogen interactions have shaped the evolution of hepatoviruses and their hosts and how they might have been involved in cross-species transmissions.

2.3 METHODS

2.3.1 Codon usage analyses of non-primate hepatoviruses

A GenBank search with the term “*Hepatovirus*” was performed on August 3, 2017, and all sequences longer than 6000 nucleotides were selected. Duplicates, cell culture-adapted strains, or viruses isolated from experimentally infected animals were excluded from the dataset, resulting in 124 final sequences. HAV genotype (gt)IV and gtVI are defined only by partial sequence information and were thus not included in the analyses (Nainan et al., 1991; Robertson et al., 1992).

Effective number of codons and relative CpG and UpA contents within the full polyprotein sequence of primate and non-primate hepatoviruses were determined using SSE V1.2 (Simmonds, 2012). GenBank accession numbers (Acc. No.) of hepatovirus sequences used for analyses are listed in **Supplementary Table S1**.

2.3.2 Codon usage analyses of a marsupial hepatovirus

Host codon usage was retrieved from the online database HIVE-CUT (Athey et al., 2017). Opossums were represented by *Monodelphis domestica* (id35487) in HIVE-CUT. Primate HAV sequences included gtIa and Ib (Acc. No. AB020564, AF268396), gtIIa and IIb (Acc. No. AY644676, AY644670), gtIIIa and IIIb (Acc. No. AB279732, AB279735) and gtV (Acc. No. D00924). The full genome of MHAV was submitted to GenBank under Acc. No. MG181943. The codon usage and relative CpG content of viruses were determined using SSE V1.2. The arginine index (AI) was calculated as the ratio of genomic CGN/(CGN + AGR) codons.

2.3.3 MAVS gene amplification

Bat (*Eidolon helvum*, *Hypsignathus monstrosus*, *Pipistrellus pipistrellus*, *Rhinolophus ferrumequinum*) and rodent (*Myodes glareolus* and *Sigmodon hispidus*) cell lines available at the institute were used for total RNA extraction using the RNeasy kit. Primers for nested- or heminested-assays covering approximately 300-900 base pairs each were designed based on MAVS sequence information of the same orders available in GenBank or the subsequently obtained Sanger sequences. Briefly, RNA was reverse transcribed for 20 min at 55°C using the SSIII One-Step Kit (Thermo Fisher) followed by 45 PCR cycles of 94°C for 15 seconds, 63°C for 15 seconds and 72°C for 30 seconds. The 2nd round PCR was performed at the same

conditions as the 1st round without reverse transcription. Gene termini were confirmed using a rapid amplification of cDNA ends (RACE) strategy based on order- or species-specific nested forward or reverse primer sets and the kit-included Oligo(dT)-Anchor primer. Primer sequence information are listed in the **Supplementary Material**.

2.3.4 *Sanger sequencing*

Sanger sequencing of purified PCR products was outsourced to MicroSynth Seqab in Göttingen. Sequencing primers (500 nM final concentration) and PCR products were provided according to the company's recommendations, with 4-30 ng/μl final PCR product concentration in 12 μl total volume. Ab1 files with obtained sequencing reads were mapped to their respective template in Geneious v 9.1.8.

2.3.5 *MAVS sequence comparison*

MAVS amino acid sequence distances were calculated with MEGA7 (Kumar et al., 2016) using the pairwise deletion option and all species per order for which MAVS sequences were available. MAVS pairwise distances of species were normalized by mean cytochrome b pairwise distance per order. Host cytochrome B sequences were obtained from GenBank.

MAVS sequence identities were plotted with SSE V1.2. program (Simmonds, 2012) using a fragment length of 200 and a step size of 100 amino acid residues. Alignment gaps were excluded from the analysis.

GenBank accession numbers of used MAVS and cytochrome b sequences are listed in **Supplementary Table S2**.

2.3.6 *MAVS pressure analyses*

Pressure analyses were performed using translation alignments of all available MAVS sequences from the 3 different orders (Primates, Rodentia and Chiroptera) that were previously generated in Geneious 6.1.8.

dN/dS ratios were calculated using six methods: fixed effects likelihood (FEL), single-likelihood ancestor counting (SLAC), random effects likelihood (REL), partitioning approach for robust inference of selection (PARRIS), MEGA7 and model M0 in CodeML

To investigate positive selection pressure in different MAVS datasets, the CodeML program implemented in the pamlX 1.3.1 software package (Xu and Yang, 2013; Yang, 1997) was used. Codon substitution models M0, M1a, M2a, M3, M7 and M8 were tested using the codon frequency model F61. Likelihood ratio tests (LRTs) on 2 pairs of site-specific models, M1a vs. M2a and M7 vs. M8, were performed to determine whether MAVS genes from each of the 3 orders are evolving under positive selection using the chi-square calculator implemented in the PAMLX package, a user interface for phylogenetic analysis by maximum likelihood (PAML). Fixed effect likelihood (FEL), single likelihood ancestor counting (SLAC), and random effect likelihood (REL) (Kosakovsky Pond and Frost, 2005) and single bayesian approximation (FUBAR) (Murrell et al., 2013) from the HyPhy package (Datamonkey.org) were also employed to detect sites under selection pressure using the HKY85 substitution mode, as suggested by automatic model selection in Datamonkey. The branch-site unrestricted statistical test for episodic diversification (BUSTED) was used to detect gene-wide positive selection in the MAVS datasets.

2.3.7 *Statistical analyses*

Statistical comparisons were carried out by 2-sided t test using Prism 6.0 (GraphPad Software, Inc.).

2.4 RESULTS

2.4.1 Codon usage of non-primate hepatoviruses

Codon usage bias (CUB) is the tendency of an organism to preferably use specific codons over other synonymous codons to encode one and the same amino acid (Grantham et al., 1980). In HAV, CUB has been attributed to a specific replication strategy aimed at avoiding virus protein overexpression triggering an immune response (Pinto et al., 2007). Whether this replication strategy is an evolutionary conserved attribute among hepatoviruses from animal hosts has never been investigated.

In order to compare codon usage of hepatoviruses from primate and non-primate hosts, a dataset of 122 hepatovirus sequences was analysed (chapter 2.3.1). 104 sequences of the dataset were from humans and four further species of non-human primates (*Homo sapiens*, *Chlorocebus aethiops*, *Macaca mulatta*, *Pan troglodytes* and *Papio anubis*), whereas 18 sequences were from 14 different non-primate hosts of the host orders Carnivora (*Phoca vitulina*), Chiroptera (*Eidolon helvum*, *Rhinolophus landeri*, *Coleura afra*, *Miniopterus cf. manavi*), Eulipotyphla (*Sorex araneus*, *Erinaceus europaeus*), Rodentia (*Marmota himalayana*, *Microtus arvalis*, *Myodes glareolus*, *Cricetulus migratorius*, *Sigmodon mascotensis*, *Lophuromys sikapusi*) and Scandentia (*Tupaia belangeri chinensis*).

One of the parameters that indicate the extent of CUB is the effective number of codons (ENC) (Wright, 1990). The highest possible value of this number is 61 if all codons are used equally, and the lowest possible is 20 if only one codon per amino acid is used. (Wright, 1990). According to previous studies, an organism shows evidence of strong CUB if its ENC value is below 40 (Jiang et al., 2008; Wright, 1990; Zhao et al., 2016). The ENC values of RNA viruses range between 38 and 58 (Belalov and Lukashev, 2013; Jenkins and Holmes, 2003) thus most RNA viruses show moderate to low CUB. Consistent with previous studies, the ENC value of HAV and hepatoviruses from primate hosts, was around 40 (**Figure 4A**) (Belalov and Lukashev, 2013; Zhang et al., 2011). This characteristic of the HAV genome was conserved within the expanded genus *Hepatovirus*, since all non-primate hepatoviruses showed relatively low ENC values, even though the variation was greater than among primate HAVs (**Figure 4A**).

Besides the ENC value, the avoidance of CpG and UpA dinucleotides presents another feature complementing to CUB and was next tested. These dinucleotides are underrepresented in many

mammalian RNA viruses, but to a variable extent (Belalov and Lukashev, 2013; Jenkins and Holmes, 2003). Depletion of CpG was attributed to pressure from a recently discovered zinc-finger antiviral protein (ZAP) (Takata et al., 2017), whereas UpA may be a target of antiviral ribonuclease (RNase) L (Player and Torrence, 1998). HAV is known to have an extremely low relative CpG content (even after correction for genomic C and G content), typically between 0.1 and 0.16 of the statistically expected value. The avoidance of CpG dinucleotides was conserved within non-primate hepatoviruses, although slightly less pronounced than in HAV (median 0.2, maximum 0.33) (**Figure 4B**). UpA dinucleotides were moderately avoided in both HAV and non-primate hepatoviruses (**Figure 4B**). Therefore, codon usage bias and the dinucleotide content bias, putatively linked in interaction with innate immune responses, seem to be similar in all hepatoviruses.

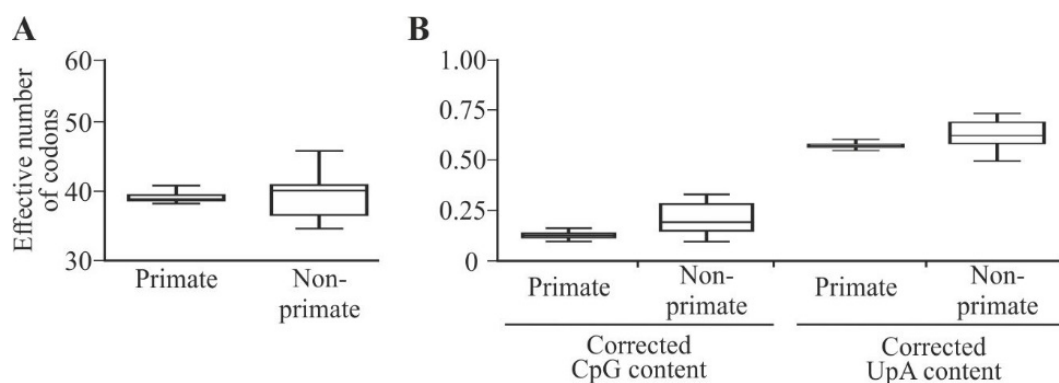


Figure 4. Effective number of codons and relative CpG and UpA dinucleotide content in primate and non-primate hepatoviruses. Median (bar) and quartiles (box and whiskers) are shown. Accession numbers of included sequences are listed in Supplementary Table S1. Adapted from (Sander et al., 2018).

Marsupials are ancient mammals and viruses from these animals may thus provide important insights into virus evolution. The discovery of a novel marsupial hepatovirus (MHAV) in the Brazilian common opossum (*Didelphis aurita*) (de Oliveira Carneiro et al., 2018), thus offered the opportunity to investigate codon usage of a hepatovirus found in one of the oldest mammalian orders.

Genome sequences of MHAV and HAV were analysed in terms of amino acid codon usage as well as arginin index, ENC value and CpG dinucleotide usage (chapter 2.3.2). As shown in **Figure 5A**, codon usages were generally comparable between HAV and MHAV and between their human and marsupial hosts. The CUBs of HAV and MHAV were very similar as evidenced by comparable ENC counts and arginine indexes (**Figure. 5B**). Rather than a preferred usage of codons that were underrepresented in their hosts, HAV and MHAV shared a pronounced avoidance of codons containing a CpG dinucleotide. In HAV and MHAV, the

CpG contents were similarly low, at 0.111 and 0.115 of the expected content given random dinucleotide composition at the actual frequencies of genomic C and G nucleotides.

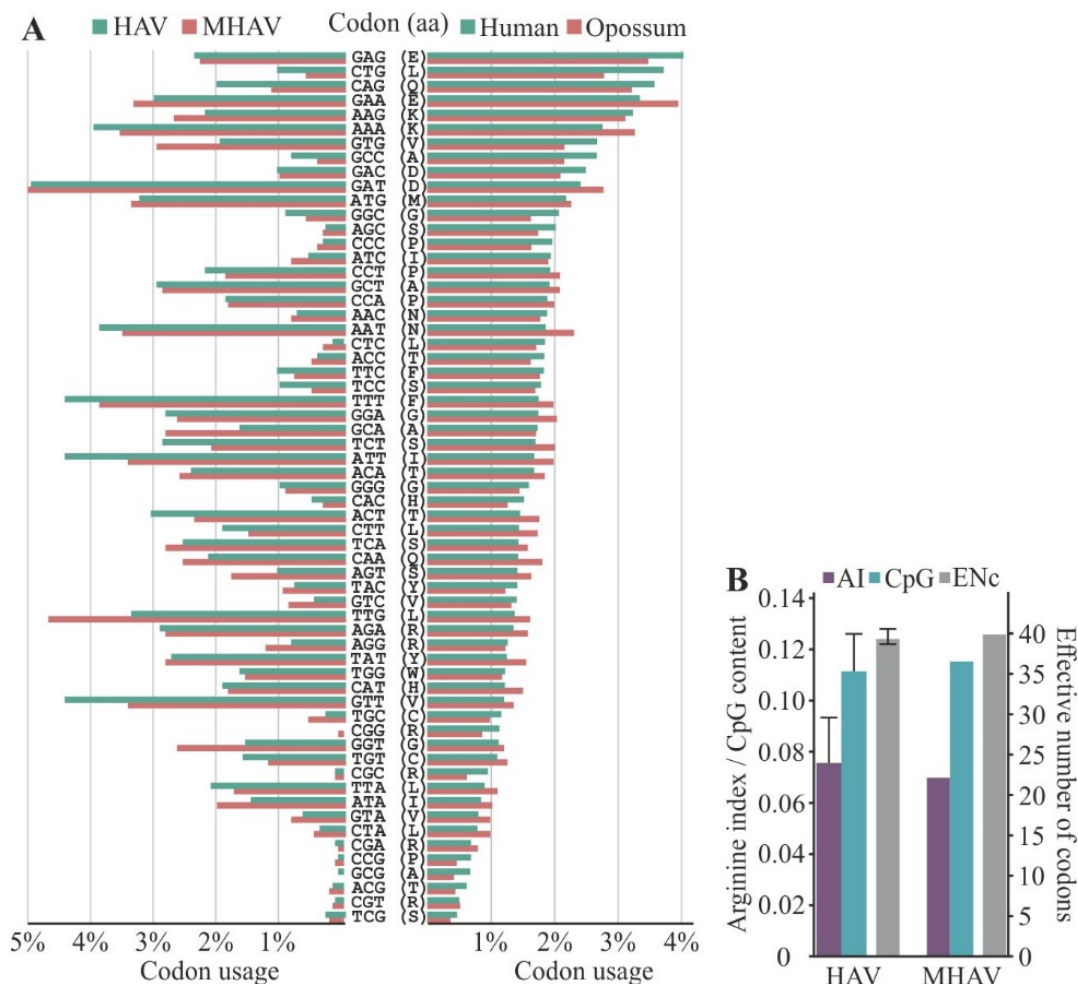


Figure 5. Codon usage bias of HAV and MHAV. (A) Codon usages of HAV and MHAV compared to those of human and marsupial hosts. Stop codons are not shown for clarity of presentation. aa, amino acid residue. (B) AIs, corrected CpG dinucleotide contents, and effective numbers of codons (ENC) of HAV and MHAV. Error bars for HAV indicate the ranges of values among HAV genotypes. Adapted from (de Oliveira Carneiro et al., 2018).

2.4.2 *Hepatitis virus 3ABC proteases and evolution of mitochondrial antiviral signalling protein (MAVS)*

Mammals diversified rapidly following the Cretaceous–Palaeogene mass extinction event ~66 million years ago resulting in ~6,400 extant mammalian species living by today (Burgin et al., 2018). Two-thirds of these species belong to the orders Rodentia and Chiroptera (Burgin et al., 2018), which host the greatest diversity of modern hepatoviruses (Drexler et al., 2015). Phylogenetic reconstructions of these viruses indicate that their evolution has involved multiple past host shifts (Drexler et al., 2015). Although the determinants of viral cross-species

transmission are not fully understood, the need to evade innate immune surveillance in a new host represents a substantial biological barrier that must be overcome. HAV does this by cleavage of the MAVS protein with its 3ABC protease (Yang et al., 2007). Experimental evidence indicated that bat and rodent MAVS proteins are relatively resistant to cleavage by their cognate 3ABC proteases. However, the bat viral proteases retain the ability to cleave human MAVS (*hsMAVS*) but at an alternative cleavage site, namely Glu⁴⁶³/Gly⁴⁶⁴ (**Figure 6A**) (Feng et al., 2019). These results suggest that the genes encoding MAVS proteins expressed by bats and rodents, which have likely hosted hepatoviruses over millions of years, have evolved under genetic diversifying selection. To test for this hypothesis, a MAVS coding sequence dataset from primate, bat and rodent species was collected (chapter 2.3.5). In genomic databases, MAVS sequences were available from 11 bat and 19 rodent species, however, none of these species was a known host of hepatoviruses. Therefore, MAVS coding sequences from additional four bat (*E. helvum*, *H. monstrosus*, *P. pipistrellus*, *R. ferrumequinum*) and two rodent (*M. glareolus* and *S. hispidus*) hepatovirus host species were amplified (chapter 2.3.3.) and characterized, providing a total of 75 MAVS sequences for analysis (**Supplementary Table S2**).

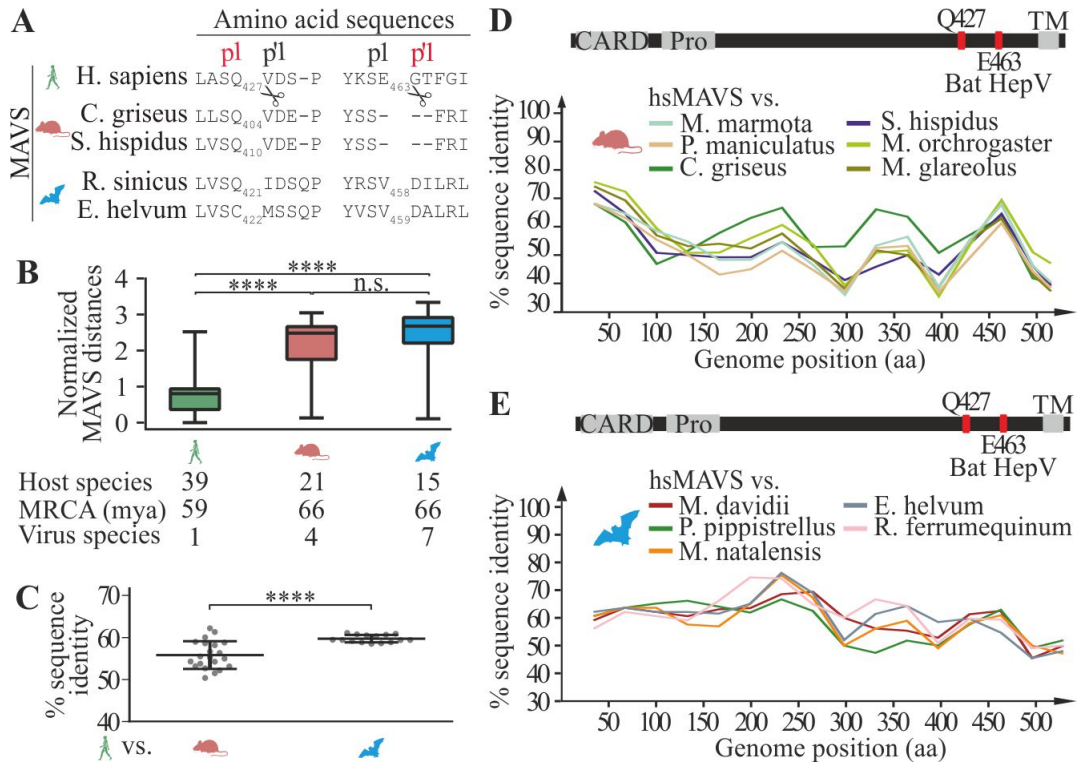


Figure 6. MAVS sequence comparison and evidence for episodic diversifying selection in MAVS. (A) Alignments of human, rodent and bat MAVS amino acid sequences at 3ABC cleavage sites in hsMAVS. (B) Cytochrome b-normalized MAVS mean pairwise distances per host order. (C) Mean pairwise sequence identities \pm SD between *hsMAVS* and rodent or bat orthologs. **** $p < 0.0001$ by t test. (D+E) Amino acid sequence identities along the length of *hsMAVS* with (D) rodent and (E) chiropteran MAVS orthologs. CARD, CARD-domain; *hsMAVS*, human MAVS; MRCA (mya), most recent common ancestor of orders (millions of years ago); Pro, proline-rich region; TM, transmembrane domain. Adapted from (Feng et al., 2019).

2.4.2.1 MAVS diversity among different host orders

First, MAVS sequence comparison was performed (chapter 2.3.5.1) in order to identify diversified regions within primate, bat and rodent MAVS genes. Although primates were overrepresented in this dataset, MAVS sequences from Chiroptera and Rodentia were significantly more diversified than those from primates (**Figure 6B**). Surprisingly, MAVS diversity was not related to overall host genetic relationships. Although rodents and primates are genetically closer than bats and primates (Foley et al., 2016), *hsMAVS* shared less identity with rodent than with bat MAVS (**Figure 6C**). Sequence distances were unevenly distributed across the MAVS reading frame (**Figure 6D and E**) with generally between 40-80% sequence identities.

2.4.2.2 Selection pressure analyses of MAVS genes

The results up to this point suggest that MAVS genes from distinct mammalian orders in general, and some regions of the MAVS gene in particular, may have been shaped by differential evolutionary forces, likely evolutionary pressure.

To assess gene-wide evidence for diversifying selection in primate, rodent and chiropteran MAVS, the ratio (ω) of non-synonymous (dN) and synonymous sites (dS) of the three datasets were calculated using six different methods (chapter 1.3.5.2). Regardless of the method used for calculation, ω was less than unity across the entire MAVS gene in all three orders (**Figure 7A**) which does not corroborate evolution of MAVS under diversifying selection. However, significant episodic diversifying selection was evident in primate, rodent and chiropteran MAVS ($p < 0.001$) according to both likelihood ratio tests (**Table 1**) and a branch-site unrestricted statistical test (BUSTED) (Murrell et al., 2015).

These incongruencies may result from the lack of power of dN/dS values to detect positive selection within closely related sequences (Kryazhimskiy and Plotkin, 2008). Furthermore, positive selection often affects only specific residues in a gene (Holmes et al., 1992; Hughes and Yeager, 1998) which the applied methods are not designed to detect. To address these limitations, the MAVS datasets were additionally tested using different models detecting site-specific selection, including Maximum-likelihood (FEL, SLAC and PAML) and Bayesian (REL and FUBAR) methods (Kosakovsky Pond and Frost, 2005; Murrell et al., 2013). In total, 90 positively selected residues, mostly in bats and rodents were identified across the entire MAVS gene (**Figure 7B, Supplementary Table S3**). Notably, 41 of these (45.6%) aligned in a contiguous fashion with other residues under positive selection. Fourteen of the 17 (87.5%) contiguous peptide segments thus identified had evidence of diversifying selection in at least 2 different orders, suggesting similar evolutionary pressures shaping MAVS across different orders.

Table 1. Likelihood ratio tests for selected PAML site-specific models.

Selected site models	$\ln L_0$	$\ln L_1$	$2\Delta\ln L$	df	<i>p</i> -value
Primates					
M1a _(nearly neutral) vs. M2a _(pos. selection)	-9125.10	-9113.53	23.13	2	<i>p</i> <0.0001
M7 _(beta) vs. M8 _(beta&omega)	-9131.50	-9113.80	35.40	2	<i>p</i> <0.0001
Rodentia					
M1a _(nearly neutral) vs. M2a _(pos. selection)	-15457.26	-15448.28	17.95	2	<i>p</i> <0.001
M7 _(beta) vs. M8 _(beta&omega)	-15448.11	-15433.70	28.82	2	<i>p</i> <0.0001
Chiroptera					
M1a _(nearly neutral) vs. M2a _(pos. selection)	-8003.66	-7983.35	40.62	2	<i>p</i> <0.0001
M7 _(beta) vs. M8 _(beta&omega)	-8009.67	-7981.22	56.90	2	<i>p</i> <0.0001

Evidence of positive selection was sought by comparing site models that allow $x > 1$ (M2a, M8) with models that disallow positive selection (M1a, M7). df, degree of freedom; $\ln L$, log-likelihood scores; PAML, phylogenetic analysis by maximum likelihood.

Using any of the 5 analytic methods, no evidence was found for diversifying selection at the site targeted by the human HAV protease, Gln⁴²⁷/Val⁴²⁸, among MAVS genes expressed by primates (**Figure 7B, Supplementary Table S3**). In contrast to this finding and as reported previously (Patel et al., 2012), PAML detected a signal for selection at residues aligning with Cys⁵⁰⁸/His⁵⁰⁹, where *hs*MAVS is cleaved by the HCV NS3/4A protease (Li et al., 2005c; Meylan et al., 2005). Consistent with the fact that rodents and bats also host diverse hepaciviruses (Drexler et al., 2013; Quan et al., 2013), evidence for diversifying selection was also noticed at residues aligning with this site in MAVS orthologs expressed by these two host orders (**Figure 7B, Supplementary Table S3**). Despite the absence of diversifying selection at the HAV 3ABC cleavage site, Gln⁴²⁷/Val⁴²⁸, among MAVS genes expressed by primates, a selection signal at MAVS residues aligning with this site in bats using PAML, and a strong signal for selection at this site in rodents was detected using multiple tests for genetic pressure (**Figure 7B, Supplementary Table S3**). Just as striking, however, was a very strong signal for diversifying selection among chiropteran MAVS orthologs at residues aligning with Lys⁴⁶¹ and Thr⁴⁶⁵ (Feng et al., 2019), which bracket the site of *hs*MAVS cleavage by the bat viral proteases (**Figure 7B, Supplementary Table S3**).

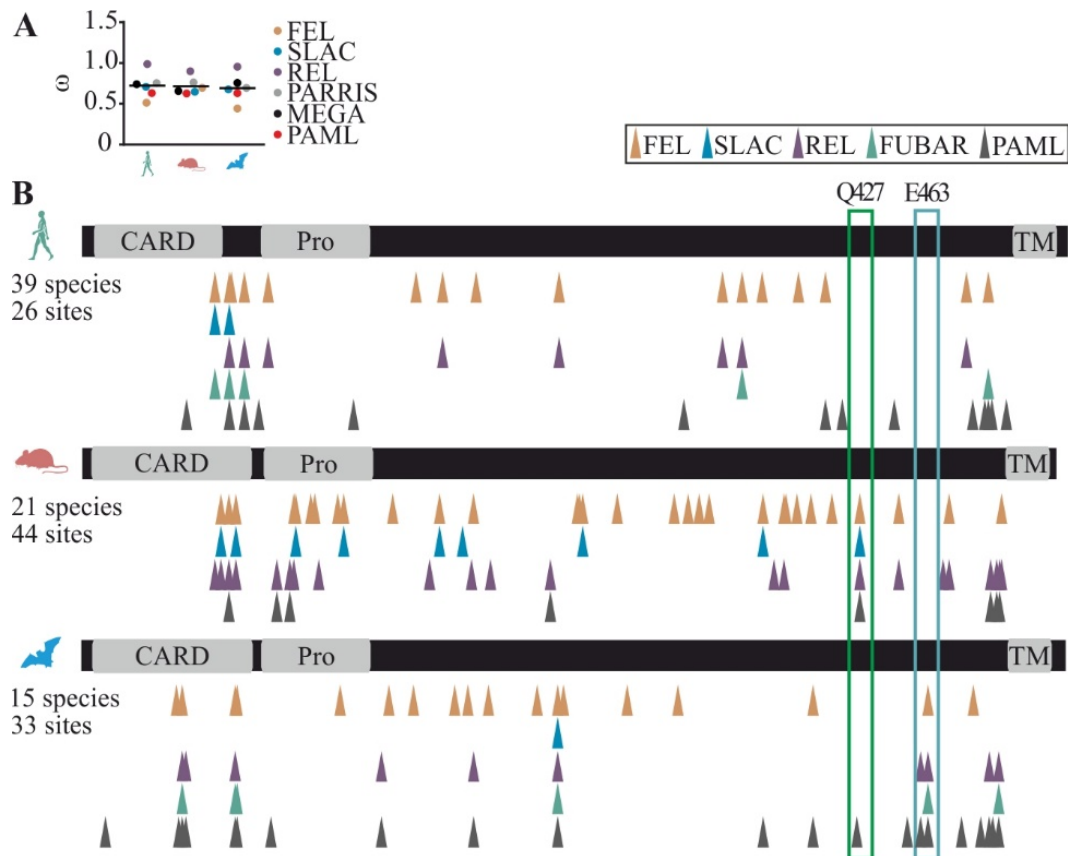


Figure 7. Episodic diversifying selection in mammalian MAVS genes. (A) Average dN/dS (ω) ratios of primate, rodent and chiropteran MAVS sequences calculated using indicated methods. M0 model was used in CodeML. FEL, SLAC, REL and PARRIS were executed in the HyPhy package on Datamonkey.org. The HKY85 substitution model was used for all methods and datasets as suggested from the automatic model selection in Datamonkey. Codon frequencies according to the F61 model were used for CodeML analyses. Horizontal line, mean dN/dS. (B) Amino acid residues identified under positive selection within primate, chiropteran and rodent MAVS. Triangles mark sites with signals for positive selection in at least 1 of the 5 methods. See Supplementary Table S3 for details. dN/dS, non-synonymous/synonymous; CARD, CARD-domain; Pro, proline-rich region; TM, transmembrane domain. Adapted from (Feng et al., 2019).

2.5 DISCUSSION

The present comprehensive genomic analyses of non-primate hepatoviruses reveal novel insights into the evolution of hepatoviruses in general and the fundamental biological properties of these viruses in particular.

Unlike hepaciviruses such as hepatitis C virus (HCV) and hepadnaviruses such as hepatitis B virus (HBV), HAV does not usually cause long-lasting infection and therefore requires a sustained high rate of transmission between individuals to maintain its presence in human populations. Potentially due to this reason, HAV likely became established among humans only with the development of population densities sufficient to support continuous chains of fecal-oral transmission while HBV and HCV appear to have infected primates for millions of years (de Carvalho Dominguez Souza et al., 2018; Littlejohn et al., 2016; Patel et al., 2012). For example, divergent hepaciviruses have been found in both New World tamarins and Old World colobus monkeys (Rasche et al., 2019b), two non-human primates that have been separated by millions of years of evolution. That divergent hepaciviruses have evolved over eons in primates explains why there is evidence of evolutionary pressure at the site of the HCV NS3/4A cleavage in primate MAVS orthologues (Patel et al., 2012). The results of chapter 2.4.2.2 confirmed this finding, while finding no pressure at the site of HAV 3ABC cleavage site (**Figure 7B**), which is consistent with the more recent establishment of HAV among primate hosts. Nonetheless, divergent genotypes of HAV (gtIV-VI) have also been recovered from different non-human primates (Lemon et al., 1982; Nainan et al., 1991; Tsarev et al., 1991). Even though most of them are defined only by partial sequence information, the fully sequenced AGM27 strain (gtV) recovered from an Old World monkey, is biologically and even antigenically distinct from human HAV (Tsarev et al., 1991) and therefore deserves further investigation.

Experimental evidence indicates that MAVS signalling is a key barrier that must be overcome for HAV to successfully infect mice (Hirai-Yuki et al., 2016). Consistent with this, the results of chapter 2.4.2.2 showed a strong signal for diversifying selection among rodent MAVS orthologs at residues aligning with the site of 3ABC cleavage in *hsMAVS*, Gln⁴²⁷/Val⁴²⁸ (**Figure 7B, Supplementary Table S3**). In contrast, diversifying selection among chiropteran MAVS was found to have occurred primarily at an alternative site in the gene that aligns with the site of *hsMAVS* cleavage by bat viral proteases: Glu⁴⁶³/Gly⁴⁶⁴. This site is highly divergent in bat MAVS, which is consistent with the inability of the bat viral proteases to cleave their cognate MAVS proteins (Feng et al., 2019). Furthermore, these data are consistent with

phylogenetic analyses indicating that hepatoviruses have existed in large and gregarious bat populations for much longer than in the human population, perhaps millions of years (Drexler et al., 2015; Sander et al., 2018), allowing more time for selective pressure on MAVS and the evolution of 3ABC resistant MAVS proteins. A lack of diversifying selection at sites of HAV 3ABC cleavage in *hsMAVS* may also be due to the limited pathogenicity HAV infections usually cause in humans. Although hepatitis A is rarely lethal in humans, and long-term, persistent HAV infection has never been described, the consequences of infection in bats, in which our data suggest the possibility of escape from 3ABC cleavage, remain undetermined. It is clear, however, that hepatoviruses use also alternative strategies to persistently infect their hosts.

For instance, codon usage analyses (chapter 2.4.1) indicate that a biased codon usage due to CpG avoidance is a conserved evolutionary mechanism used by primate and non-primate hepatoviruses to temper the host immune response. As CUB may largely result from selective pressure against CpG dinucleotides (Belalov and Lukashev, 2013), which were recently shown to be a main target of ZAP (Takata et al., 2017), it is therefore likely that HAV and non-primate hepatoviruses are under similar selective pressure, hypothetically mediated by ZAP.

Furthermore, recent research revealed that a variety of classically considered non-enveloped viruses, including HAV; circulate in the blood in a second type of particle form that are cloaked in host-derived membranes, thereby protecting the virion from antibody recognition. A very recently published study now provided first evidence, that this quasi-envelopment is also evolutionary conserved among animal hepatoviruses (Shirasaki et al., 2022). There is limited knowledge about the underlying reasons for this process, but it was previously speculated, that a C-terminal extension of the VP1 capsid protein, namely pX, is involved. In some of the recently found bat hepatoviruses pX could not be unambiguously identified (Drexler et al., 2015; Sander et al., 2018). However, other bat hepatoviruses that share pX, share as well the same endosomal sorting complexes required for transport (ESCRT)-interacting activity as pX in the human virus. These results suggest that the bat viruses share quasi-envelopment as a means of cellular egress with human HAV. However, isolation of infectious virus in cell culture from hepatoviruses other than HAV is still pending and would provide evidence for the existence of quasi-enveloped hepatoviruses in the blood of naturally infected small mammals.

In sum, the conservation of human HAV biological properties revealed in this chapter substantiates a previously suggested long evolutionary association of hepatoviruses with small mammals (Drexler et al., 2015; Sander et al., 2018). The results of comparing genomic sequences point out how animal homologs of human viruses offer unprecedented possibilities to unravel fundamental biological properties defining these unique pathogens.

Outlook:

In addition to immune evasion strategies, information on the host cell receptor usage are of indispensable importance for viral risk assessment and surveillance. Even 50 years after its discovery, the cell entry receptor of HAV remains unidentified. Prospective virological studies should focus on identifying this receptor and the repertoire of diverse mammalian hepatovirus genomes may offer fundamental insights. A prerequisite for the conduction of entry studies is the presence of infectious virus. However, none of these animal hepatoviruses has yet been isolated in infectious form. Isolation of viruses should be attempted directly from specimens of infected animals or by rescue of infectious virus from cells using *in vitro* transcripts of the complete virus genome. Since virus rescue from cell culture is not yet established for animal hepatoviruses, parameters that still need to be optimized comprise the choice of cell lines and transfection methods. Furthermore, the accumulation of cell culture-associated mutations has to be overcome (Cohen et al., 1989; Emerson et al., 1992), but genetically stable virus growth may be generated by the insertion of a blasticidin resistance gene into the hepatovirus genome as it has been shown for human HAV previously (Konduru and Kaplan, 2006).

The availability of infectious virus particles of animal hepatoviruses may additionally enable investigations into whether bat and rodent hepatoviruses cause infection patterns in their natural hosts similar to those induced by HAV in humans. Classical symptoms of HAV infections in humans are weakness, vomiting and fever, histological signs are mainly liver inflammation (reviewed in (Shin and Jeong, 2018)). However, HAV-related viruses from several different mammals show high variability in their pathogenesises in their natural hosts. For example, a HAV-like virus from seals did not cause inflammation or histopathological changes in the liver of infected animals (Anthony et al., 2015). Woodchuck hepatovirus infection caused no severe symptoms except fever in woodchucks but was associated with liver damage (Yu et al., 2018). Patterns of infection in small mammals suggested possible acute infection but no severe liver pathology was observed in liver specimens. Additionally, virus was consistently enriched in livers, except for bats in which the spleen showed equally high hepatovirus concentration (Drexler et al., 2015). The underlying reasons for this great variability in clinical outcome and severity of symptoms are not well understood. Extensive *in vitro* and *in vivo* infection studies of the recently detected hepatoviruses from small mammals could thus provide useful information to further examine hepatovirus infection and immune evasion in their natural hosts and could offer the possibility for establishing new animal models to investigate hepatovirus pathogenesis.

CHAPTER 3

SARS-COV-2 EVOLUTION IN BENIN, WEST AFRICA¹

3.1 INTRODUCTION

In principle, evolution in general can be defined as biological change over time. The term macroevolution describes the overarching evolutionary process which causes sorting of interspecies variation over the long term by culminating changes produced by a variety of patterns and processes, including microevolutionary processes (Erwin, 2000). Microevolution thus confines processes that happen over comparatively shorter time spans creating genetic variation for a species gene pool (Bell, 2009). Microevolutionary processes are driven by four main mechanisms, namely mutation, gene flow, genetic drift and natural selection. Mutations create new alleles for the gene pool, while the other evolutionary forces have significant effects on allele frequencies (Prentis et al., 2008).

Mutations are naturally produced during viral replication and give rise to new genetic variants of a virus. Unlike bacteria or parasites, which evolve with around 10^{-6} - 10^{-8} substitutions per genomic site per year (ssy) (Duchene et al., 2016) and 10^{-8} - 10^{-9} (ssy) (Sundararaman et al., 2016), respectively, the estimated substitution rates for viruses are typically at around 10^{-4} - 10^{-6} ssy (Duffy et al., 2008). In contrast to many other RNA viruses, coronaviruses show much slower mutation rates due to their exonuclease activity, which is functionally analog to proofreading DNA polymerases. It thus results that the genomic mutation rate of the newly emerged SARS-CoV-2 is estimated to be 6×10^{-4} substitutions per site per year, which corresponds to approximately two mutations per month (Wang et al., 2022).

¹ Parts of this chapter have been previously published in:

Sander AL, Yadouleton A, Moreira-Soto A, Tchiboza C, Hounkanrin G, Badou Y, et al. An Observational Laboratory-Based Assessment of SARS-CoV-2 Molecular Diagnostics in Benin, Western Africa. *mSphere*. 2021;6(1).

Sander AL, Yadouleton A, de Oliveira Filho EF, Tchiboza C, Hounkanrin G, Badou Y, et al. Mutations Associated with SARS-CoV-2 Variants of Concern, Benin, Early 2021. *Emerg Infect Dis*. 2021;27(11):2889-903.

Yadouleton A, Sander AL, Adewumi P, de Oliveira Filho EF, Tchiboza C, Hounkanrin G, et al. Emergence of SARS-CoV-2 Delta Variant, Benin, May-July 2021. *Emerg Infect Dis*. 2021;28(1).

Copyright permissions are granted to the authors of the publications.

Host switching events force pathogens to adapt to the new environment in order to survive and adaptive evolution is one key mechanism (Prentis et al., 2008). Due to its recent nature in the human population, SARS-CoV-2 is a great example to investigate and trace the dynamic process of microevolutionary adaptation to the human host after the host switch. As each replication cycle provides an opportunity for mutations to occur, intense SARS-CoV-2 transmission increases the frequency of viral mutations and thus promotes the emergence of variants (Grenfell et al., 2004). This is particularly the case in densely populated geographical regions such as Sub-Saharan Africa. Within the Sub-Saharan African region which is one of the most densely populated and underdeveloped regions globally, Benin is a small West-African country with about 12 million inhabitants (**Figure 8**).



Figure 8. Location of Benin in Africa. Africa map was generated by using the rworldmap package in R.

3.2 OUTLINE AND AIMS OF THIS CHAPTER

The true SARS-CoV-2 case numbers in Benin may be significantly higher than officially reported, as WHO suspects only 14% of the SARS-CoV-2 infections in Africa to be reported (WHO, 2021b). Undetected transmissions may lead to the emergence or circulation of new viral genetic variants with partial or complete immune escape, but genomic surveillance in West African countries is notoriously weak compared to other African countries such as South Africa (Tegally et al., 2022). However, tracing the molecular epidemiology and monitoring of new mutations in the viral genome is key to our understanding of the adaptive evolution of SARS-CoV-2 and allows for timely surveillance in terms of risk assessment.

The following chapter aimed to monitor adaptive mutations of SARS-CoV-2 over the course of approximately 1.5 years. Therefore, I first investigate the SARS-CoV-2 genomic diversity introduced into Benin by returning travellers until end of April 2020 to identify imported lineages and mutations during the beginning of the pandemic. To further monitor SARS-CoV-2 evolution, I conducted genomic surveillance one year after the start of the pandemic in order to discover predominating lineages and mutations. Finally, the last part of this chapter focuses on the epidemiology and genetic variation of the Delta variant of concern (VOC) in Benin during early 2021.

In this context, I aim at answering the following questions:

- Which SARS-CoV-2 lineages were primarily imported into Benin within the first months of the pandemic?
- Are there any predominant mutations in these Benin-derived SARS-CoV-2 sequences, and if so, which ones?
- How does the SARS-CoV-2 genomic diversity in Benin change over time?
- Which mutations occur at which frequency in circulating lineages?
- At which time points were VOCs introduced into Benin?

The results of this chapter contribute to the understanding of the microevolutionary diversification of SARS-CoV-2 and form the basis for further, per example phylodynamic or phylogenetic investigations that are needed to understand the spread and evolutionary dynamics of SARS-CoV-2.

3.3 METHODS

3.3.1 *Collection of samples*

Oro-nasopharyngeal swab samples of patients were collected from different cities in Benin between March 2020 and July 2021. Anonymized data sets were used. All analysis of personally identifiable data took place only in the Laboratoire des Fièvres Hémorragiques Virales du Benin (LFHB), Benin's central laboratory for respiratory diseases in Cotonou.

3.3.2 *Laboratory testing in Benin*

Initial SARS-CoV-2 Diagnostics of all samples took place at the LFHB, or in associated satellite laboratories in the country of Benin. Here, viral RNA was extracted from oro-nasopharyngeal swabs suspended in 140 µl of viral transport medium using the viral RNA mini kit (Qiagen, Germany) following the manufacturer's instructions. Over the course of the pandemic, a heat/proteinase-K-based RNA extraction method without usage of dedicated kits was implemented and used further on. Briefly, 50µl sample aliquots were treated with 125ug Proteinase K. The solution was vortexed at 3,200 rpm for 1 minute followed by thermal inactivation at 95°C for 5 minutes. Prior to RT-PCT testing, samples were cooled down at room temperature for a min of 2 minutes.

Samples were screened for SARS-CoV-2 RNA at the LFHB using different kits: (1) the kit donated by the Jack Ma Foundation termed 2019 Novel Coronavirus RNA detection kit (Da An Gene Co., Ltd. Of SunYat-sen University, China) which targets the ORF1ab and N genes of SARS-CoV-2 or (2) the RealStar SARS-CoV-2 RT-PCR kit 1.0 (Altona Diagnostics, Germany), which targets the E and S genes of SARS-CoV-2, or (3) the SarbecoV E-gene kit and SARS-CoV-2 RNA-dependent RNA polymerase (RdRP) kit (TIB Molbiol, Germany). All PCR's were performed according to the manufacturer's protocols.

3.3.3 *Whole genome amplification*

SARS-CoV-2 full genomes (chapter 3.4.1) were generated using the ARTIC Consortium PCR-based protocol (ARTIC Network, 2020) using primer set version 2 (https://github.com/artic-network/artic-ncov2019/tree/master/primer_schemes/nCoV-2019/V2)(2020). Library preparation and Illumina MiSeq sequencing was done using the KAPA Frag Kit and KAPA Hyper Prep kit (Roche Molecular Diagnostics, Switzerland) and MiSeq reagent v2 chemistry

(Illumina, U.S.A) according to the manufacturers' protocols. Genome assembly was done by mapping MiSeq reads to a representative African SARS-CoV-2 sequence (HCoV-19/Senegal/611/2020|EPI_ISL_420076). Genome annotations were made in analogy to a SARS-CoV-2 reference sequence (NC045512) using Geneious 9.1.8 (<https://www.geneious.com>).

SARS-CoV-2 full genomes in chapter 3.4.2 and 3.4.3 were generated using a Nimagen/Illumina-based workflow including the following steps. First, viral RNA was extracted from oro-nasopharyngeal swabs using the MagNA Pure 96 DNA and viral nucleic acid (NA) small-volume kit (Roche, Switzerland). cDNA was subsequently generated using the LunaScript[®] RT SuperMix Kit (New England Biolabs Inc., MA, USA) with 5.5 μ l RNA input. Whole genome sequencing was performed using the EasySeq[™] SARS-CoV-2 WGS Library Prep Kit V3 (NimaGen B.V., The Netherlands) according to the manufacturer's instructions. Deep sequencing was subsequently performed using the Illumina MiSeq platform.

3.3.4 *Lineage designation*

Lineage designation of the generated SARS-CoV-2 full genomes was performed using the Pangolin COVID-19 Lineage Assigner version 3.0.2 (<https://pangolin.cog-uk.io>).

Only sequences with a genome coverage exceeding 70% of the coding region were ascribed to lineages as proposed earlier (Rambaut et al., 2020).

3.3.5 *In-silico analyses*

Time-stamped Bayesian phylogeny based on sampling dates was performed in BEAST2 (<https://www.beast2.org/>). A codon position-specific general time reversible (GTR) substitution model with γ -distributed rates among sites was used. A SARS-CoV-2 sequence from Wuhan (MT019529) was used as an outgroup. The majority consensus of 10,000 trees from the posterior distribution with mean branch lengths was sampled. Subsequently, the phylogeny was annotated with TreeAnnotator and visualized in FigTree from the BEAST package (<https://beast.community/programs>).

Gaussian kernel smoothed violin plots (chapter 3.4.3) were generated using the ggplot2 package in R (R Foundation for Statistical Computing, <https://www.r-project.org>).

3.3.6 *Data availability*

All full genomes generated were deposited to GISAID. **Supplementary Table S4** provides all GISAID accession numbers of SARS-CoV-2 genomes sequenced during the studies.

3.4 RESULTS

3.4.1 SARS-CoV-2 genomic surveillance, early 2020

Genomic sequencing of SARS-CoV-2 is important to elucidate transmission chains and to identify mutations potentially changing the viral phenotype (Dietzel et al., 2017). By 10 November 2020, 195, 771 SARS-CoV-2 fully or partially sequenced genomes have been deposited in GISAID by the global scientific community. While affluent countries published more than 83% of those sequences, only 1.8% (n=3,580) originated from Africa (**Figure 9**).

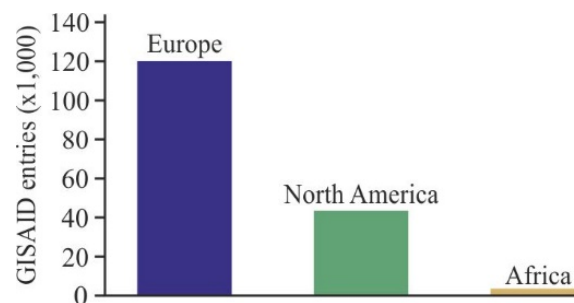


Figure 9. SARS-CoV-2 sequence entries from Africa and affluent settings in GISAID on 10 November 2020.

3.4.1.1 Sample characteristics March-April 2020

The first SARS-CoV-2 case in the country was detected 14 March 2020. To investigate the SARS-CoV-2 diversity in Benin during the onset of the pandemic, 12 SARS-CoV-2 genomes from Beninese citizens were determined. Samples were collected between 15 March and 15 April 2020 in southern and northern Benin, namely the capital city Cotonou, Porto-Novo and Natitingou (**Figure 10**). Only one of the 12 individuals was a patient from a local hospital, whereas the other 11 were returning travellers from European (France, Belgium, Italy) or Central-West African (Burkina Faso, Togo, DR Congo) countries who underwent mandatory SARS-CoV-2 testing upon entering Benin (**Table 2**).



Figure 10. Map of Benin showing the three sampling sites of oro-nasopharyngeal swabs in March-April 2020. Benin maps were obtained from The Humanitarian Data Exchange (<https://data.humdata.org>) and were plotted with the ggplot2 package in R (R Foundation for Statistical Computing, <https://www.r-project.org>).

3.4.1.2 *Generation of SARS-CoV-2 full genome sequences*

Full genomes were generated as described in chapter 3.3.3. Next-generation sequencing (NGS) read counts were between 1,237,688 and 443,672 reads of which between 440,720 and 130,691 mapped to the SARS-CoV-2 reference genome. 100% complete full genomes were assembled for 6/12 samples. For the other 6 samples, only 88-98% genome completeness was achieved (**Table 2**). Since NGS read count was in general sufficiently high to assemble full genome sequences, this may have resulted from a combination of (1) insufficient viral load (Ct-value above 30) and (2) amplicon sequencing with potential primer dropouts.

Table 2. Characteristics of SARS-CoV-2-positive patients from which SARS-CoV-2 full genomes were generated, March-April, 2020, Benin

Sample-ID	Gender	Age	Location	Travel history	Symptoms	Sampling date	Ct-value*	Genome completeness (%)
197	M	34	Porto-Novo	none	None	15.03.2020	28.2	100
260	F	29	Cotonou	France	None	17.03.2020	30.1	100
314	F	42	Cotonou	France	None	17.03.2020	35.9	88
461	F	45	Cotonou	France	None	17.03.2020	31.0	100
501	F	61	Cotonou	France	None	17.03.2020	32.7	100
843	F	41	Natitingou	Burkina Faso	None	19.03.2020	31.8	95
1022	F	29	Cotonou	Togo	None	25.03.2020	32.9	100
1092	M	38	Cotonou	Belgium	None	27.03.2020	37.0	90
1408	M	39	Cotonou	Italy	None	04.04.2020	34.1	98
1409	M	44	Cotonou	DR Congo	None	06.04.2020	34.7	94
1950	M	52	Cotonou	Belgium	Fever	15.04.2020	33.5	96
2012	M	34	Cotonou	France	None	15.04.2020	30.8	100

3.4.1.3 Phylogenetic analysis of generated SARS-CoV-2 genomes, March-April 2020

By 10 November 2020, 13 different SARS-CoV-2 lineages have been detected globally, namely A.1-A.5 and B.1-B.8 (Rambaut et al., 2020). In a full genome-based phylogeny representing this global diversity of SARS-CoV-2 lineages, the full genomes from Benin from early 2020 clustered with both globally circulating lineages A and B (**Figure 11**)

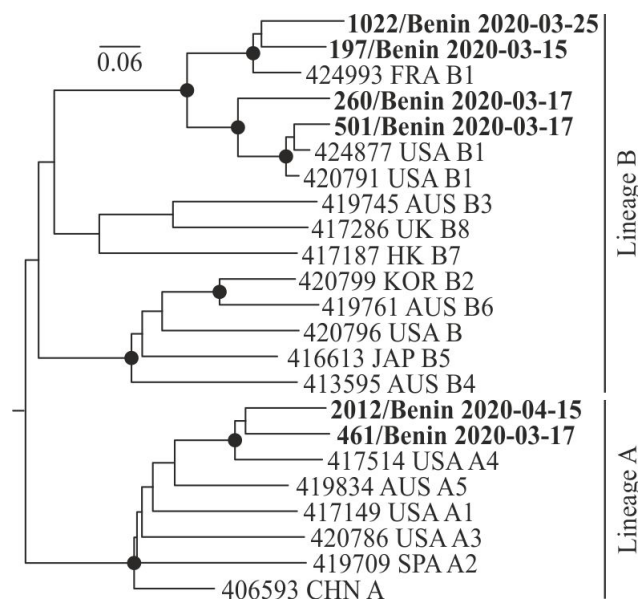


Figure 11. Phylogenetic analyses of SARS-CoV-2 lineages circulating in Africa by early 2020. Phylogenetic tree inferred using BEAST2 showing 23 complete SARS-CoV-2 genomes globally sampled from humans. Posterior support >0.9 is shown at nodes as filled circles. Benin-derived sequences are shown in bold. Only sequences without missing information were used for this analysis. Sequences are designated with GISAID accession IDs/country of origin/lineage according to (Rambaut et al., 2020). Adapted from (Sander et al., 2021b).

3.4.1.4 Mutation analysis of SARS-CoV-2 full genomes, March-April 2020

To identify genetic variation within those two lineages, nucleotide differences within the 12 Benin-derived sequences were analysed (Moreira-Soto et al., 2018). In total, 12 variable nucleotide positions across the complete SARS-CoV-2 genome were observed, resulting in seven amino acid exchanges known from other strains circulating globally (**Figure 12**) (Yin, 2020). According to this mutational analysis, four Benin-derived strains showed mutational patterns most similar to lineage A.4, which has been reported predominantly from the USA until mid-September 2020. Since none of those travellers returned from the USA, this finding highlights the global dispersion of SARS-CoV-2 variants and likely sample biases in the existing datasets (**Table 2**). Another eight strains showed mutational patterns similar to lineage B.1, a globally circulating lineage originating in Europe (Rambaut et al., 2020), which was consistent with the individual travel histories (**Table 2**). Of note, all eight strains belonging to lineage B.1 harboured a nucleotide exchange resulting in the D614G spike protein variant, which may be associated with increased transmissibility (Zhang et al., 2020).

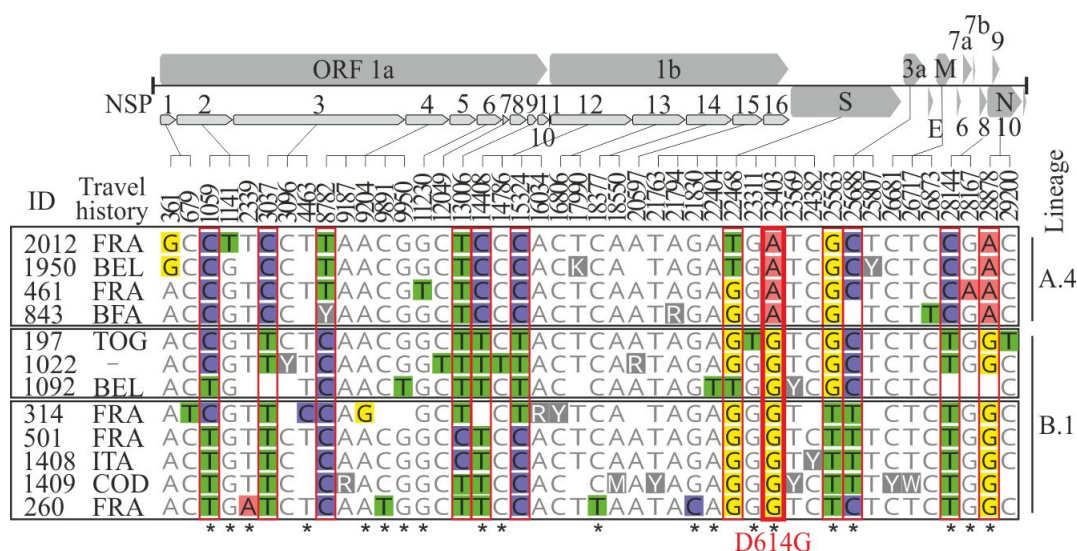


Figure 12. Mutation analysis of SARS-CoV-2 full genomes from Benin collected early 2020. Alignment showing all variable sites across Benin derived SARS-CoV-2 genomes from this study. Empty spaces indicate lack of sequence information. Red boxes denote variable nucleotide positions in at least three or more sequences. Gray boxes denote groups belonging to lineage A.4 or B.1 according to the work of (Rambaut et al., 2020). Asterisks indicate nonsynonymous substitutions. Nucleotide positions correspond to the SARS-CoV-2 reference genome (GenBank accession number MN908947). BEL, Belgium; BFA, Burkina Faso; COD, Democratic Republic of the Congo; FRA, France; ITA, Italy; TOG, Togo; NSP, nonstructural protein; ORF, open reading frame. Adapted from (Sander et al., 2021b).

3.4.2 SARS-CoV-2 genomic surveillance, early 2021

Numerous genetic variants of SARS-CoV-2 have emerged globally since the start of the COVID-19 pandemic (<https://cov-lineages.org>). While most of the emerging lineages have not contributed significantly to a more severe course of the pandemic, three categories of variants, so called variants of concern (VOC) with specific mutations associated with a) increased transmissibility, b) increased disease severity or c) partial or complete immune evasion, are monitored with caution. By May 2021 the World Health Organization (WHO) defined four lineages as VOCs: B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma), and B.1.617.2 (Delta) (WHO, 2021a).

As of 25 May 2021, a total of 55 SARS-CoV-2 lineages were described in West Africa considerably fewer than the >350 lineages in affluent regions (**Figure 13B**), demonstrating the notoriously weak genomic surveillance in sub-Saharan Africa (**Figure 13A**). From Benin, no further full genomic sequences than the 12 genomes described in chapter 3.4.2 were generated to that date.

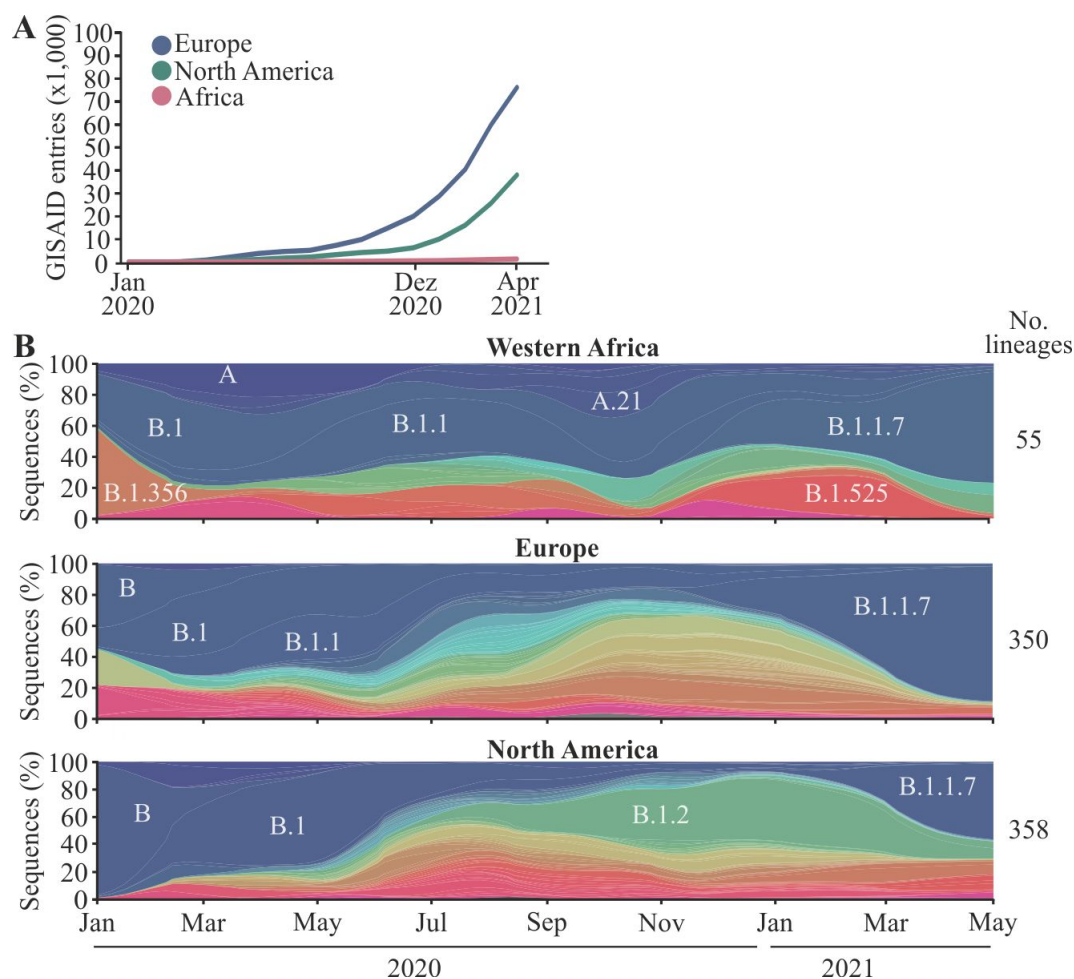


Figure 13. Globally available SARS-CoV-2 genomic data by early 2021. (A) SARS-CoV-2 sequence entries from Africa and affluent settings in GISAID on 24 April 2021. (B) SARS-CoV-2 genomic epidemiology in West Africa and affluent settings during January 2020–May 2021. Frequency plots were attained from Nextstrain (<https://nextstrain.org/ncov/global>) on May 25, 2021. The global dataset was filtered for the 3 different regions and their associated countries, including 11 countries for West Africa, 45 countries for Europe, and the United States and Canada for North America. Frequencies are coloured by PANGO lineages and normalized to 100% at each time point for 705 sequences from West Africa, 2,400 sequences for Europe, and 1,515 sequences for North America. Adapted from (Sander et al., 2021a).

3.4.2.1 SARS-CoV-2 full genome generation of Benin-derived samples in early 2021

Continuing genomic surveillance in Benin, 68 SARS-CoV-2–positive diagnostic respiratory samples tested at the reference laboratory in Benin during 30 January – 3 March 2021 were sequenced using a Nimagen/Illumina-based whole genome sequencing (WGS) workflow (chapter 3.3.3). All samples used had a cycle threshold ≤ 36 (Sarbeco E-gene assay; TIB Molbiol) (**Supplementary Table S5**). NGS read coverage was sufficiently high in order to generate sequences of 91–100% genome completeness (**Supplementary Table S5**)

3.4.2.2 Lineage assignment of SARS-CoV-2 full genomes, January–March 2021

Lineage assignment of the generated SARS-CoV-2 genomes was performed as described in chapter 3.3.4. In total, the 68 Benin-derived near-complete genomes were designated to 10

unique lineages (**Table 3**), two A lineages and eight B lineages. Lineage B.1.1.7 (22%), A.27 (19.1%), B.1.525 (17.6%) and B.1.1.318 (16.2%) were the most prominent ones (**Table 3**). These results are consistent with recent online sequence reports from Western Africa (Augustin Anoh et al., 2021; Ozer et al., 2022).

Table 3. SARS-CoV-2 lineages to which the 68 Benin-derived full genomes were designated and the frequency of sequences within each lineage

Lineage	Number of sequences (%)
B.1.1.7	15 (22.0)
A.27	13 (19.1)
B.1.525	12 (17.6)
B.1.1.318	11 (16.2)
B.1	8 (11.8)
B.1.1.10.3	5 (7.4)
B.1.1.420	1 (1.5)
B.1.160	1 (1.5)
A.23.1	1 (1.5)
B.1.214.2	1 (1.5)

3.4.2.3 Mutation analysis of SARS-CoV-2 full genomes, January-March 2021

To identify genomic regions with mutation prevalence, a 100% consensus sequence of all 68 Benin-derived sequences was generated. In total, 229 non-synonymous nucleotide substitutions across the whole genome were present, of which 57 (24.9%) occurred in the S protein (**Figure 14A**). Variants with mutations in the S gene are of particular interest since they may alter the transmissibility and antigenicity of the virus (Harvey et al., 2021). Internationally recognized VOCs by May 2021 share 16 S mutations in unique combinations (<https://covariants.org/shared-mutations>) (**Supplementary Table S6**). The Benin-derived SARS-CoV-2 strains shared eight unique S mutations reported in VOCs, although most of those strains were not defined as any VOC other than Alpha (**Figure 14B**). Despite presence of the mutation P681R (associated with the Delta VOC) in one sequence, that sequence was typed as lineage A.23.1, and no Delta variant was found. Mutations associated with immune escape, such as the E484K mutation harboured by the Beta, Gamma and Delta sublineage AY.2 (Baj et al., 2021; Voloch et al., 2021; Wang et al., 2021), appeared in three distinct circulating SARS-CoV-2 lineages in Benin. Predominant lineages harboured apparently more mutations than those found at lower frequencies (**Figure 14B**). Of note, the D614G mutation was present in all sublineages of the parental lineage B, which are now the predominant circulating lineages.

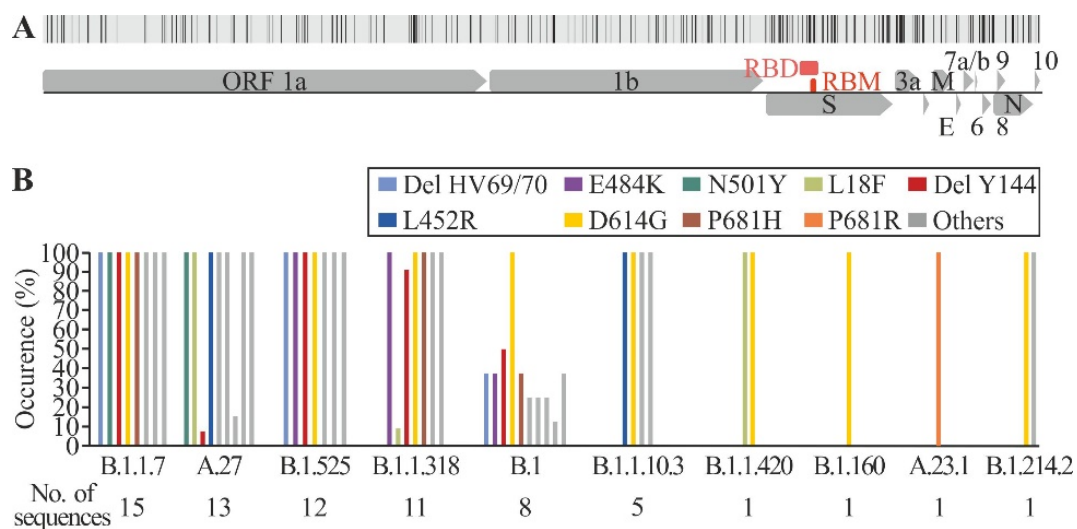


Figure 14. Genomic surveillance of SARS-CoV-2 lineages in Benin, 2021. (A) Nonsynonymous mutations of Benin-derived SARS-CoV-2 sequences across the full genome. (B) Spike mutations occurring in the SARS-CoV-2 lineages circulating in Benin. Hallmark mutations of variants of concern are shown in colour. Other mutations occurring in the Benin-derived sequences are depicted in grey and summarized as ‘others’. ORF, open reading frame; RBD, receptor-binding domain. Adapted from (Sander et al., 2021a).

3.4.3 Surveillance of the SARS-CoV-2 Delta VOC in Benin, mid 2021

In October 2020, a new SARS-CoV-2 lineage (B.1.617.2) was first detected in India, which showed an increased transmissibility compared to other lineages (Li et al., 2022). On 11 May 2021, WHO declared this lineage a VOC because its mutations likely facilitated potential immune escape and intense replication (Li et al., 2022; Planas et al., 2021) which is consistent with the global spread of the Delta VOC (WHO, 2021a), and rapid outcompetition of other lineages, such as Alpha and Kappa, in India (Mlcochova et al., 2021). By September 2021, more than 33 sublineages (AY.1–AY.33) of the Delta VOC have been detected globally (<https://cov-lineages.org>). By August 2021, the Africa Centres for Disease Control and Prevention had reported Delta VOC infections from 30 countries (Africa CDC, 2021). Nevertheless, epidemiologic information on the emergence and dissemination of the Delta VOC in Africa is scarce. As detailed in chapter 3.4.2, no Delta VOC was detected in Benin until end of March 2021. To detect the first emergence and monitor the spread of this SARS-CoV-2 lineage in Benin, West Africa, genomic surveillance was further investigated between May–July 2021.

3.4.3.1 Generation of SARS-CoV-2 full genomes sequences, mid 2021

For genomic surveillance, the reference laboratory in Benin kept a subset of 17.0% (200/1181; 20% of all positive samples in May, 17% in June and 13% in July) of all SARS-CoV-2–positive respiratory samples. From those, a total of 166 Samples with a cycle threshold below 35

(Sarbeco E-gene assay; TIB Molbiol, <https://www.tib-molbiol.de>) were selected for genomic sequencing using a NimaGen/Illumina-based workflow as conducted previously (Sander et al., 2021a) (chapter 3.3.3). Near-complete viral genomes were attained for 67.8% (114/166) of the SARS-CoV-2 positive samples reaching 9.7% of all SARS-CoV-2 positive samples in Benin. The remaining 52 samples did not provide sufficient genomic data for genome coverage of more than 70%.

3.4.3.2 Lineage assignment of SARS-CoV-2 full genomes, mid 2021

Lineage analysis (chapter 3.3.4) resulted in the assignment of the generated SARS-CoV-2 full genomes to 12 distinct SARS-CoV-2 lineages (**Figure 15**); the most common were B.1.1.318(50/114) and the Delta VOC B.1.617.2 (38/114) (**Table 4**).

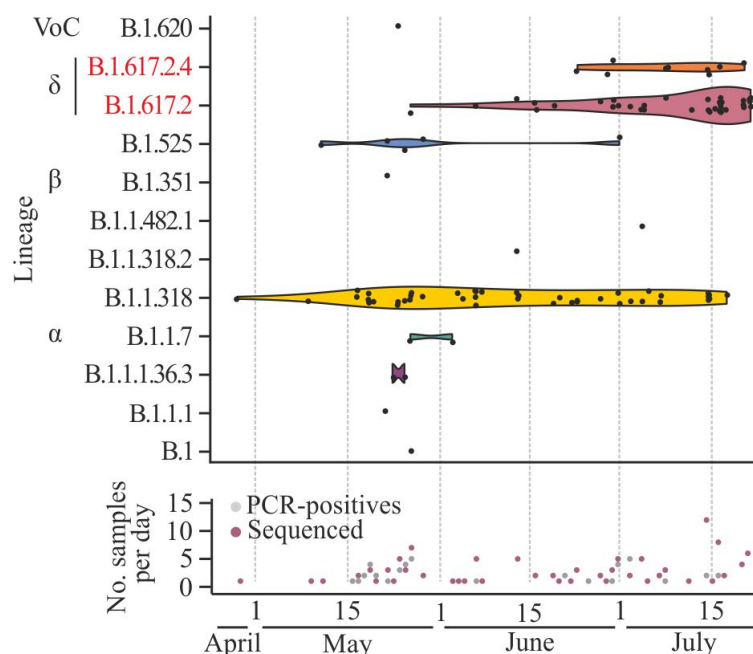


Figure 15. Spread of SARS-Cov-2 lineages in Benin, April-July 2021. Top Gaussian kernel smoothed violins representing the density of observed occurrences per SARS-CoV-2 lineage at a given time point during the sampling period from the end of April until mid-July 2021. Black dots represent lineage occurrences of 114 generated genomes. Height of the violin plot corresponds to density of lineage in time. Bottom of graph shows collection date of the 200 SARS-CoV-2-positive samples collected for this study. Red indicates the subset for which near-full genomes were generated. Adapted from (Sander et al., 2021a).

Table 4. Designated lineages, number of samples, and sample location and dates of 114 detected SARS-CoV-2 genomes, Benin, May–July 2021.

Lineage	WHO label	No. samples	Alias	All locations detected	Earliest sample	
					Date collected	Location
B.1.1.318	NA	50	NA	Cotonou, Come, Grand-Popo	28.04.2021	Grand-Popo
B.1.617.2	Delta	38	NA	Cotonou, Come, Lokossa	27.05.2021	Cotonou
AY.4	Delta sublineage	9	B.1.617.2.4	Cotonou	24.06.2021	Cotonou
B.1.525	NA	5	NA	Cotonou, Grand-Popo	12.05.2021	Grand-Popo
B.1.1.7	Alpha	3	NA	Cotonou	27.05.2021	Cotonou
C.36.3	NA	3	B.1.1.1.36.3	Cotonou	24.05.2021	Cotonou
AV.1	NA	1	NA	Cotonou	05.07.2021	Cotonou
AZ.2	NA	1	NA	Cotonou	14.06.2021	Cotonou
B.1	NA	1	NA	Cotonou	27.05.2021	Cotonou
B.1.1.1	NA	1	NA	Cotonou	23.05.2021	Cotonou
B.1.351	Beta	1	NA	Cotonou	23.05.2021	Cotonou
B.1.620	NA	1	NA	Cotonou	25.05.2021	Cotonou

*NA, not applicable; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; WHO, World Health Organization.

3.4.3.3 SARS-CoV-2 epidemiology in Benin during April-July 2021

Compared with the molecular SARS-CoV-2 epidemiology in Benin in the beginning of 2021 (see chapter 3.4.2), the situation changed drastically within only two months. The only commonality between these two time points was the continuous detection of the B.1.1.318 lineage and detection of the Alpha VOC (Sander et al., 2021a). Differently from the SARS-CoV-2 lineage distribution beginning of 2021 (chapter 3.4.2), the Beta and Delta VOC were detected now in mid-2021. Notably, the Alpha and Beta VOC, as well as six other lineages, were only detected sporadically in this study (**Figure 15**), highlighting lack of intense transmission of both the Alpha and the Beta VOC during the study period. To assure that no Delta strains were missed in early 2021 (Sander et al., 2021a), another 88 genomes from January-March 2021 were sequenced, but as expected, no Delta VOC was found (**Supplementary Table S4**).

3.4.3.4 Emergence and dissemination of the SARS-CoV-2 Delta VOC in Benin

On May 27, 2021, the Delta VOC was first detected in Cotonou, the capital of Benin (**Table 4**). The frequency of the Delta VOC increased rapidly, from 3.4% in May to 37.9% in June and 63.6% in July (**Table 5**). Two sublineages of Delta VOC were detected in the dataset, namely B.1.617.2 and the AY.4. The Delta VOC sublineage AY.4 was first detected in Cotonou on 24 June 2021, co-circulating with B.1.617.2 henceforth. Of note, no detection of AY.4 outside of the capital Cotonou was observed (**Table 4**). However, it cannot be excluded that a sampling bias due to low sample numbers from other cities affected the apparent spatial distribution of AY.4.

Table 5. Detection rates of major SARS-CoV-2 lineages over time, Benin, May–mid-July 2021.

Pangolin lineages	% Lineages		
	May	June	July
B.1.1.7 Alpha	3.4	3.4	1.8
B.1.351 Beta	3.4	0	0
B.1.617.2 Delta	3.4	27.6	52.7
AY.4 Delta	0	10.3	10.9
B.1	3.4	0	0
B.1.1.1	3.4	0	0
B.1.1.318	58.6	55.2	29.1
B.1.525	13.8	0	1.8
B.1.620	3.4	0	0
AV.1	0	0	1.8
AZ.2	0	3.4	0
C.36.3	6.9	0	1.8

*Lineages according to Pangolin software (<https://github.com/cov-lineages/pangolin>). SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

3.4.3.5 Adaptive mutation of the SARS-CoV-2 Delta lineage in Benin

Given the widespread convergent evolution of different spike protein mutations observed across clades, the occurrence of mutations associated with partial or complete immune escape or reduced vaccine efficacy, such as E484K (Baj et al., 2021) would be cause for concern. Therefore, the frequency of occurring mutations within the 47 full genomes assigned to the two Delta lineages were analysed (**Figure 16**). Neither the Delta VOC sublineage AY.1 signature mutations (V70F, W258L and K417N) (Kannan et al., 2021), nor the E484K mutation associated with immune escape in the Delta VOC sublineage AY.2 (Baj et al., 2021) were present in any of the Delta VOC genomes from Benin (**Figure 16**). Additionally, not all mutations of the Delta VOC occurred in SARS-CoV-2 strains circulating in Benin at identical frequencies (**Figure 16**).

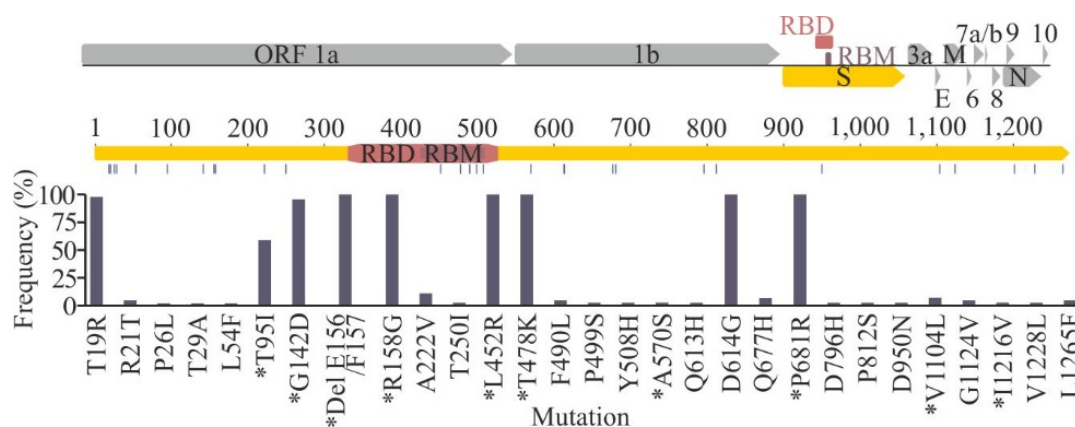


Figure 16. Nonsynonymous substitutions within the spike protein occurring within 47 SARS-CoV-2 Delta variant genomes sampled, Benin, May–July 2021. For clarity of presentation, a SARS-CoV-2 full genome illustration is included at the top. Asterisks (*) indicate mutations for which genomic information was not available in all 47 Delta variant genomes, potentially due to assay failures or relatively lower DNA concentrations. Blue bars represent frequency of a given mutation in 46 genomes (for T95I, L452R, T478K, and P681R mutations); in 45 genomes (for A570S, V1140L, and I1216V mutations); and in 23 genomes (for G142D, Del E156/F157, and R158G mutations). Adapted from (Yadouleton et al., 2021).

3.5 DISCUSSION

This chapter analysed the genetic diversity and microevolutionary adaptation of SARS-CoV-2 in humans in Benin, Western Africa in the period of March 2020 – July 2021.

After its first emergence in December 2019 in Wuhan China, SARS-CoV-2 quickly spread around the globe and quickly led to the emergence of multiple genetic variants. By November 2020, already 13 different SARS-CoV-2 lineages descending from the early lineages A and B have been reported globally (Rambaut et al., 2020). The genomic surveillance conducted in Benin, West Africa, showed multiple introductions of at least two of those distinct SARS-CoV-2 lineages termed A.4 and B.1 into Benin before mid-April 2020. Sampling of those divergent SARS-CoV-2 strains from travellers within only four weeks was consistent with global mixing of SARS-CoV-2 variants early during the pandemic and with sequence reports from other African countries such as South Africa, Kenya and Nigeria (Giandhari et al., 2021; Githinji, 2020; Happi et al., 2020).

Of particular importance within these early genomes was the detection of the D614G mutation in the gene encoding the S protein, which occurred in more than 60% of the sequences already by mid-April 2020 (**chapter 3.4.1**) and which became predominant in all circulating B lineages until the following year (**chapter 3.4.2**). This mutation first emerged late January 2020 (Isabel et al., 2020) in the B.1 lineage and became one of the first established mutations in the SARS-CoV-2 S gene due to a replication fitness advantage resulting in higher viral loads and thus higher transmissibility (Plante et al., 2021). Cross-species transmission events may lead to rapid adaptation to the new host, typically resulting in species-specific adaptations (Bashor et al., 2021). However, the acquisition of the D614G mutation in experimentally SARS-CoV-2-infected cats suggests substantial adaptive pressures shaping viral evolution at this position in early stages after host switching (Bashor et al., 2021). This results also substantiates the likely disadvantage of lineages lacking this mutation, such as the early A lineage, which was indeed nearly displaced by the B.1 lineage until mid-2021 globally.

However, specific lineage A variants A (A.27 and A.23.1) were abundant in Benin until spring 2021 (**chapter 3.4.2.3**) which is in accordance with other reports from African countries (Augustin Anoh et al., 2021; Bugembe et al., 2021; Chouikha et al., 2022) as well as the likely origin of these lineages on the African continent (Bugembe et al., 2021; Calvignac-Spencer et al., 2021). Lineage A.27 for example likely originated in April 2020 in Niger (neighbouring country of Benin) (Chouikha et al., 2022; Kaleta et al., 2022) and remained the second most

circulating lineage in Benin by spring 2021 (**chapter 3.4.2.2**). Although these variants of lineages A do not show the D614G mutation typical for lineage B variants, they do harbour several amino acid changes in the S protein. Among those, a mutation at S position 613 in A.23.1 (Bugembe et al., 2021) or the N501Y S mutation in A.27 (Calvignac-Spencer et al., 2021). Furthermore, both lineages harbour mutations associated with immune escape, such as the Delta P681R (A.23.1) or L452R (A.27) (**chapter 3.4.2.3**), suggesting acquisition of fitness through accumulation of mutations altering transmissibility (Liu et al., 2022) and immune escape.

The increased emergence of SARS-CoV-2 variants with mutations of concern during late 2020 quickly required a categorization system for variants to prioritize global monitoring and research. Therefore, WHO is monitoring variants with specific mutations for which there is clear evidence of increased transmissibility, increased disease severity, or partial or complete immune evasion (VOCs). By May 2021 the World Health Organization (WHO) defined four lineages as VOCs: B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma), and B.1.617.2 (Delta) (WHO, 2021a). The Benin-derived SARS-CoV-2 strains circulating by early 2021 shared eight unique S mutations reported in VOCs, although most of those strains were not defined as any VOC other than Alpha (**Figure 11B**). Apparently, there were more S mutations in predominant lineages than in those found at lower frequencies (**Figure 11B**), suggesting a putative higher fitness mediated by genomic change. However, mutations of the Delta VOC did not occur at identical frequencies in SARS-CoV-2 strains circulating in Benin, (Appendix Figure 1), suggesting that the Delta variant expanded in Benin irrespective of mutations that are hypothesized to confer partial immune escape. In sum, the convergent evolution of such mutations suggests selective advantages of key mutations associated with partial or complete immune escape across different SARS-CoV-2 lineages (Cherian et al., 2021; Martin et al., 2021). This may likely be the result of viral adaptation due to an increasing immune population (Harvey et al., 2021). The exact SARS-CoV-2 seroprevalence in Benin is unclear but expected to be much higher than assumed by reported cases (WHO, 2021b). Robust investigation of the extent of SARS-CoV-2 transmission in Benin and adjacent countries could clarify the immune status of the population in the Sub-Saharan African region and elucidate potential immune escape by SARS-CoV-2 lineages such as the Delta VOC.

The data presented within this chapter have potential limitations. First of all, the tested sample subsets were limited by a laboratory-based sampling that does not represent the total population of Benin. Samples were not received from all regions across the country, and additionally only

a small number of all positive cases in the country were used for full genomic sequencing. These reasons may have led to failure to detect other circulating lineages. Ideally, all SARS-CoV-2 positive samples in the country would be sequenced, but this was practically not possible due to financial and logistical challenges during the pandemic, such as restraint access to state-of-the-art reagents, missing NGS device on-site in Benin or severe travel restrictions impeding sample transport. Moreover, the NGS strategy used for full genome sequencing has limitations in terms of accuracy. Although the amplicon-based sequencing approach using the ARTIC or Nimagen protocols is highly robust for degraded and low concentrations of RNA, which was expected for the samples received from Benin due to sampling performed several weeks-months before sample processing and unknown freeze-thaw cycles, it may nevertheless produce decreased coverage in specific genomic regions or even incomplete full genomes due to primer drop-outs caused by nucleotide mutations in the binding regions (Chiara et al., 2021). However, to limit any potential false lineage-designation using the Pangolin COVID-19 Lineage Assigner, only sequences with a genome coverage exceeding 70% of the coding region were ascribed to lineages as proposed by the inventors of the program (Rambaut et al., 2020).

In conclusion, the SARS-CoV-2 genomic surveillance performed in this chapter emphasize the importance of genomics-based surveillance of emerging infectious diseases, especially in global regions with high transmissions. Timely identification of lineages or mutations with altered fitness remains challenging but generating genetic sequences provides a fundamental step for functional characterization studies of mutations of concern and in-depth analyses of viral adaptation (Obermeyer et al., 2022).

CHAPTER 4

GENOMIC DETERMINANTS OF FURIN CLEAVAGE IN EUROPEAN SARS-RELATED BAT CORONAVIRUSES¹

4.1 INTRODUCTION

Until today, there are seven known human pathogenic CoVs of recent or regular zoonotic origin. Four of them are endemic and known as “common cold” or “seasonal” HCoV-229E, HCoV-OC43, HCoV-NL63 and HCoV-HKU1) whereas the other three are highly pathogenic coronaviruses, including the betacoronaviruses SARS-CoV, MERS-CoV as well as SARS-CoV-2 (**Figure 17**) (Jo et al., 2021). Among the highly pathogenic CoVs, SARS-CoV-2 is unique in its high transmissibility between humans (Jones et al., 2021). SARS-CoV and SARS-CoV-2 are conspecific by belonging to the species *Severe acute respiratory syndrome-related coronavirus*, subgenus *Sarbecovirus* and both use the host cell protein angiotensin-converting enzyme 2 (ACE2) as the main receptor for cellular entry (Hoffmann et al., 2020b).

¹ Parts of this chapter have been previously published in:

Sander AL, Moreira-Soto A, Yordanov S, Toplak I, Balboni A, Ameneiros RS, et al. Genomic determinants of Furin cleavage in diverse European SARS-related bat coronaviruses. *Commun Biol.* 2022;5(1):491.

Copyright permissions are granted to the authors of the publications.

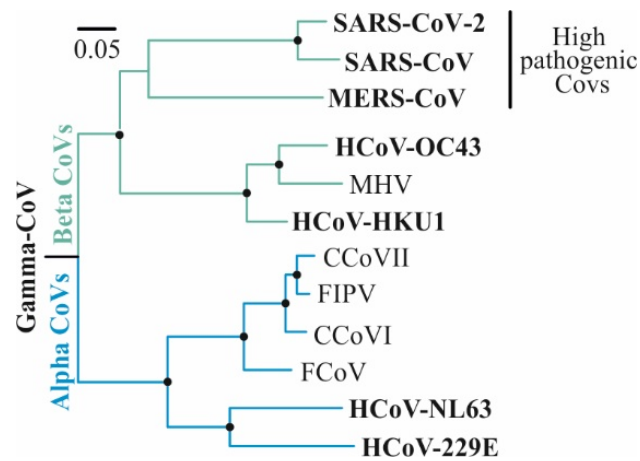


Figure 17. Phylogenetic relationship of human pathogenic coronaviruses. Displayed is a maximum-likelihood phylogeny based on 816-bp alignment of the RNA-dependent RNA-polymerase (RdRp) of selected alpha- (blue) and betacoronaviruses (green). Human pathogenic CoVs are shown in bold. Circles at nodes indicate support of grouping in $\geq 90\%$ of 1,000 bootstrap replicates. Scale bar represents nucleotide substitutions per site. Accession numbers: GQ477367, AY994055, KP849472, FJ938051, NC005831, NC002645, NC006213, NC048217, NC006577, MT019529, NC004718, NC019843. The Beluga whale gammacoronavirus was used as outgroup (Acc. Number NC010646). CCoV, Canine Coronavirus; FCoV, Feline Coronavirus; FIPV, Feline infectious peritonitis virus; HCoV, Human coronavirus; MHV, Mouse hepatitis virus.

Viral entry into the target cell is performed by the S protein, which mediates host cell receptor binding as well as fusion of the viral envelope with the host cell membrane. In order to do so, the S protein needs to become proteolytically activated by cleavage through host cell proteases at specific cleavage sites located between the S1 and S2 subunits (S1/S2 site) and at the N-terminus of the S2 subunit (S2' site) (Belouzard et al., 2009). For coronaviruses in general, several proteases have been detected to effectively cleave and thus activate the S proteins, including transmembrane protease serine subtype 2 (TMPRSS2) (Glowacka et al., 2011; Hoffmann et al., 2020b), cathepsin L (Bosch et al., 2008) and trypsin (Belouzard et al., 2009; Bertram et al., 2011). Whereas SARS-CoV is cleaved by TMPRSS2 at the S1/S2 site (Glowacka et al., 2011; Matsuyama et al., 2010), SARS-CoV-2 contains a polybasic furin cleavage site (FCS) at the S1/S2 boundary (Coutard et al., 2020). Efficient enzymatic processing at the FCS is thought to be essential for entry into human lung cells and may also determine the efficiency of infection of the upper respiratory tract and consequent transmissibility of SARS-CoV-2 (Hoffmann et al., 2020a).

The FCS in SARS-CoV-2 is atypical in that it does not follow the canonical R-X-K/R-R amino acid residue pattern that furin preferentially cleaves (Thomas, 2002) but is defined by P/H-R-R-A-R with additional upstream basic residues occurring in some variants (Coutard et al., 2020). The existence of the FCS has led to various hypotheses regarding the evolution of SARS-CoV-2, including conjectures about the possibility of a non-natural origin from laboratory experiments (Bloom et al., 2021; Calisher et al., 2021). FCS naturally occur in different other

coronaviruses (Wu and Zhao, 2020), such as MERS-CoV, which also harbours an atypical furin cleavage motif, namely P-R-S-V-R (Duckert et al., 2004; Millet and Whittaker, 2014). However, among its species *SARS-related coronavirus*, SARS-CoV-2 is the only virus with this genomic attribute, as even its closest known relatives, such as the bat coronaviruses RaTG13 from China and BANAL-20-52 from Laos, as well as the pangolin coronaviruses, lack an FCS (Liu et al., 2020; Temmam et al., 2022b; Wu and Zhao, 2020).

The natural hosts of SrC are horseshoe bats (order Chiroptera, family Rhinolophidae, genus *Rhinolophus*) (Li et al., 2005b), which are widely distributed in the Old World. Next to a broad genetic diversity of SrC found in Asian rhinolophid bats (Delaune et al., 2021; Lau et al., 2005; Temmam et al., 2022a; Wacharapluesadee et al., 2021; Zhou et al., 2021), several studies previously found that European horseshoe bats also carry highly diversified SrC. Genomic and phylogenetic analyses revealed that they are conspecific with, but phylogenetically distinct from SrC detected in Asia (Balboni et al., 2011; Drexler et al., 2010; Muth et al., 2018; Rihtaric et al., 2010). However, detailed characterization of their spike glycoprotein genes in terms of protease cleavage and potential mechanisms leading to the acquisition of cleavage sites has not been performed for either of them.

4.2 OUTLINE AND AIM OF THIS CHAPTER

The FCS in SARS-CoV-2 is unique within the species *SARS-related coronavirus* and its origin remains elusive. Therefore, it is highly critical to identify how, and at which step in the evolution of SARS-CoV-2 the FCS was acquired.

In this chapter, I re-assessed diverse SrC from European horseshoe bats and analysed the spike-encoding genomic region harbouring the FCS in SARS-CoV-2. I compare their S1/S2 genomic region to other sarbecoviruses and mammalian coronaviruses to better understand the genealogy of furin cleavage at the S1/S2 site within HCoV-2. In addition, *in silico* modelling will be used to test whether there are molecular features of influenza viruses, a prototypic example of RNA viruses acquiring an FCS during host switches that facilitate FCS acquisition. In the last part of this chapter, the potential acquisition of FCS through molecular mechanisms such as nucleotide substitution, insertion, or recombination is discussed.

4.3 METHODS

4.3.1 *Sample collection and processing*

Bats (*Rhinolophus euryale*, *R. blasii*, *R. ferrumequinum* and *R. hipposideros*) were sampled in four countries (Bulgaria, Italy, Slovenia, Spain) (Balboni et al., 2011; Muth et al., 2018; Rihtaric et al., 2010). All animals were handled according to national and European legislation for the protection of animals (EU council directive 86/609/EEC). No animals were sacrificed during this study. All animal handling and sampling were done by trained personnel, with animal safety and comfort as the first priority during minimally invasive sampling (collection of faeces). Captured bats were freed from nets immediately and put into cotton bags for 2 to 15 min to allow them to calm down before examination. While being kept in bags, bats produced faecal pellets that were transferred to 500 µl RNAlater RNA stabilization solution (Qiagen, Hilden, Germany) for sample processing. Licenses for sampling of bats using mist nets, hand nets or harp traps were obtained from the respective countries and authorities: Bulgarian Ministry of Environment and Water, permit No. 192/26.03.200913, Italian Ministry of the Environment, permit No. 192/26.03.200952, Slovenian Environment Agency, permit No. 35701-80/200453, Service for the Biodiversity Conservation of the Rural Counselling of the Xunta de Galicia, Spain, permit No. 52/2010 n.s. 13697.

4.3.2 *Analysis of samples by reverse-transcription PCR*

Specimens were screened for the presence of coronaviral RNA by using nested reverse transcription-PCR (RT-PCR) amplifying 455 bp fragments of the RdRp gene. Briefly, RT-PCR was performed using the QIAGEN (Hilden, Germany) one-step RT-PCR kit with 5 µl of RNA, 200 nM of primer PC2S2, 900 nM of primer PC2As1 and 1 µl QIAGEN one-step RT-PCR kit enzyme mix. The amplification protocol comprised 30 min at 50°C; 15 min at 95°C; 10 cycles of 20 s at 94°C, 30 s starting at 62°C with a decrease of 1°C per cycle, and 40 s at 72°C; and 30 cycles of 20 s at 95°C, 30 s at 52°C, and 40 s at 72°C (de Souza Luna et al., 2007). For further phylogenetic analyses these amplicons were extended to >800 bp fragments towards the 5'-end using the 5' primers: SP3080, SP3195, SP3374, GrISP1, and GrISP2, (Drexler et al., 2010). One bat SrC spike glycoprotein gene (termed BB99-04; GenBank Acc No. KR559017) was fully characterized using nested RT-PCR primer sets (**see Supplementary Material**).

S1/S2 genomic regions (346 bp; positions 23,422-23,767 in SARS-CoV-2 Wuhan strain GenBank Acc. Number MT019529) of eight bat SrC (GenBank Acc No. KC633198, KC633201, KC633203-205, KC633209, KC633212 and KC633217) were characterized using a hemi-nested RT-PCR assay using the following oligonucleotides: panSARS-S1S2-F1, panSARS-S1S2-F2 and panSARS-S1S2-R. An extended S2 region (691 bp; positions 23,422-24,112 in SARS-CoV-2 Wuhan strain) for phylogenetic analyses was amplified using the same forward primers but a panSARS-S1S2-R2 reverse primer. Briefly, RNA was reverse transcribed for 30 min at 50°C using the SSIII One-Step Kit (Thermo Fisher) followed by 45 PCR cycles of 94°C for 15 seconds, 58°C for 20 or 45 seconds, respectively and 72°C for 1 minute. The 2nd round PCR was performed at the same conditions as the 1st round without reverse transcription. PCR amplicons were Sanger sequenced.

To detect single nucleotide variants within the S1/S2 site, PCR amplicons of the S1/S2 genomic region (346 bp fragments) were sequenced using the MiSeq System with the MiSeq Reagent Kit v2 (500-cycles) according to the manufacturer's instructions. Sequence reads obtained from the library were mapped against their corresponding S1/S2 genomic sequence in Geneious 9.1.8. A custom script using the python module pysam (<https://github.com/pysam-developers/pysam>) was utilized to determine the nucleotide distribution on specific genome positions from the read alignment.

4.3.3 Phylogenetic analyses

A tblastx search of the complete spike sequence of the Bulgarian SrC (GenBank Acc No. GU190215) within the taxonomy ID 11118 (*Coronaviridae*) excluding taxonomy ID 2697049 (SARS-CoV-2) was performed on 28 December 2021. Hits with percentage identities below 80% were non-SrC sequences and were thus not included in the dataset. SARS-CoV sequences from experimental infections or clones as well as sequences with less than 27,000 nt or gaps in the spike-encoding region were excluded, resulting in 88 sequences. One reference sequence of each SARS-CoV (NC004718) and SARS-CoV-2 (MT019529) as well as the nine sequences from European bats newly generated within this study were added, resulting in a final dataset of 99 sequences. Because coronaviruses frequently recombine (Vakulenko et al., 2021b) only the S2 region (495 nt) of the 691 nucleotide fragment was used for the phylogenetic analysis. Maximum-likelihood phylogenies were generated using FastTree (Price et al., 2009) Version 2.1.10 using a GTR substitution model and 1,000 bootstrap replicates. Local support values are based on the Shimodaira-Hasegawa (SH) test (Shimodaira and Hasegawa, 1999).

4.3.4 *In silico analyses*

Secondary structures were modelled using the UNAFold web server (Zuker, 2003). Furin cleavage sites were predicted using the ProP v.1.0b ProPeptide Cleavage Site Prediction software (Duckert et al., 2004). Sequences were retrieved from the NCBI Taxonomy website downloading all sequences of the subgenera *Duvinacovirus* (HCoV-229E), *Merbecovirus* (MERS-CoV), *Setracovirus* (HCoV-NL63) and *Embecovirus* (HCoV-HKU1 and HCoV-OC43). Duplicates, non-complete genomes, sequences from experimental infections and clones were excluded from all datasets. Only the NCBI reference sequences of human genomes were left in the dataset at this stage of analysis. For Sarbecoviruses, the dataset of the tBlastx search used for the SrC-phylogeny of the **Supplementary Figure S1** was used. Translation alignments were conducted in Geneious 9.1.8. Accession numbers of all viruses tested in ProP are listed in the **Supplementary Table S7**. Amino acid identity in translated 816 nt RdRp fragments was calculated with MEGA11 (Tamura et al., 2021) using the pairwise deletion option.

4.3.5 *Data collection of HCoV sequences for examining conservation of FCS in human sequences*

Datasets used for the ProP FCS prediction included only one human reference sequence. However, conservation of FCS in human-derived sequences was examined as follows. Blastp searches of partial spike sequences flanking the S1/S2 site (SARS-CoV, NC004718, positions (pos.) 23,139-24,191; HCoV-229E, NC002645, pos. 21,923-22,954; HCoV-NL63, NC005831, pos. 22,329-23,414; HCoV-OC43, NC006213, pos. 25,524-26,675; HCoV-HKU1, NC006577, pos. 24,823-25,977 and MERS-CoV, NC019843, pos. 23,358-24,449) was performed using the *nr* database on 3 January 2022. To restrict results to human-derived sequences only, the blast search was restricted to the taxonomy IDs 111137 (HCoV-229E), 290028 (HCoV-HKU1), 1335626 (MERS-CoV), 277944 (HCoV-NL63) and 31631 (HCoV-OC43). In the case of SARS-CoV, the species *Severe acute respiratory syndrome-related coronavirus* (taxonomy ID 694009) was chosen and non-human viruses were excluded according to their taxonomy IDs (1283333, 2709072, 1283332, 442736, 349342, 349343, 347537, 347536, 349344, 1508227, 1415834, 1415851, 1415852, 1699360, 1699361, 285945, 698398, 1487703, 1503296, 1503299, 1503300, 1503301, 1503302, 1503303, 2042697, 2042698, 2697049). Blastp search resulted in 157, 60, 378, 102, 348, 94 hits for HCoV-229E, HCoV-HKU1, MERS-CoV, HCoV-NL63, HCoV-OC43 and SARS-CoV respectively. Full records were downloaded and manually curated in Geneious 9.1.8, excluding experimental infections, clones, genetically modified

sequences as well as sequences only partially covering the reference sequence resulting in a final dataset encompassing 104, 58, 180, 55, 264 and 40 protein sequences for HCoV-229E, HCoV-HKU1, MERS-CoV, HCoV-NL63, HCoV-OC43 and SARS-CoV respectively. Proteins sequences were aligned in Geneious 9.1.8 using *Mafft* (Kato et al., 2002) with an auto algorithm and a BLOSUM62 scoring matrix. Conservation of the FCS was examined manually.

Due to the immense amount of SARS-CoV-2 available sequences, the metadata of all available sequences in GISAID on 11 January 2022, 10:55 am CET were downloaded. Only metadata of sequences fulfilling the *complete* and *high* coverage criteria in GISAID were used, resulting in a final dataset encompassing metadata of 4,824,313 sequence entries. A custom python script was used to extract GISAID sample IDs of sequences with complete or partial deletions within the SARS-CoV-2 FCS motif (amino acid positions/motif ${}_{682}\text{RRAR}_{685}$) or of sequences which had at least one amino acid substitution at position R682 or R685, which all likely result in a non-functional FCS. These criteria matched 304 SARS-CoV-2 sequences, which were all of human origin. Sequences were downloaded from the GISAID database and aligned to the Wuhan reference strain MT019529 in Geneious 9.1.8 using *Mafft*.

4.4 RESULTS

4.4.1 Taxonomic classification of European bat SARS-related coronaviruses

Stored original fecal samples from four horseshoe bat species (*Rhinolophus hipposideros*, *R. euryale*, *R. ferrumequinum* and *R. blasii*) sampled in Italy, Bulgaria, Spain, and Slovenia during 2008-2009 were re-accessed (Balboni et al., 2011; Drexler et al., 2010; Muth et al., 2018; Rihtaric et al., 2010). Specimens were screened for the presence of coronaviral RNA by using a nested reverse transcription-PCR (RT-PCR) amplifying 455 bp fragments of the RdRp gene (chapter 4.3.2). Nine unique coronaviruses were chosen and an 816 base pair fragment of the viral RdRp gene was amplified (chapter 4.3.2) in addition to the bat-associated BG31 SrC that has been previously described (Drexler et al., 2010; Muth et al., 2018).

In a previous proposal that tentatively classified CoVs into RdRp-grouping units (RGU) which are predictive of species classification, viruses within the species *SARS-related coronavirus* showed maximal 6.3% difference on the amino acid level in this sequence fragment (Drexler et al., 2010). The translated amino acid sequence identity between the European bat SrC and SARS-CoV/SARS-CoV-2 within this fragment ranged between 96.7-99.3% (**Table 6**), suggesting that all ten bat coronaviruses can be taxonomically classified within the coronavirus species *SARS-related coronavirus*.

Table 6. Amino acid identities of a partial RdRp fragment of European bat coronaviruses and prototype betacoronaviruses for species delineation.

% amino acid sequence identity (range):					
Within European bat CoVs	European bat CoVs compared to				SARS-CoV* Vs. SARS-CoV-2*
	SARS-CoV* (AY274119)	SARS-CoV-2* (MT019529)	Hp-BetaCoV (NC025217)	MERS-CoV* (NC019843)	
97.8-100	97.4-99.3	96.7-97.8	82.4-82.7	75.4-75.7	98.2

*For clarity of presentation, only one reference strain of SARS-CoV and SARS-CoV-2 as representatives of the species *SARS-related coronavirus*, Hp-Betacoronavirus as representative of the species *Bat Hp-Betacoronavirus Zhejiang2013* and MERS-CoV as a representative of the species *Middle East respiratory syndrome-related coronavirus* was included in the analysis.

4.4.2 *Phylogenetic analysis of the European bat SARS-related coronaviruses*

SARS-CoV-2 harbours a functional FCS at the boundary between the S1/S2 spike subunits (Coutard et al., 2020). The partial or full spike-encoding genomic region of European bat SrC containing this S1/S2 genomic region was next amplified (chapter 4.3.2) and subsequently sequenced. In a representative partial S2-based phylogeny that covered the complete genetic diversity of known SrC, European bat-associated SrC formed a sister clade to South-East Asian bat-associated SrC (**Figure 18, Supplementary Figure S1**).

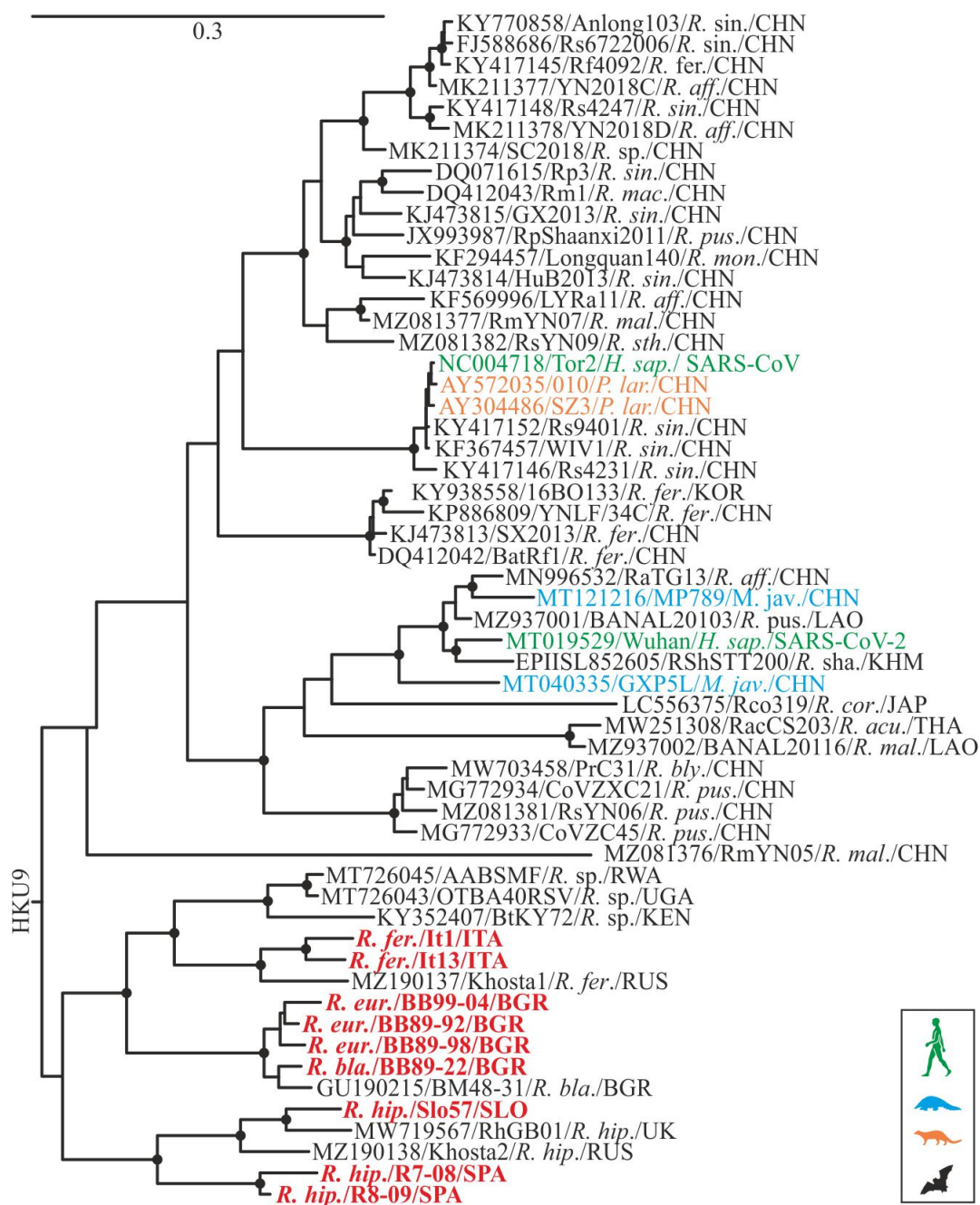


Figure 18. Evolutionary relationships of SARS-related coronaviruses. Phylogeny based on a partial S2-fragment (495 nt) of representative sequences of the complete diversity of SrC shown in Supplementary Figure S1. European horseshoe bat SrC generated within this study are shown in bold red and without accession numbers. Sequences are named as followed: GenBank accession number/strain name/host species/country of detection. Circles at nodes indicate support of grouping in $\geq 90\%$ of 1,000 bootstrap replicates. Scale bar represents nucleotide substitutions per site. H. sap., Homo sapiens; M. jav., Manis javanica; P. lar., Paguma larvata; R. acu., Rhinolophus acuminatus; R. aff., Rhinolophus affinis; R. bla., Rhinolophus blasii; R. bly., Rhinolophus blythi; R. cor., Rhinolophus cornutus; R. eur., Rhinolophus euryale; R. fer., Rhinolophus ferrumequinum; R. hip., Rhinolophus hipposideros; R. mac., Rhinolophus macrotis; R. mal., Rhinolophus malayanus; R. mon., Rhinolophus monoceros; R. pea., Rhinolophus pearsonii; R. pus., Rhinolophus pusillus; R. sha., Rhinolophus shameli; R. sin., Rhinolophus sinicus; R. sp., Rhinolophus species; R. sth., Rhinolophus stheno; BGR, Bulgaria; CHN, China; ITA, Italy; JPN, Japan; KEN, Kenya; KHM, Cambodia; KOR, Korea; LAO, Laos; RUS, Russia; RWA, Ruanda; SLO, Slovenia; SPA, Spain; THA, Thailand; UGA, Uganda; UK, United Kingdom.

4.4.3 *Sequence comparison of the S1/S2 genomic region in SARS-related coronaviruses*

Next, the genomic region encompassing the S1/S2 subunit boundary was compared between the diversity of known SrC.

SARS-CoV-2 is unique among the species *SARS-related coronavirus* in containing a polybasic FCS between the S1 and S2 subunits of the viral spike glycoprotein. While in bat-associated SrC from Asia and Africa remnants of single amino acid of a potential FCS motif can be observed, six bat-associated SrC from Spain, Slovenia and Bulgaria harbour near-complete sequence motifs (R-X-R) that partially resemble the atypical polybasic FCS of SARS-CoV-2 at the S1/S2 boundary (**Figure 19**).

4.4.4 *Furin cleavage sites in animal homologues of human coronaviruses*

The results up to this point indicate that despite incomplete FCS motifs, no other member of the *SARS-related coronavirus* species harbours an intact FCS at the S1/S2 boundary. The emergence of the furin cleavage site during the genealogy of SARS-CoV-2 thus remains unclear. To better understand the evolution of furin cleavage at the S1/S2 site, the S1/S2 genomic region within HCoVs and their animal homologues was examined for the existence of potential FCS. Therefore, complete datasets of the subgenera *Duvinacovirus* (HCoV-229E), *Merbecovirus* (MERS-CoV), *Setracovirus* (HCoV-NL63) and *Embecovirus* (HCoV-HKU1 and HCoV-OC43) were generated (chapter 4.3.4 and 4.3.5.). Conservation of FCS at the S1/S2 boundary was examined manually and potential motifs were checked by using a furin cleavage prediction algorithm (ProP).

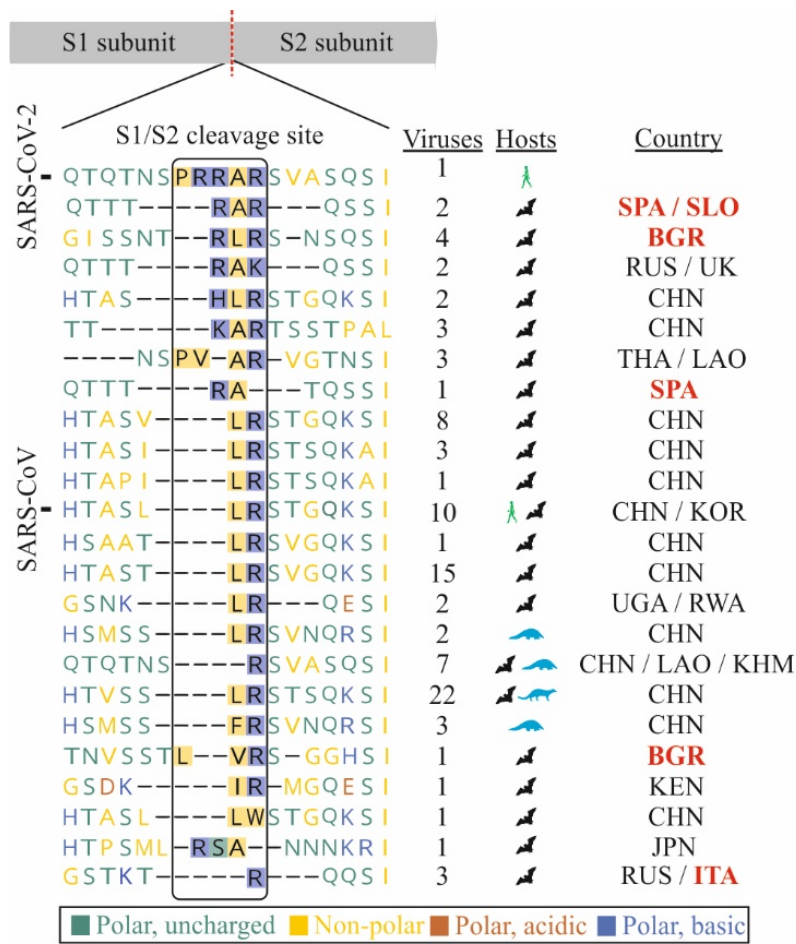


Figure 19. Comparison of the S1/S2 genomic region in SARS-related coronaviruses. A scheme of the spike protein and its subunits S1 and S2 shows the position of the furin cleavage site (FCS) in SARS-CoV-2. Amino acid residues of the polybasic FCS in SARS-CoV-2 are highlighted within the box. European horseshoe bat SrC sequences generated within this study are indicated by country names in bold red. BGR, Bulgaria; CHN, China; ITA, Italy; JPN, Japan; KEN, Kenya; KHM, Cambodia; KOR, Korea; LAO, Laos; RUS, Russia; RWA, Rwanda; SLO, Slovenia; SPA, Spain; THA, Thailand; UGA, Uganda; UK, United Kingdom.

Within the animal homologues of bat-associated human coronaviruses, a furin cleavage motif in the S1/S2 genomic region only exists in MERS-related coronaviruses. Here, an FCS motif was present in 10/27 (37%) of bat-associated and within all camel-associated MERS-related viruses (**Figure 20**). In six out of those 10 bat-associated MERS-related viruses this motif followed the canonical R-X-K/R-R motif. This motif in the remaining four bat- as well as all the camel- and human-derived sequences however, followed the R-X-X-R motif that allows furin cleavage in the human MERS-CoV (**Supplementary Table S8**) (Millet and Whittaker, 2014). Within the two HCoVs likely originating from a rodent reservoir, HCoV-OC43 and HCoV-HKU1, a canonical R-X-K/R-R at the S1/S2 site was present in all human-related animal CoVs, including all viruses known from the likely rodent reservoir.

The hypothetical turnover of a functional FCS in the animal reservoir and fixation after the host switch is reminiscent of the avian Influenza A virus (AIV), which is a prototypic example of an

RNA virus gaining pathogenicity via acquisition of an FCS. AIVs can be distinguished into low-pathogenic avian influenza virus (LPAI) and high-pathogenic avian influenza virus (HPAI) but only HPAs are defined by the existence of a polybasic FCS. A major determinant of the conversion from an LPAI into an HPAI virus variant within the reservoir or the new host is the acquisition of an FCS in the hemagglutinin (HA) protein (Lee et al., 2017b; Monne et al., 2014). This polybasic HA cleavage site can be introduced through three different molecular mechanisms: recombination with cellular or other RNA molecules, multiple nucleotide insertions, or nucleotide substitutions (Lee et al., 2017b; Nao et al., 2017; Suarez et al., 2004). Both insertions and nucleotide substitutions in the HA gene of AIVs are facilitated by a stem-loop secondary RNA structure enclosing the FCS and a high adenine/guanine content in the external loop structure (Nao et al., 2017).

Thus, it was next to be tested whether genomic surrogates of these mechanisms present in AIVs which may facilitate the acquisition of an FCS are given in the European bat-associated SrC. Therefore, RNA secondary structures of the genomic regions enclosing the FCS of the European bat-associated SrC were modelled (chapter 4.3.4), and genetic mutations or recombination were *in-silico* inserted to the sequences to see if the generation of an FCS may occur.





























	Reservoir		Related CoVs		Human	
	FCS	S1/S2	S1/S2	S1/S2	S1/S2	S1/S2
SARS-CoV	79 	–	12 	–	40 	–
SARS-CoV-2		–	6 	–	* 	+ [#]
MERS-CoV	27 	10	241  3 	+ –	180 	178
HCoV-NL63	3 	–	?		55 	–
HCoV-229E	6 	–	33 	–	104 	–
HCoV-OC43			97  4  1  1  4  13  2  9  1 	+ –	264 	262
HCoV-HKU1			?		58 	+

Figure 20. Furin cleavage sites at the S1/S2 interface in human and related animal coronaviruses. Predicted furin cleavage sites with Propp scores are shown in Supplementary Table S7. If not all tested sequences of one host category were predicted to contain an FCS, counts are given. *, 4,824,313 SARS-CoV-2 GISAID sequence entries; #304 SARS-CoV-2 sequences showing a loss of the RRAR furin cleavage site motif.

4.4.5 Potential acquisition of furin cleavage through nucleotide exchange or insertion

In accordance with the generally high level of secondary structures within coronavirus genomes (Simmonds, 2020) three European bat-associated SrC predictively contain RNA secondary structures similar to HPAI sequences within the region corresponding to the S1/S2 genomic region in SARS-CoV-2 (**Figure 21B** and **Supplementary Figure S2** and **Supplementary Figure S3**). Within this stem-loop structure, one European SrC (termed It1) showed a more than 60% adenine/guanine content. Both determinants together may facilitate an insertion of an adenine/guanine stretch and could thus accelerate the acquisition of a S1/S2 FCS motif within this virus (R-K-T-R; ProP cleavage site score of 0.69) (**Figure 21A and B**

and Supplementary Figure S2A) comparable to the insertion leading to HPAI outbreaks in the US in 2016/2017 (Lee et al., 2017a; Lee et al., 2017b).

Next, a single non-synonymous nucleotide substitution (a C to G transversion) in the external loop of the RNA secondary structure would already suffice to create a motif resembling the SARS-CoV-2 R-R-X-R FCS motif in two European bat-associated SrC (T-R-L-R to R-R-L-R in BB99-04 and BB89-98). Using a furin cleavage prediction algorithm (ProP), both of these motifs are likely to be cleaved by furin (ProP cleavage site scores of 0.65 and 0.52 for BB99-04 and BB89-98, respectively; the program's threshold is 0.5) (**Figure 21A and B and Supplementary Figure S2A and Supplementary Figure S3B**).

4.4.6 *Detection of single-nucleotide variants by deep sequencing*

The previous analyses suggest that an FCS in some of the European bat-associated SrC may naturally be acquired through single nucleotide substitutions. To explore if a viral quasispecies harbouring this mutation exist within the bat host, PCR amplicons of the S1/S2 genomic regions of these SrC were subjected to ultra-deep sequencing (chapter 4.3.2). NGS read coverage at the genomic position was sufficiently high with read counts between 147,459 and 117,192 reads at the corresponding S1/S2 genomic sequence position in the two viruses BB99-04 and BB89-89, respectively. The C to G transversion within the genome position leading to a non-synonymous nt exchange generating an FCS motif in the two viruses was present in 0.004% and 0.006% of the total reads, respectively (**Table 7**).

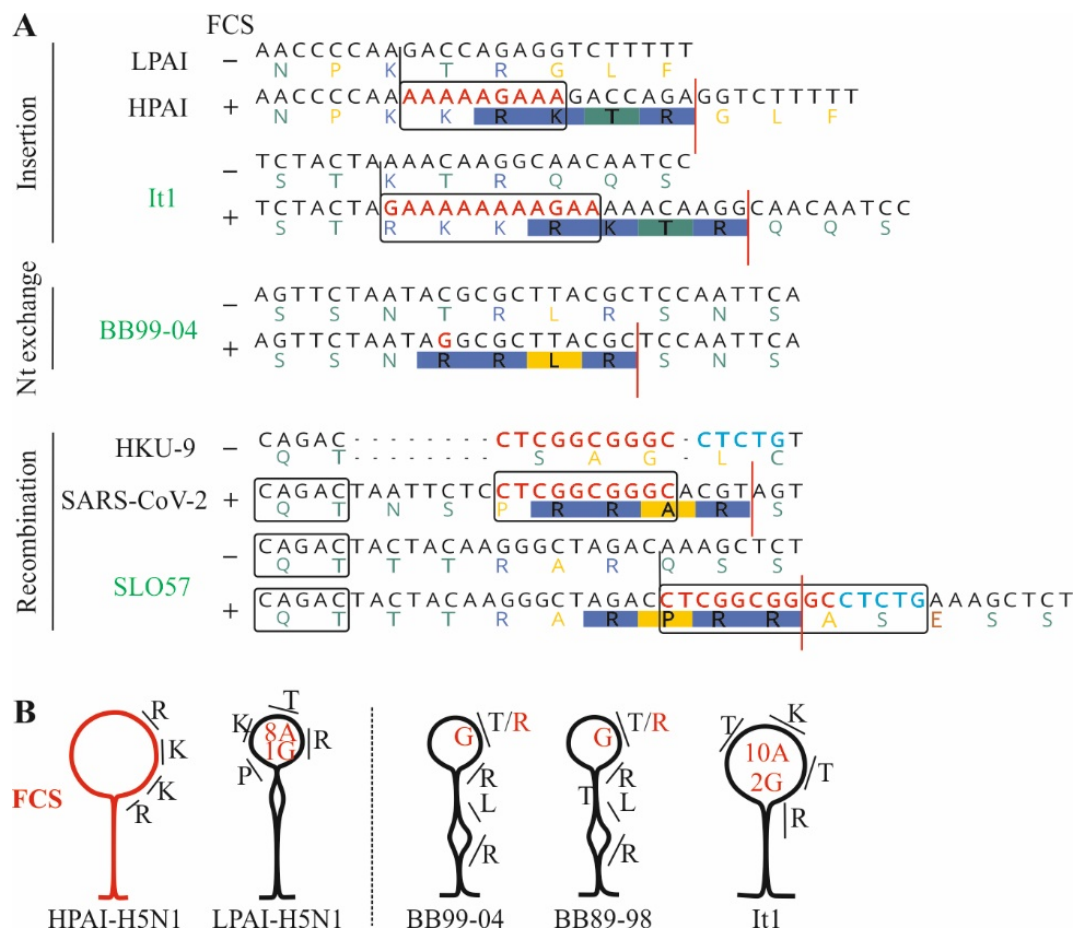


Figure 21. Predictive acquisition of furin cleavage sites in European bat SARS-related coronaviruses through different molecular mechanisms. (A) Potential generation of FCS in different European bat SrC by nucleotide exchange, insertion due to external stem loop structures or recombination with HKU9 (15 nucleotides, 10 of which occur identically in SARS-CoV-2). Sequences of BB99-04 and It1 were trimmed in the figure for graphical reasons to highlight the FCS. The complete sequences used for FCS prediction were AKYGISSNRRLRSNSQSIV for BB99-04 and AKFGSTRKKRKTRQQSIL for It1. (B) Simplified predicted RNA secondary structures of the polybasic cleavage site regions of AIV and SrC that acquire FCS through nucleotide substitutions or insertions. Amino acids corresponding to codons forming the FCS are shown. Nucleotides and corresponding amino acids that make a change to an FCS through nucleotide substitutions or insertions are marked in red. Corresponding structures showing complete nucleotide sequences are shown in Supplementary Figure S2.

Table 7. Single nucleotide variants within two European SrC at spike position Thr672§.

BB99-04	Consensus sequence	A	C	G
	Corresponding amino acid	Thr (T)		
	A (%)	146506 (99.801)	8 (0.005)	159 (0.108)
	Corresponding amino acid	Thr (T)	Lys (K)	Thr (T)
	T (%)	19 (0.013)	187 (0.127)	13 (0.009)
	Corresponding amino acid	Ser (S)	Met (M)	Thr (T)
	C (%)	7 (0.005)	147250 (99.858)	10 (0.007)
	Corresponding amino acid	Pro (P)	Thr (T)	Thr (T)
	G (%)	259 (0.176)	6* (0.004)	146817 (99.872)
	Corresponding amino acid	Ala (A)	Arg (R)	Thr (T)
	Total reads	146798	147459	147005
BB89-98	Consensus sequence	A	C	A
	Corresponding amino acid	Thr (T)		
	A (%)	117032 (99.593)	1 (0.001)	116893 (99.745)
	Corresponding amino acid	Thr (T)	Lys (K)	Thr (T)
	T (%)	47 (0.0400)	38 (0.032)	13 (0.011)
	Corresponding amino acid	Ser (S)	Ile (I)	Thr (T)
	C (%)	10 (0.009)	117540 (99.961)	8 (0.007)
	Corresponding amino acid	Pro (P)	Thr (T)	Thr (T)
	G (%)	421 (0.358)	7## (0.006)	278 (0.237)
	Corresponding amino acid	Ala (A)	Arg (R)	Thr (T)
	Total reads	117510	117586	117192

§Spike position Thr672 corresponds to the European SrC BB99-04 (GenBank Acc. No. KR559017).

#One of the reads was not paired-end

*This single nucleotide variant leads to a non-synonymous exchange generating an FCS motif in this virus

4.4.7 Potential acquisition of furin cleavage through recombination

The acquisition of multibasic cleavage sites by recombination and the determinants thereof in AIV remain unclear (Boni et al., 2008), but may be facilitated by host nucleolar RNAs (Gulyaev et al., 2021). In coronaviruses however, it is a common microevolutionary mechanism to exchange genetic material (Drexler et al., 2014). For the origin of the S1/S2 FCS of SARS-CoV-2 for example, it is speculated that it may have arisen by recombination with other bat coronaviruses such as HKU9, likely facilitated by short palindromic sequences within the SARS-CoV-2 genome serving as breakpoint signals (Gallaher, 2020). Thus, it was finally investigated whether European bat-associated SrC may acquire an FCS at the S1/S2 boundary by this mechanism as well.

Therefore, the sequences of the European bat SrC were checked manually for palindromic sequences and one European SrC (termed SLO57) was found to harbour the palindromic sequence CAGAC close to the S1/S2 boundary. Subsequent *in-silico* insertion of the HKU9-derived sequence, that has been found in SARS-CoV-2 (Gallaher, 2020) 5' to the palindromic sequence resulted in a predictively functional FCS motif (R-Q-S-S to R-P-R-R-A-S-E-S-S including the corresponding HKU9-derived sequence; ProP cleavage site score of 0.73) (**Figure 21A**). This suggests that recombination between the European bat SrC and other bat coronaviruses could result in the acquisition of an FCS.

4.5 DISCUSSION

SARS-CoV-2 contains an FCS at the S1/S2 site of the spike protein that is likely the reason for its high transmissibility between humans (Hoffmann et al., 2020a). The FCS in SARS-CoV-2 is unique within the *SARS-related coronavirus* species for which reason speculations concerning the origin of this novel CoV arose, including hypotheses of a laboratory-leak source. However, the evolution and potential natural acquisition of such FCS in CoVs has not been studied to date. The data presented in this chapter are the results of the first comprehensive *in-silico* assessment of whether and how FCS may evolve in bat-associated SrC and reveal several discoveries that likely support a natural origin of SARS-CoV-2 and its FCS.

First of all, the results in chapter 4.4.4 revealed that FCS naturally occur at the spike S1/S2 genomic site in other HCoVs and some of their animal homologues. These data are consistent with previous studies (Liu et al., 2021; Stout et al., 2021; Wu and Zhao, 2020) showing that FCS at the S1/S2 interface occurred independently for several times during the evolution of CoVs, substantiating the insertion of an FCS as a natural process of evolution. However, the existence of FCS in some bat MERS-related CoVs but in none of the other bat homologues of bat-associated HCoVs suggest that an FCS may not be selected in bat hosts, as compared to viruses observed in humans or intermediate hosts. In humans, furin is a ubiquitously expressed membrane-bound protease localized to organelles of the constitutive protein secretory pathway, such as the Trans-Golgi network or endosomes and is also present at the cell surface and in extracellular space (reviewed in (Seidah et al., 2008)). Albeit little knowledge exists about furin protease expression in bats, previous research found that some fruit bats contain a furin protease homologue that is, at least, capable of cleaving the furin R-X-K/R-R motif in the parainfluenza virus fusion protein (El Najjar et al., 2015), presuming that a non-selection of FCS in bat CoVs is not due to a lack of furin-like enzymes in these hosts. However, differences in cellular localization and activity of the furin proteases in humans and bats potentially exist and this may determine location and mode of entry of viruses that depend on furin-mediated glycoprotein activation. Notably, while replication of HCoVs primarily takes place in the respiratory tract, bat CoVs are enteric (Corman et al., 2014; Watanabe et al., 2010) and bats show limited signs of disease. However, clinical (Cholankeril et al., 2020) and experimental (Lamers et al., 2020; Zhou et al., 2020b) evidence imply also gastrointestinal infection and a faecal–oral transmission route of SARS-CoV-2 in humans. Therefore, it seems plausible that the ubiquitous furin expression in humans is likely the reason that SARS-CoV-2 can spread systemically and cause more severe disease than per example SARS-CoV. Likewise, SARS-CoV-2 effectively

replicates also in bat intestinal cells suggesting a potential infection of bats with SARS-CoV-2 or its precursor (Zhou et al., 2020b). Whether the intestinal infection in bats depends on furin-cleavage or not remains unclear and requires further investigation. In conclusion, the FCS seem to be a central virulence factor altering pathogenicity of SARS-CoV-2 but its role in bat SrC remains equivocal and requires further investigation.

That FCS are not typical in bat CoV spike proteins stands in stark contrast to the evolutionary conservation of FSC in the reservoir of rodent-borne HCoV_s (Stout et al., 2021); recently proved by the discovery of a novel betacoronavirus from Chinese rats (Li et al., 2021). This finding potentially hints at differences in organ tropism and/or transmission routes between bat and rodent coronavirus reservoirs. Whether bat CoVs employ different host cell proteases for spike protein cleavage than humans or rodents or if spike protein cleavage in general is not essentially needed for successful infection of bat cells remains to be investigated in future studies. Those may include experiments with pseudotyped vesicular stomatitis virus (VSV) containing bat CoV spike glycoproteins or reverse genetic systems to investigate viral entry *in-vitro* including epithelial and primary airway cells from different bat and rodent species.

Secondly, the data presented within this chapter of this thesis suggest that acquisition of an FCS in European bat-associated SrC is conceivable through natural molecular mechanisms. Even if the mutational changes are hypothetical, they resemble prototypic molecular mechanisms leading to the generation of HPAI strains in their avian reservoir. It should not be disregarded that in AIV, the acquisition of an FCS is only one of several determinants in the transition from LPAI to HPAI. Other molecular changes that have been found in HPAI strains include for example deletions in the neuraminidase or additional mutations in the HA protein (Beerens et al., 2020). It is therefore likely that additional genetic changes would have to occur before European bat-associated SrC achieve overall higher fitness in a new host and acquire transmissibility via the respiratory tract for example. Cell entry via the host cell receptor ACE2 as well as spike protein activation by TMPRSS2 via the S2' cleavage site are both essential for SARS-CoV, and SARS-CoV-2 pathogenesis (Belouzard et al., 2009; Bertram et al., 2011; Hoffmann et al., 2020b; Li et al., 2003). The TMPRSS2 cleavage site KR|S in the S2' site seems to be highly conserved among bat SrC (Hoffmann et al., 2020a), indicating that the genomic prerequisites for cleavage by TMPRSS2 may be given already. Of particular interest in this context is, that the cleavage site at S2' can also be proteolytically activated by furin (Zhang et al., 2022), indicating that in the absence of TMPRSS2, SARS-CoV-2 infection may be alternatively mediated by furin. Comparable to the potential FCS acquisition through a single nucleotide exchange in some bat-associated SrC (chapter 4.4.5), a recent study revealed that

also ACE2 binding is readily evolvable in different bat-associated SrC through only single amino acid substitutions (Starr et al., 2022). The European bat SrC studied in this thesis were, however, not among these candidate viruses, suggesting greater mutational adaptation is required in these viruses before human ACE2 binding may be acquired. As coronaviruses recombine extensively, especially between spike genes (Vakulenko et al., 2021a), ACE-binding virus variants may be quickly generated by this mechanism for example (Xiong et al., 2022). In addition, it must be considered that the effect of single mutations, can vary greatly from virus to virus, as shown by the N501Y mutation, which increases the human ACE2-binding affinity of several SARS-CoV-2 variants but significantly decreases it in SARS-CoV (Liu et al., 2022; Starr et al., 2022). Such traits of bat SrC increasing their zoonotic potential beyond FCS acquisition thus merit additional investigation.

Finally, the data of this chapter showed the discovery of genomic features such as RNA secondary structures as well as purine-richness in the external loop structure of the S1/S2 genomic region in European bat SrC. Since such genomic features are genetic predispositions that accelerate insertions or substitutions of nucleotides in the HA gene of AIV, this result suggests the S1/S2 junction in bat SrC to be an evolutionary “hot spot” vulnerable for insertion events. This is further supported by the observation of an insertion of multiple amino acids at this site in another bat SrC from China (RmYN02) (Zhou et al., 2020a) and highlights once more the feasibility of a random but natural occurrence of an FCS.

The analyses performed in this chapter of the thesis come with some limitations. Firstly, the *in-silico* genetic modifications to test for FCS acquisition assumed that the FCS turnover happened in the bat reservoir and excluded possibilities of adaptation in intermediate hosts or humans directly. For the 2002 SARS-CoV, for example, it has been speculated that the mutational adaptation that improved binding to human ACE2 has not happened directly in the bat reservoir but only in the intermediate host, the palm civet (Graham and Baric, 2010; Song et al., 2005). A similar stepwise adaptation would be conceivable for SARS-CoV-2, but requires further investigation into virus discovery within potential intermediate hosts. Additionally, a direct spillover from bats to humans must be considered. A recent study identified bat SrC in China with the potential to infect humans (Temmam et al., 2022a) and acquisition of an FCS in the S protein through adaptive evolution during the initial human-to-human transmission chains seems another possible scenario. However, several indications such as (i) the high genetic diversity paired with frequent recombination within single bat colonies, (ii) the incomplete FCS motifs found in European bat SrC, and (iii) the occurrence of FCS motifs in some but not all

bat MERS-related CoVs, indicate that coronaviruses harbouring an FCS may evolve and exist within the reservoir gene pool and hypothesize that an FCS may be fixed / selected after switching hosts in analogy to the transition from low-pathogenic to high-pathogenic AIV.

Secondly, prediction and verification of FCS motifs in the bat SrC sequences were solely performed using one computational furin prediction algorithm (ProP). Better accuracy could have been achieved by validation through a second program, for example PiTou (Tian et al., 2012). However, since both programs are based upon different methods and no comparison of the tools has been performed due to the lack of an independent testing dataset, the results using both programs might not have been easy to compare. Transfection experiments could add to confirm that the experimentally generated FCS in the European bat SrC are substrates proteolytically processed by furin *in-vivo*.

Finally, by reaching out to find quasispecies circulating in nature, deep sequencing results detected single nucleotide variants (SNV) that delineate nucleotide exchanges generating an FCS motif in the two European bat-associated SrC at a comparatively low frequency (0.0004% and 0.006%) compared to other studies (Braun et al., 2021; Lythgoe et al., 2021) that only call variants with a frequency above 2-3%. Inferring viral genetic diversity using NGS technologies is rapid and cost-effective, however, this approach comes with relatively high error rates induced by for example reverse transcription of viral RNA into cDNA, PCR amplification or the NGS process itself (reviewed in (Beerenwinkel et al., 2012)). The detected SNV frequency was indeed within the error rate of viral polymerases used for cDNA synthesis and amplification (Orton et al., 2015) (**Table 7**). Entrainment of a control sample consisting of an amplicon with the same genomic region from a clonal sample or synthesized DNA throughout the complete experiment would have allowed to account for polymerase error and thereby distinguish biological variants from process errors. Furthermore, computational analysis tools developed to identify SNVs could have increased the power and accuracy of SNV calling substantially (Jayasundara et al., 2015; Posada-Cespedes et al., 2021). Nevertheless, SNVs affording the emergence of furin cleavage in LPAI were found at a comparatively low frequency (0.0028%) in influenza virus strains, which may imply the potential for emergence of FCS in those European bat SrC strains (Nao et al., 2017). These intermediate variants are to be discovered and may provide further insights into the viral diversity that evolves and exists in bats. However, it may be possible that these variants are not transmitted in the bat reservoir and go extinct before zoonotic transmission, as they may not be selectively advantageous in the bat

host as discussed before. Whether bat SrC quasispecies harbouring an FCS exist within European bat hosts thus requires careful examination.

In conclusion, the results of this part of the thesis at hand present several possible ways for the natural acquisition of an FCS in SrC that circulate in bats. Even though the entire evolutionary path to the maintenance of the functional cleavage motif in SARS-CoV-2 remains presently elusive, the comprehensive results support a natural evolutionary origin of SARS-CoV-2 from bats with or without the involvement of intermediary hosts. As discussed, future studies of viral diversity in bats and potential intermediate hosts may identify other sarbecoviruses harbouring functional FCS. The zoonotic potential of such sarbecoviruses then deserves profound investigation in follow-up studies to identify variants potentially posing threats to human health.

CHAPTER 5

FINAL DISCUSSION AND OUTLOOK

In this thesis I presented a series of studies that shed light on the adaptation processes that shaped the evolutionary trajectory of emerging viruses before and after host switching. The application of computational and molecular virological methods for analysing comprehensive viral genomic sequence datasets enabled me to (i) reveal that MAVS-cleavage and a biased codon usage strategy may contribute to and facilitate cross-species transmissions of hepatoviruses (**chapter 2**), (ii) elucidate the adaptation of mutations of concern in different SARS-CoV-2 lineages circulating in a geographic region with intense transmission (**chapter 3**) and (iii) reveal that furin cleavage sites in SARS-related coronaviruses can be acquired in the bat reservoir via conserved molecular mechanisms, supporting a natural origin of SARS-CoV-2 (**chapter 4**).

In the next sections, I will outline the application of the findings identified in the studies for future research and action for pandemic preparedness strategies.

5.1 LOOKING FORWARD: PROSPECTIVE ACTIONS FOR PANDEMIC PREVENTION

5.1.1 Animal reservoir and population surveillance

The present work identified a number of hepatoviruses and SARS-related coronaviruses that already have or may acquire genetic attributes that may facilitate cross-species transmissions in the future, including into humans. Especially in the tropics, where humans hunt and consume both bats and rodents for nutritional reliance (Chausson et al., 2019; Schulte-Herbruggen et al., 2013), exposure to these pathogens with zoonotic potential is critically high. Appropriately tailored prevention activities should thus include (i) surveillance of people at risk and animals of frequent contact (ii) educational-driven strategies to increase awareness of zoonotic spillover in the communities and (iii) development-based approaches through the provision of bushmeat substitutes to decrease demand on small mammals (Bachmann et al., 2020). Contemporaneously, governments must address the problems that contribute to the rise of emerging virus infections at source by for example reducing drivers of climate change (Mora et al., 2022).

Comparison of FCS occurrence in bat- and rodent counterparts of human CoVs revealed that FCS are not typical in bat CoVs spike proteins but dominant in rodent-associated sequences. This result suggests a possible selective advantage of CoV spike cleavage by furin in rodents. Investigating the role of furin cleavage in rodent-associated CoVs in future studies may provide important knowledge about the role of FCS in zoonotic transmissions. Furthermore, this result implies that cross-species transmissions of CoVs harbouring FCS from bats to rodents should be considered possible. In this sense, studies investigating potential bridging hosts of SARS-CoV-2 should include rodents (Montagutelli et al., 2022) and could thus bring hidden hosts or the yet undiscovered intermediate host of SARS-CoV-2 to light (Hale et al., 2021; Peacock et al., 2022).

Among the >6,400 extant mammalian species, about 60% are rodents or bats (Burgin et al., 2018). Both mammalian orders have played a major role in transmission of severe pathogens, exemplified by epidemics, such as COVID-19 (Latinne et al., 2020), SARS (Ge et al., 2013), Ebola (Swanepoel et al., 1996), Marburg (Towner et al., 2009), Nipah (Drexler et al., 2012; Eaton et al., 2006), hantavirus cardiopulmonary syndrome (Nunez et al., 2014) and plague (Nguyen et al., 2018). To mitigate future emerging zoonotic diseases that may end in devastating epi- or pandemics, further virus discovery and enhanced understanding of virus evolution in these two mammalian orders is a meaningful task (Mollentze and Streicker, 2020).

Beyond the SARS-CoV-2 pandemic, recent public health emergencies imply that current surveillance strategies need to be reconsidered. The recent discovery of porcine Deltacoronavirus in children in Haiti (Lednicky et al., 2021) for example; a genus of coronaviruses that has not yet caused zoonotic outbreaks, suggests that also viruses outside of the to-date human-associated virus groups should be put to the agenda for virus surveillance and risk assessment studies. Furthermore, the recent monkeypox outbreaks in several previously non-endemic countries (Isidro et al., 2022) highlight that only continuous strategies monitoring genetic changes of known viruses can help to prevent “old” viruses from re-emerging in the human population.

5.1.2 Collection and sharing of research data

In terms of risk assessment of newly identified pathogens, one of the main challenges arising is to combine and correlate sequencing data together with data of comprehensive risk assessments. Novel tools employ artificial intelligence to predict the emergence or risk of potentially zoonotic viruses (Becker et al., 2022; Mollentze et al., 2021). These programs rely on genomic

sequencing and primary research data on genomic signatures that enable host switches, giving the research data obtained within this thesis another, yet new, application. However, the implementation and performance in practice remains to be determined.

Sharing the obtained data is subsequently key to inform about newly identified pathogens that may have pandemic or epidemic potential. However, filtering the mass of information on newly discovered wildlife viruses and studies about their zoonotic potential in order to timely identify those pathogens that pose the greatest threat to global public health is one big challenge scientists as well as governments face. To this end, the opensource platform “SpillOver” (<https://spillover.global/>) provides a risk-ranking of viruses prioritized for surveillance based on data entered by experts aiming at free information access and exchange not only for the scientific community (Grange et al., 2021).

Free and timely access to these information should be prioritized as interdisciplinary collaboration and exchange among researchers will become more and more important, with the goal to develop and implement One Health solution strategies to achieve healthy ecosystems, animals, and people (One Health High-Level Expert et al., 2022).

5.1.3 *Global genomic surveillance strategy*

The crucial role genomic sequencing plays in pandemic preparedness and response outlines the need for a global genomic surveillance strategy that assures that each country can generate, analyse, and interpret genomic data in order to react quickly in case of a public health emergency (Wilkinson et al., 2021). Undeniably, the advances in NGS technologies, software tools, data-storage systems and development of global databases that have been achieved over the last decade have made genomic surveillance much easier, faster and (cost) efficient (Tegally et al., 2022). However, while being standard practice in affluent global regions such as North America and Europe, genomic sequencing is, even after substantial capacity building during the SARS-CoV-2 pandemic, still not accomplishable in one out of three countries (WHO, 2022c). Especially resource-limited countries deal with financial and logistical challenges to undertake genome sequencing independently (Inzaule et al., 2021). As of March 28, 2022, the WHO therefore released a global genomic surveillance strategy with the key message “By 2032, all 194 WHO Member States have, or have access to, timely genomic sequencing for pathogens with pandemic and epidemic potential” (WHO, 2022b). It is indeed the first step to build and strengthen capacities as the strategies objectives strive to improve access to laboratory and computational technology, provide education and training in the use of (new) software as well

as facilitate (inter)national networking of researchers for information sharing and data exchange. However, profound challenges remain with respect to lack of infrastructure such as stable power supply, internet access or data storage capacities, deficient supply guarantee of materials, service and device maintenance due to poor manufacturer infrastructures outside affluent countries (Kanteh et al., 2022). As these are fundamental prerequisites for the successful implementation of the global genomic surveillance strategy, regional and international stakeholders, and donors as well as the manufacturers are called upon to take immediate action to address these shortcomings (Brito et al., 2022).

Box 2. Future tasks to prevent emergence of zoonotic viruses

Society

- Increase public awareness globally about how to prevent zoonotic transmissions More precisely, educate vulnerable societies about
 - o the risk of zoonotic spillover infections by consumption of bushmeat such as rodents and bats
 - o the significance of surveillance strategies to early identify spillover events

Research

- Intensify virus discovery in animal reservoirs and alternative materials such as sewage or animal waste, also with help of artificial intelligence
- Enhance identification and genetic as well as phenotypic characterization of genetic signatures that facilitate zoonotic transmissions of viruses

Policy

- Strengthen genomic sequencing capacities globally
- Intensify/Enhance scientific communication to the society, interdisciplinary and policymakers
- Involve and correlate virological and socio-ecological data to identify vulnerable groups for zoonotic spillover events
- Reduce animal-human contact (meat/bushmeat, animal farming etc) with the help of other stakeholders/authorities

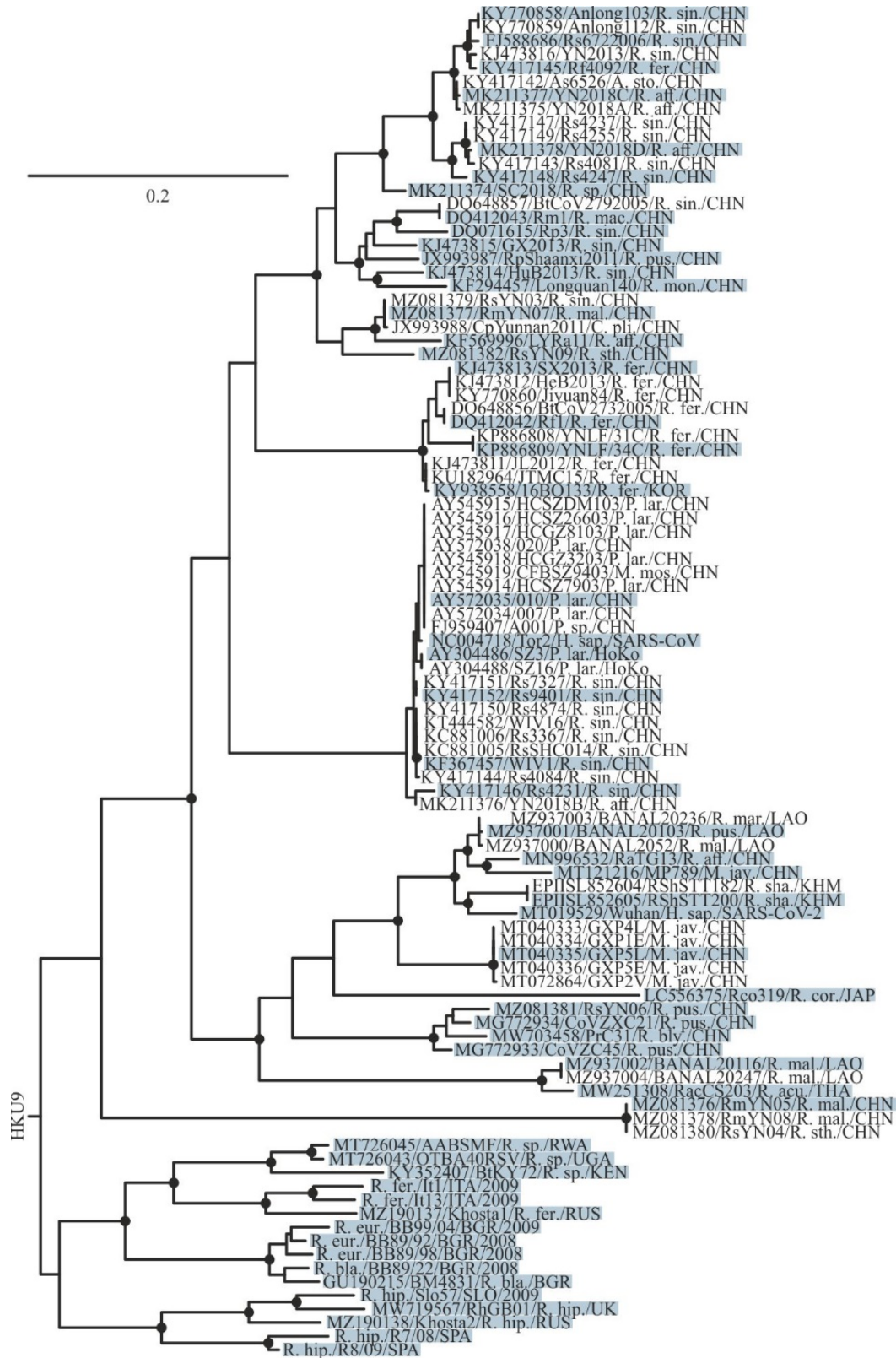
5.2 CONCLUDING REMARKS

The emergence and re-emergence of zoonotic viruses causing severe outbreaks will continuously pose a threat to global health in the future (Baker et al., 2022). Scientists warned of the emergence of a new bat-borne coronavirus outbreak since years already (Hu et al., 2017). However, the SARS-CoV-2 pandemic could not be prevented. The timely detection of naturally circulating viruses with pandemic potential prior to their outbreak remains one of the main challenges of our time. Therefore, not only virus discovery but detailed genomic characterizations of the viral genetic diversity circulating in animal reservoirs are of fundamental importance (Mollentze et al., 2021) and part of recently published action plans to better prepare for future pandemics (Keusch et al., 2022). Filling gaps in knowledge about the role of prior adaptation in the animal host and its responsibility for cross-species transmissions, or molecular signatures of viral genotypes with pandemic potential, has been of multiple request (Drexler et al., 2010; Hu et al., 2017; Ruiz-Aravena et al., 2022). Therefore, the results of the diverse studies performed in this thesis at hand, revealing evolutionarily conserved host-defense mechanisms as well as genetic characteristics associated with host tropism including MAVS or furin cleavage, provide highly important insights into the evolution and genetic adaptation of hepato- and coronaviruses before and after the host switch into humans.

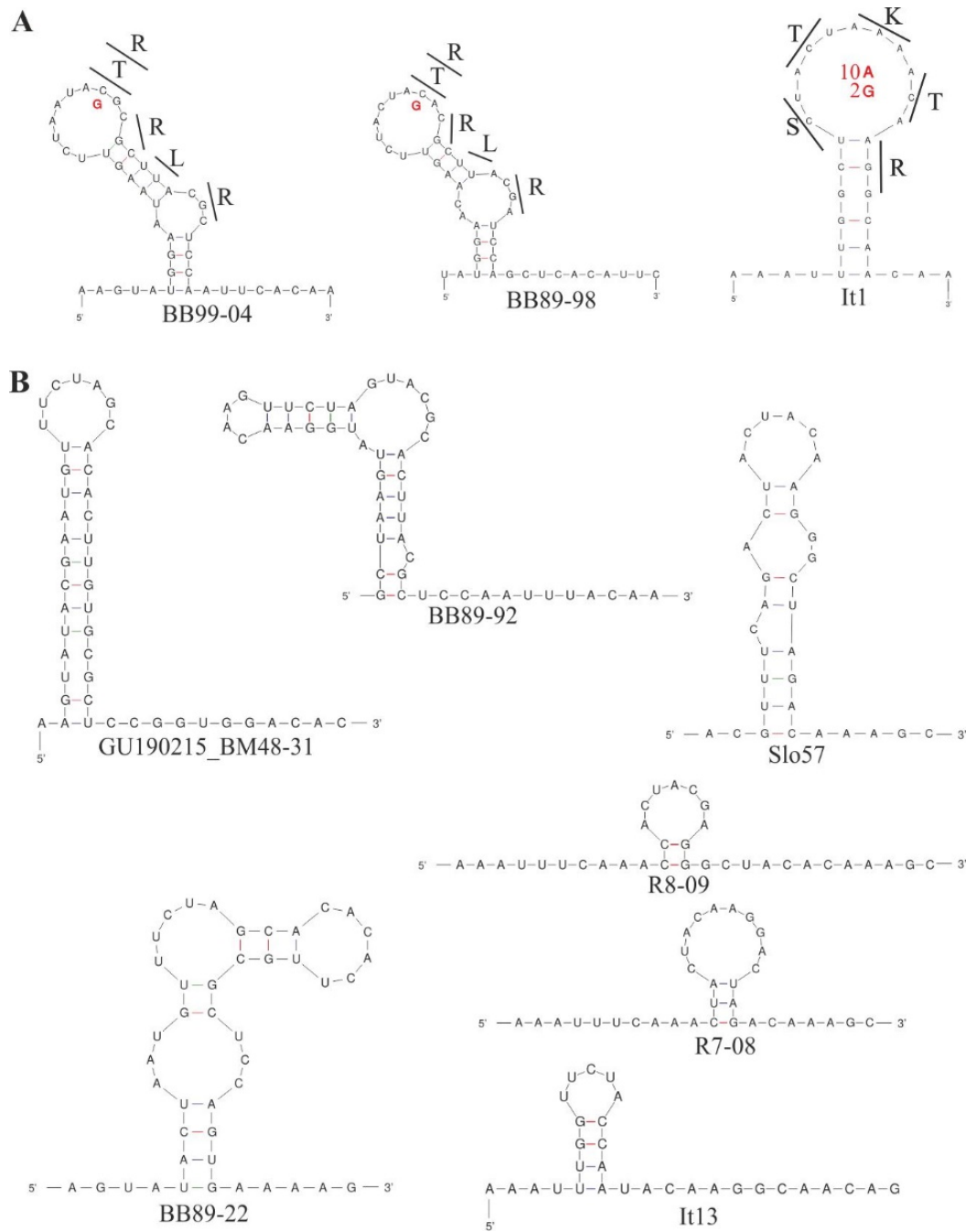
Beyond that, the comprehensive results obtained within this thesis underline the solid foundation of information that genomic surveillance coupled with evolutionary and genetic risk assessment can provide for subsequent phenotypical characterizations. For instance, primary identification of amino acid residues to be prioritized for experimental assessments is highly beneficial (Rasche et al., 2019a; Schroeder et al., 2021), because virus isolation of reservoir-borne viruses and establishment of reverse genetic systems are limited by low rates of success or high expenditure of time, respectively, making it unfeasible to check every individual amino acid of viral genes. The *in-silico* prediction of mutations of interest or genomic risk markers can thus facilitate and promote follow-up *in vitro* studies tremendously by saving time and consumables and additionally contributes to the 3R principles (Russell and Burch, 1959), to avoid poorly informative animal experiments. In sum, sequence-based genomic characterizations of human and animal-associated viruses can brief risk assessment studies at the earliest stage of viral surveillance and provide the fundamental basis to confer pandemic preparedness substantially.

SUPPLEMENTARY INFORMATION

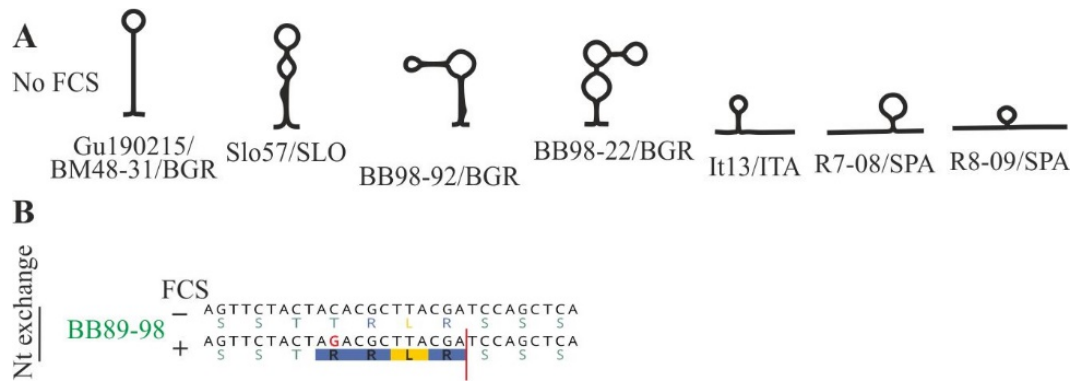
Figures



Supplementary Figure S1. Evolutionary relationships of SARS-related coronaviruses. Approximately-maximum-likelihood phylogeny showing the complete diversity of members of the species *SARS-related coronavirus*. The final dataset comprised 99 sequences (2 human, 12 civet, 6 pangolin and 79 bat-associated SrC). Sequences are named as followed: GenBank Acc. number/strain name/host species/country of detection. Circles at nodes indicate support of grouping in $\geq 90\%$ of 1,000 bootstrap replicates. Scale bar represents nucleotide substitutions per site. Sequences highlighted in blue were chosen for phylogenetic analysis shown in Figure 18, representing the complete diversity of SrC. *A. sto.*, *Aselliscus stoliczkanus*; *C. pli.*, *Chaerephon plicatus*; *H. sap.*, *Homo sapiens*; *M. jav.*, *Manis javanica*; *M. mos.*, *Melogale moschata*; *P. lar.*, *Paguma larvata*; *P. sp.*; *Paguma species*; *R. acu.*, *Rhinolophus acuminatus*; *R. aff.*, *Rhinolophus affinus*; *R. bla.*, *Rhinolophus blasii*; *R. bly.*, *Rhinolophus blythi*; *R. cor.*, *Rhinolophus cornutus*; *R. Eur.*, *Rhinolophus Euryale*; *R. fer.*, *Rhinolophus ferrumequinum*; *R. hip.*, *Rhinolophus hipposideros*; *R. mac.*, *Rhinolophus macrotis*; *R. mal.*, *Rhinolophus malayanus*; *R. mar.*, *Rhinolophus marshali*; *R. mon.*, *Rhinolophus monoceros*; *R. pea.*, *Rhinolophus pearsonii*; *R. pus.*, *Rhinolophus pusillus*; *R. sha.*, *Rhinolophus shameli*; *R. sin.*, *Rhinolophus sinicus*; *R. sp.*, *Rhinolophus species*; *R. sth.*, *Rhinolophus stheno*; BGR, Bulgaria; CHN, China; ITA, Italy; JPN, Japan; KEN, Kenya; KHM, Cambodia; KOR, Korea; LAO, Laos; RUS, Russia; RWA, Ruanda; SLO, Slovenia; SPA, Spain; THA, Thailand; UGA, Uganda; UK, United Kingdom.



Supplementary Figure S2. Detailed predicted RNA secondary structures of the polybasic cleavage site of European bat SARS-related coronaviruses. (A) European bat SARS-related CoVs that may acquire FCS through nucleotide substitutions or insertions (shown in main text Figure 4b). Amino acids corresponding to codons forming the potential FCS are shown. Nucleotides and corresponding amino acids that change to a predicted FCS through nucleotide substitutions or insertions are marked in red. **(B)** European bat SARS-related CoVs that are not predicted to acquire FCS through nucleotide substitutions or insertions.



Supplementary Figure S3. (A) Predicted RNA secondary structures of the S1/S2 genomic regions of bat SrC that are not predicted to acquire an FCS through nucleotide substitutions or insertions. (B) Potential generation of FCS for bat SrC BB89-98 by nucleotide exchange. Sequence was trimmed in the figure for graphical reasons to highlight the FCS. The sequence used for FCS prediction was AKYGTSSSTRRLRSSSHSIV; ProP score was 0.52.

*Tables***Supplementary Table S1.** GenBank accession numbers of hepatovirus sequences used for codon usage analyses of primate and non-primate hepatoviruses in chapter 2.4.1.

Host order	Accession number
Primate	AB020564, AB020565, AB020566, AB020567, AB020568, AB020569, AB258387, AB279732, AB279733, AB279734, AB279735, AB300205, AB425339, AB618529, AB618531, AB623053, AB793725, AB793726, AB819869, AB819870, AB839692, AB839693, AB839694, AB839695, AB839696, AB839697, AB909123, AB969748, AB973400, AB973401, AF268396, AF314208, AF357222, AF485328, AF512536, AJ299464, AY644337, AY644670, AY974170, D00924, DQ646426, DQ991029, DQ991030, EF207320, EF406357, EF406358, EF406359, EF406360, EF406361, EF406362, EF406363, EU011791, EU131373, EU140838, EU251188, EU526088, EU526089, FJ227135, FJ360730, FJ360731, FJ360732, FJ360733, FJ360734, FJ360735, HM769724, HQ246217, HQ437707, JQ425480, JQ655151, K02990, KC182587, KC182588, KC182589, KC182590, KF569906, KF773842, KJ427799, KP177964, KP177965, KP879216, KP879217, KT819575, KX035096, KX088647, KX228694, KX523680, KY003229, LC035020, LC049337, LC049338, LC049339, LC049340, LC049341, LC049342, LC128713, LC191189, M14707, M16632, M20273, M59808, M59809, M59810, X75215, X83302
Carnivora	KR703607
Rodentia	KT229611, KT229612, KT452637, KT452641, KT452644, KT452685, KT452735
Chiroptera	KT452714, KT452729, KT452730, KT452742
Eulipotyphla	KT452658, KT452661, KT452691, KT452695, KT452698
Scandentia	KT877158

Supplementary Table S2. GenBank accession numbers of MAVS protein and cytochrome b sequences used for pressure analyses in chapter 2.4.2.

Mammalian order	Species name	MAVS Accession number	Cytochrome b Accession number
Primates	<i>Allenopithecus nigroviridis</i>	KC415015	NC023965
	<i>Cercocebus atys</i>	XM012052177	NC028592
	<i>Mandrillus leucophaeus</i>	XM011975135	NC028442
	<i>Papio anubis</i>	KC415017	NC020006
	<i>Papio hamadryas</i>	GAAH01023167	NC001992
	<i>Macaca fascicularis</i>	XM005568426	NC005943
	<i>Macaca mulatta</i>	NM001042666	NC012670
	<i>Macaca nemestrina</i>	XM011741236	NC026976
	<i>Chlorocebus aethiops</i>	KC415012	NC007009
	<i>Chlorocebus sabaues</i>	XM008019736	NC008066
	<i>Erythrocebus patas</i>	KC415013	NC021947
	<i>Miopithecus talapoin talapoin</i>	KC415014	JQ256996
	<i>Ptilocolobus tephrosceles</i>	XM023220219	NC008219
	<i>Colobus guereza</i>	KC415019	NC006901
	<i>Rhinopithecus bieti</i>	XM017878251	NC008218
	<i>Rhinopithecus roxellana</i>	XM010378951	NC018058
	<i>Trachypithecus francoisi</i>	KC415018	NC023970
	<i>Gorilla gorilla</i>	KC415007	NC001645
	<i>Homo sapiens</i>	NM020746	NC012920
	<i>Pan paniscus</i>	XM003821318	NC001644
	<i>Pan troglodytes</i>	GABE01002441	NC001643
	<i>Hylobates alibarbis</i>	KC415008	GU321306
	<i>Symphalangus syndactylus</i>	KC415009	NC014047
	<i>Nomascus leucogenys</i>	KC415010	NC021957
	<i>Pongo abelii</i>	XM003779305	NC002083
	<i>Pongo pygmaeus pygmaeus</i>	KC415011	X97717
	<i>Aotus nancymae</i>	XM012455374	NC018116
	<i>Aotus trivirgatus</i>	KC415023	DQ098873
	<i>Ateles geoffroyi</i>	KC415021	HM222708
	<i>Lagothrix lagotricha</i>	KC415020	NC021951
	<i>Callithrix jacchus</i>	KC415025	NC025586
	<i>Saguinus labiatus</i>	KC415022	HM367996
	<i>Saimiri boliviensis boliviensis</i>	XM003941086	NC018096
	<i>Cebus capucinus imitator</i>	XM017504835	FJ529110
	<i>Plecturocebus moloch</i>	KC415024	KU694272
	<i>Carlito syrichta</i>	XM008066568	NC012774
	<i>Microcebus murinus</i>	XM012755130	NC028718
	<i>Propithecus coquereli</i>	XM012639278	NC011053
	<i>Otolemur garnettii</i>	XM003788212	AF271412
Rodentia	<i>Peromyscus maniculatus bairdii</i>	XM006984556	JF489123
	<i>Heterocephalus glaber</i>	XM004840746	NC015112
	<i>Fukomys damarensis</i>	XM010610975	NC027742
	<i>Chinchilla lanigera</i>	XM005380805	NC021386
	<i>Cavia porcellus</i>	XM003476409	KP100656
	<i>Octodon degus</i>	XM023718584	NC020661
	<i>Castor canadensis</i>	XM020171181	NC033912
	<i>Dipodomys ordii</i>	XM013011081	AF172836
	<i>Marmota marmota marmota</i>	XM015482592	AF100711
	<i>Ictidomys tridecemlineatus</i>	XM005320511	NC027278

	<i>Meriones unguiculatus</i>	XM021655146	NC023263
	<i>Mesocricetus auratus</i>	XM005068650	NC013276
	<i>Cricetulus griseus</i>	XM007623651	NC007936
	<i>Sigmodon hispidus</i>	MK096787	NC035572
	<i>Microtus ochrogaster</i>	XM005365578	NC027945
	<i>Myodes glareolus</i>	MK096788	NC024538
	<i>Rattus norvegicus</i>	NM001005556	NC001665
	<i>Mus musculus</i>	DQ174271	NC010339
	<i>Mus pahari</i>	XM021193616	NC036680
	<i>Nannospalax galili</i>	XM008844340	NC020754
	<i>Jaculus jaculus</i>	XM004661444	NC005314
Chiroptera	<i>Myotis davidii</i>	KU161110	NC025568
	<i>Eidolon helvum</i>	MK096783	JN398200
	<i>Pteropus alecto</i>	XM006921628	NC023122
	<i>Pteropus vampyrus</i>	XM011369865	NC026542
	<i>Hypsignathus monstrosus</i>	MK096784	KX823011
	<i>Rousettus aegyptiacus</i>	XM016136111	NC007393
	<i>Rhinolophus ferrumequinum</i>	MK096786	NC016191
	<i>Rhinolophus sinicus</i>	XM019733790	EF517303
	<i>Myotis brandtii</i>	XM005874684	NC025308
	<i>Myotis lucifugus</i>	XM006096420	NC029849
	<i>Eptesicus fuscus</i>	XM008141050	AF376835
	<i>Pipistrellus pipistrellus</i>	MK096785	AJ504443
	<i>Miniopterus natalensis</i>	XM016212068	KX548057
	<i>Desmodus rotundus</i>	XM024559566	NC022423
	<i>Hipposideros armiger</i>	XM019644596	NC018540

		Primates:					Rodentia:						Chiroptera:					
AA in <i>hs</i> MAVS		SLAC p-value <0.1	FEL p-value <0.1	REL BF>50	FUBAR PPr>0.9	M8 BEB PPr>0.9	AA in <i>P. maniculatus</i> MAVS	SLAC p-value <0.1	FEL p-value <0.1	REL BF>50	FUBAR PPr>0.9	M8 BEB PPr>0.9	AA in <i>M. davidii</i> MAVS	SLAC p-value <0.1	FEL p-value <0.1	REL BF>50	FUBAR PPr>0.9	M8 BEB PPr >0.9
N18		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	K18	n. s.	n. s.	n. s.	n. s.	0.959
T45		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	H45	n. s.	0.086	n. s.	n. s.	n. s.
T47		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	G47	n. s.	n. s.	n. s.	n. s.	0.901
L48		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	L48	n. s.	0.031	107.498	0.937	0.947
S49		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	H49	n. s.	n. s.	65.588	n. s.	0.954
N60		n. s.	n. s.	n. s.	n. s.	0.928		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
Y71		n. s.	n. s.	n. s.	n. s.	n. s.	C71	n. s.	n. s.	78.924	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
A74		0.096	0.043	n. s.	0.906	n. s.	H74	0.049	0.012	315.279	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
G78		n. s.	n. s.	n. s.	n. s.	n. s.	I78	n. s.	0.089	2886.540	n. s.	0.951		n. s.	n. s.	n. s.	n. s.	n. s.
V82		0.088	0.005	269.739	0.971	0.957	P82	0.044	0.003	184.839	n. s.	n. s.	A82	n. s.	0.086	76.996	0.907	0.969
D83		n. s.	0.091	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	A83	n. s.	0.028	n. s.	0.909	0.936
R98		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	P98	n. s.	n. s.	n. s.	n. s.	0.910
S90		n. s.	0.010	209.857	0.947	0.969		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
S100		n. s.	n. s.	n. s.	n. s.	0.900		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
R102		n. s.	n. s.	n. s.	n. s.	n. s.	R103	n. s.	n. s.	282.378	n. s.	0.902		n. s.	n. s.	n. s.	n. s.	n. s.
P104		n. s.	0.061	53.480	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
P109		n. s.	n. s.	n. s.	n. s.	n. s.	S110	n. s.	n. s.	398.282	n. s.	0.906		n. s.	n. s.	n. s.	n. s.	n. s.
S111		n. s.	n. s.	n. s.	n. s.	n. s.	A112	n. s.	0.041	57.118	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
L112		n. s.	n. s.	n. s.	n. s.	n. s.	I113	0.073	0.014	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
P120		n. s.	n. s.	n. s.	n. s.	n. s.	S121	n. s.	0.027	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
P122		n. s.	n. s.	n. s.	n. s.	n. s.	C123	n. s.	0.022	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
A124		n. s.	n. s.	n. s.	n. s.	n. s.	P125	n. s.	n. s.	52.306	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
R134		n. s.	n. s.	n. s.	n. s.	n. s.	Q135	n. s.	0.037	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
P138		n. s.	n. s.	n. s.	n. s.	n. s.	T138	0.084	0.010	n. s.	n. s.	n. s.	T138	n. s.	0.032	n. s.	n. s.	n. s.
A149		n. s.	n. s.	n. s.	n. s.	0.976		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
A160		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	S160	n. s.	n. s.	62.509	n. s.	0.902
L164		n. s.	n. s.	n. s.	n. s.	n. s.	n.a.	n. s.	0.095	n. s.	n. s.	n. s.	D164	n. s.	0.051	n. s.	n. s.	n. s.
L178		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	Q177	n. s.	0.087	n. s.	n. s.	n. s.
S182		n. s.	0.075	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
A185		n. s.	n. s.	n. s.	n. s.	n. s.	Q183	n. s.	n. s.	71.530	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
L190		n. s.	n. s.	n. s.	n. s.	n. s.	Q188	0.026	0.012	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
G194		n. s.	0.090	51.719	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
T200		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	K199	n. s.	0.064	n. s.	n. s.	n. s.
L202		n. s.	n. s.	n. s.	n. s.	n. s.	L200	0.098	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.

T207		n. s.	n. s.	n. s.	n. s.	n. s.	T205	n. s.	n. s.	74.788	n. s.	n. s.	I206	n. s.	0.058	n. s.	n. s.	n. s.
A208		n. s.	0.076	n. s.	n. s.	n. s.	A206	n. s.	0.028	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
A210		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	T209	n. s.	n. s.	134.688	n. s.	0.995
S217		n. s.	n. s.	n. s.	n. s.	n. s.	P215	n. s.	n. s.	82.932	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
R218		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	H217	n. s.	0.048	n. s.	n. s.	n. s.
G245		n. s.	0.081	53.922	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
V247		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	A243	n. s.	0.066	n. s.	n. s.	n. s.
V248		n. s.	n. s.	n. s.	n. s.	n. s.	P246	n. s.	n. s.	994.932	n. s.	0.964		n. s.	n. s.	n. s.	n. s.	n. s.
A262		n. s.	n. s.	n. s.	n. s.	n. s.	A260	n. s.	0.080	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
S263		n. s.	n. s.	n. s.	n. s.	n. s.	F261	n. s.	0.073	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
A264		n. s.	n. s.	n. s.	n. s.	n. s.	T262	0.087	0.010	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
G265		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	G254	0.063	0.043	181.787	0.938	0.975
S275		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	S257	n. s.	0.082	n. s.	n. s.	n. s.
A289		n. s.	n. s.	n. s.	n. s.	n. s.	V281	n. s.	0.075	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
L310		n. s.	n. s.	n. s.	n. s.	0.964		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
S318		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	L291	n. s.	0.089	n. s.	n. s.	n. s.
L326		n. s.	n. s.	n. s.	n. s.	n. s.	L311	n. s.	0.088	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
P333		n. s.	n. s.	n. s.	n. s.	n. s.	T318	n. s.	0.097	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
N339		n. s.	n. s.	n. s.	n. s.	n. s.	T324	n. s.	0.036	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
P344		n. s.	n. s.	n. s.	n. s.	n. s.	T329	n. s.	0.025	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
A345		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	T318	n. s.	0.067	n. s.	n. s.	n. s.
R355		n. s.	0.036	154.063	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
T365		n. s.	0.020	235.402	0.905	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
S373		n. s.	0.066	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
A374		n. s.	n. s.	n. s.	n. s.	n. s.	A357	0.096	0.015	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
D380		n. s.	n. s.	n. s.	n. s.	n. s.	D363	n. s.	n. s.	252.006	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
R384		n. s.	n. s.	n. s.	n. s.	n. s.	R368	n. s.	0.016	87.346	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
N385		n. s.	n. s.	n. s.	n. s.	n. s.	A369	n. s.	0.027	n. s.	n. s.	n. s.	S370	n. s.	n. s.	n. s.	n. s.	0.937
A391		n. s.	n. s.	n. s.	n. s.	n. s.	L375	n. s.	0.085	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
T398		n. s.	0.081	n. s.	n. s.	n. s.	T382	n. s.	0.063	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
W404		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	Y390	n. s.	0.045	n. s.	n. s.	0.933
G413		n. s.	n. s.	n. s.	n. s.	n. s.	H393	n. s.	0.067	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
L414		n. s.	0.048	n. s.	n. s.	0.905		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
V423		n. s.	n. s.	n. s.	n. s.	0.984		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
Q427	HAV cleavage site	n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	Q413	n. s.	n. s.	n. s.	n. s.	0.953
V428		n. s.	n. s.	n. s.	n. s.	n. s.	L408	0.029	0.025	699.855	n. s.	0.957		n. s.	n. s.	n. s.	n. s.	n. s.
G448		n. s.	n. s.	n. s.	n. s.	n. s.	N428	n. s.	0.091	58.164	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
M449		n. s.	n. s.	n. s.	n. s.	0.917		n. s.	n. s.	n. s.	n. s.	n. s.	S436	n. s.	n. s.	n. s.	n. s.	0.952
K461	alternative HAV	n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	T448	n. s.	n. s.	68.472	n. s.	0.975
T465		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	S452	n. s.	0.051	81.778	0.918	0.936

		cleavage site																	
	N473		n. s.	n. s.	n. s.	n. s.	n. s.	D450	n. s.	n. s.	110.053	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
	I476		n. s.	n. s.	n. s.	n. s.	n. s.	A453	n. s.	0.051	70.320	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
	P486		n. s.	0.059	53.119	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
	A487		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	A474	n. s.	n. s.	n. s.	n. s.	n. s.
	D490		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	0.928
	A497		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
	D498		n. s.	0.060	n. s.	0.922	0.975		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
	R499		n. s.	n. s.	n. s.	n. s.	0.917		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.
	K500		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	L478	n. s.	0.087	n. s.	n. s.	n. s.
	F501		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	K479	n. s.	n. s.	n. s.	n. s.	0.951
	V506	HCV cleavage site	n. s.	n. s.	n. s.	n. s.	0.986		n. s.	n. s.	n. s.	n. s.	n. s.	A484	n. s.	n. s.	119.767	n. s.	0.994
	P507		n. s.	n. s.	n. s.	n. s.	n. s.	L476	n. s.	n. s.	4495.200	n. s.	0.986	R485	n. s.	n. s.	n. s.	n. s.	0.978
	R510		n. s.	n. s.	n. s.	n. s.	n. s.	S479	n. s.	n. s.	183.704	n. s.	0.931		n. s.	n. s.	n. s.	n. s.	n. s.
	P511		n. s.	n. s.	n. s.	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.	E489	n. s.	n. s.	202.475	0.914	0.990
	S512		n. s.	n. s.	n. s.	n. s.	n. s.	I481	n. s.	n. s.	87.298	n. s.	0.952		n. s.	n. s.	n. s.	n. s.	n. s.
	P513		n. s.	n. s.	n. s.	n. s.	n. s.	P482	n. s.	0.027	146.233	n. s.	n. s.		n. s.	n. s.	n. s.	n. s.	n. s.

BF, Bayes Factor

Supplementary Table S3. GISAID accession numbers of SARS-CoV-2 genomes sequenced during March 2020 and July 2021, Benin*

Sample collection dates	No. sequenced	GISAID Accession No.†
March-April 2020	12	EPI_ISL_476822, EPI_ISL_476823, EPI_ISL_476824, EPI_ISL_476825, EPI_ISL_476826, EPI_ISL_476827, EPI_ISL_476828, EPI_ISL_476829, EPI_ISL_476830, EPI_ISL_476831, EPI_ISL_476833, EPI_ISL_476834
January – April 2021	68	EPI_ISL_2932532, EPI_ISL_2932533, EPI_ISL_2932534, EPI_ISL_2932535, EPI_ISL_2932536, EPI_ISL_2932537, EPI_ISL_2932538, EPI_ISL_2932539, EPI_ISL_2932540, EPI_ISL_2932541, EPI_ISL_2932542, EPI_ISL_2932543, EPI_ISL_2932544, EPI_ISL_2932545, EPI_ISL_2932546, EPI_ISL_2932547, EPI_ISL_2932548, EPI_ISL_2932549, EPI_ISL_2932550, EPI_ISL_2932551, EPI_ISL_2932552, EPI_ISL_2932553, EPI_ISL_2932554, EPI_ISL_2932555, EPI_ISL_2932556, EPI_ISL_2932557, EPI_ISL_2932558, EPI_ISL_2932559, EPI_ISL_2932560, EPI_ISL_2932561, EPI_ISL_2932562, EPI_ISL_2932563, EPI_ISL_2932564, EPI_ISL_2932565, EPI_ISL_2932566, EPI_ISL_2932567, EPI_ISL_2932568, EPI_ISL_2932569, EPI_ISL_2932570, EPI_ISL_2932571, EPI_ISL_2932572, EPI_ISL_2932573, EPI_ISL_2932574, EPI_ISL_2932575, EPI_ISL_2932576, EPI_ISL_2932577, EPI_ISL_2932578, EPI_ISL_2932579, EPI_ISL_2932580, EPI_ISL_2932581, EPI_ISL_2932582, EPI_ISL_2932583, EPI_ISL_2932584, EPI_ISL_2958658, EPI_ISL_2958659, EPI_ISL_2958660, EPI_ISL_2958661, EPI_ISL_2958662, EPI_ISL_2958663, EPI_ISL_2958664, EPI_ISL_2958665, EPI_ISL_2958666, EPI_ISL_2958667, EPI_ISL_2958668, EPI_ISL_2958669, EPI_ISL_2958670, EPI_ISL_2958671, EPI_ISL_2958672
January–March 2021	88	EPI_ISL_4567011, EPI_ISL_4567013, EPI_ISL_4567015, EPI_ISL_4567016, EPI_ISL_4567018, EPI_ISL_4567020, EPI_ISL_4567022, EPI_ISL_4567023, EPI_ISL_4567025, EPI_ISL_4567027, EPI_ISL_4567029, EPI_ISL_4567031, EPI_ISL_4567032, EPI_ISL_4567034, EPI_ISL_4567036, EPI_ISL_4567038, EPI_ISL_4567039, EPI_ISL_4567041, EPI_ISL_4567043, EPI_ISL_4567045, EPI_ISL_4567046, EPI_ISL_4567048, EPI_ISL_4567050, EPI_ISL_4567052, EPI_ISL_4567054, EPI_ISL_4567056, EPI_ISL_4567058, EPI_ISL_4567059, EPI_ISL_4567061, EPI_ISL_4567063, EPI_ISL_4567065, EPI_ISL_4567067, EPI_ISL_4567069, EPI_ISL_4567071, EPI_ISL_4567072, EPI_ISL_4567074, EPI_ISL_4567076, EPI_ISL_4567078, EPI_ISL_4567080, EPI_ISL_4567081, EPI_ISL_4567083, EPI_ISL_4567085, EPI_ISL_4567087, EPI_ISL_4567089, EPI_ISL_4567090, EPI_ISL_4567092, EPI_ISL_4567094, EPI_ISL_4567096, EPI_ISL_4567098, EPI_ISL_4567099, EPI_ISL_4567101, EPI_ISL_4567103, EPI_ISL_4567105, EPI_ISL_4567107, EPI_ISL_4567108, EPI_ISL_4567110, EPI_ISL_4567112, EPI_ISL_4567114, EPI_ISL_4567116, EPI_ISL_4567118, EPI_ISL_4567120, EPI_ISL_4567122, EPI_ISL_4567124, EPI_ISL_4567126, EPI_ISL_4567128, EPI_ISL_4567130, EPI_ISL_4567132, EPI_ISL_4567133, EPI_ISL_4567135, EPI_ISL_4567137, EPI_ISL_4567139, EPI_ISL_4567141,

		EPI_ISL_4567143, EPI_ISL_4567144, EPI_ISL_4567146, EPI_ISL_4567148, EPI_ISL_4567150, EPI_ISL_4567152, EPI_ISL_4567153, EPI_ISL_4567155, EPI_ISL_4567157, EPI_ISL_4567158, EPI_ISL_4567160, EPI_ISL_4567162, EPI_ISL_4567164, EPI_ISL_4567166, EPI_ISL_4567168, EPI_ISL_4567169
April–July 2021	114	EPI_ISL_4566811, EPI_ISL_4566813, EPI_ISL_4566814, EPI_ISL_4566816, EPI_ISL_4566818, EPI_ISL_4566820, EPI_ISL_4566821, EPI_ISL_4566823, EPI_ISL_4566825, EPI_ISL_4566827, EPI_ISL_4566829, EPI_ISL_4566830, EPI_ISL_4566832, EPI_ISL_4566834, EPI_ISL_4566836, EPI_ISL_4566837, EPI_ISL_4566839, EPI_ISL_4566841, EPI_ISL_4566843, EPI_ISL_4566845, EPI_ISL_4566846, EPI_ISL_4566848, EPI_ISL_4566850, EPI_ISL_4566852, EPI_ISL_4566854, EPI_ISL_4566855, EPI_ISL_4566857, EPI_ISL_4566859, EPI_ISL_4566861, EPI_ISL_4566863, EPI_ISL_4566865, EPI_ISL_4566867, EPI_ISL_4566869, EPI_ISL_4566871, EPI_ISL_4566872, EPI_ISL_4566874, EPI_ISL_4566876, EPI_ISL_4566878, EPI_ISL_4566880, EPI_ISL_4566882, EPI_ISL_4566883, EPI_ISL_4566885, EPI_ISL_4566887, EPI_ISL_4566889, EPI_ISL_4566891, EPI_ISL_4566893, EPI_ISL_4566894, EPI_ISL_4566896, EPI_ISL_4566898, EPI_ISL_4566900, EPI_ISL_4566901, EPI_ISL_4566903, EPI_ISL_4566905, EPI_ISL_4566907, EPI_ISL_4566908, EPI_ISL_4566910, EPI_ISL_4566912, EPI_ISL_4566914, EPI_ISL_4566915, EPI_ISL_4566917, EPI_ISL_4566919, EPI_ISL_4566921, EPI_ISL_4566922, EPI_ISL_4566924, EPI_ISL_4566926, EPI_ISL_4566928, EPI_ISL_4566929, EPI_ISL_4566931, EPI_ISL_4566933, EPI_ISL_4566935, EPI_ISL_4566936, EPI_ISL_4566938, EPI_ISL_4566940, EPI_ISL_4566942, EPI_ISL_4566943, EPI_ISL_4566945, EPI_ISL_4566947, EPI_ISL_4566949, EPI_ISL_4566950, EPI_ISL_4566952, EPI_ISL_4566954, EPI_ISL_4566956, EPI_ISL_4566958, EPI_ISL_4566959, EPI_ISL_4566961, EPI_ISL_4566963, EPI_ISL_4566964, EPI_ISL_4566966, EPI_ISL_4566968, EPI_ISL_4566970, EPI_ISL_4566971, EPI_ISL_4566973, EPI_ISL_4566975, EPI_ISL_4566977, EPI_ISL_4566978, EPI_ISL_4566980, EPI_ISL_4566982, EPI_ISL_4566984, EPI_ISL_4566985, EPI_ISL_4566987, EPI_ISL_4566989, EPI_ISL_4566990, EPI_ISL_4566992, EPI_ISL_4566994, EPI_ISL_4566996, EPI_ISL_4566997, EPI_ISL_4566999, EPI_ISL_4567001, EPI_ISL_4567002, EPI_ISL_4567004, EPI_ISL_4567006, EPI_ISL_4567008, EPI_ISL_4567009, EPI_ISL_4572438

*SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

†GISAID, <https://www.gisaid.org>.

Supplementary Table S4. Characteristics of severe acute respiratory syndrome coronavirus 2–positive samples from which full genomes were generated, Benin, January-March 2021

ID	Collection date	C _t E-gene	Pango lineage	Genome completeness, %
256146	2021 Feb 4	35.74	B.1	91
315551	2021 Mar 3	17.88	B.1.1.7	99
314235	2021 Mar 2	15.76	B.1.1.7	100
255701	2021 Feb 4	17.85	B.1.1.7	100
255138	2021 Feb 3	26.62	B.1.1.7	99
253094	2021 Feb 2	18.84	B.1.1.7	100
251917	2021 Feb 1	20.57	B.1.1.7	100
251354	2021 Feb 1	19.88	B.1.1.7	100
251326	2021 Feb 1	15.97	B.1.1.7	100
311929	2021 Mar 2	23.94	B.1.1.7	98
256208	2021 Feb 4	25.25	B.1.1.7	100
256046	2021 Feb 4	24.86	B.1.1.7	98
249234	2021 Jan 30	33.83	B.1	97
251307	2021 Feb 1	17.21	B.1	99
250990	2021 Feb 1	19.01	B.1	96
312964	2021 Mar 2	31.63	B.1.525	96
312950	2021 Mar 2	29.43	B.1.525	98
312266	2021 Mar 1	17.89	B.1.525	99
312198	2021 Mar 1	19.66	B.1.525	99
254242	2021 Feb 3	17.73	B.1.525	99
253408	2021 Feb 2	16.49	B.1.525	99
250541	2021 Jan 31	21.35	B.1.525	99
315465	2021 Mar 3	15.61	B.1.525	99
254128	2021 Feb 3	18.45	A.23.1	99
253832	2021 Feb 2	24.05	B.1	100
249868	2021 Jan 31	25.14	B.1	100
249713	2021 Jan 31	25.52	B.1	95
250814	2021 Feb 1	22.26	B.1	100
250323	2021 Jan 31	31.74	B.1.1.420	94
250412	2021 Feb 1	32.49	B.1.160	97
254278	2021 Feb 3	29.65	L.3	99
250924	2021 Feb 1	21.61	L.3	100
250772	2021 Feb 1	28.08	L.3	99
249964	2021 Jan 31	31.88	L.3	95
249110	2021 Jan 30	29.33	L.3	98
250699	2021 Feb 1	22.88	A.27	100
312648	2021 Mar 1	28.07	B.1.1.318	97
312572	2021 Mar 1	21.97	B.1.1.7	100
311995	2021 Mar 2	25.31	B.1.1.7	99
314058	2021 Mar 2	21.58	B.1.1.7	100
312541	2021 Mar 1	23.00	B.1.214.2	98
248661	2021 Jan 30	32.81	B.1.525	92
314176	2021 Mar 2	28.31	B.1.525	98
253620	2021 Feb 2	21.22	B.1.525	98
254286	2021 Feb 3	23.32	B.1.525	99
315530	2021 Mar 3	18.83	B.1.1.318	99
312262	2021 Mar 1	18.44	B.1.1.318	99
312182	2021 Mar 1	16.95	B.1.1.318	99
311979	2021 Mar 2	18.25	B.1.1.318	94

254375	2021 Feb 3	20.15	B.1.1.318	99
253228	2021 Feb 2	21.04	B.1.1.318	99
249971	2021 Jan 31	20.47	B.1.1.318	99
249944	2021 Jan 31	22.11	B.1.1.318	99
248922	2021 Jan 30	21.60	B.1.1.318	99
251411	2021 Feb 1	24.71	B.1.1.318	96
255062	2021 Feb 3	17.53	A.27	100
252348	2021 Feb 2	21.44	A.27	98
312239	2021 Mar 1	24.61	A.27	100
255170	2021 Feb 3	25.66	A.27	100
254323	2021 Feb 3	19.68	A.27	100
253312	2021 Feb 2	19.71	A.27	100
251455	2021 Feb 1	19.76	A.27	99
251296	2021 Feb 1	20.07	A.27	100
250498	2021 Jan 31	20.67	A.27	100
250471	2021 Jan 31	31.86	A.27	98
249839	2021 Jan 31	22.66	A.27	100
248651	2021 Jan 30	23.22	A.27	98
314080	2021 Mar 2	21.90	B.1.1.7	100

Supplementary Table S5. Spike mutations shared by SARS-CoV-2 lineages defined as Variants of Concern (VOC) by the World Health Organization by May 2021.

Mutation	VOC			
	B.1.1.7 (Alpha)	B.1.351 (Beta)	P.1 (Gamma)	B.1.617.2 (Delta)
1		L18F	L18F	
2				T19R
3			P26S	
4	DelHV69/70			
5				G142D
6	DelY144			
7				Del F157
8		K417N	K417T	
9				L452R
10				T478K
11		E484K	E484K	
12	N501Y	N501Y	N501Y	
13	D614G	D614G	D614G	D614G
14			H655Y	
15	P681H			P681R
16				D950N

Supplementary Table S6. Accession numbers of viruses included for analysis in Figure 19. Amino acid sequences of related coronaviruses and human NCBI reference strains were tested for FCS prediction in ProP. Translated sequence fragments encompassing 100 amino acids up- and downstream of the putative cleavage site in the respective human reference sequences were analyzed. The corresponding amino acid positions in the spike coding sequences were: MERS-CoV (NC038294) aa 651-851, HCoV-229E (NC002645) aa 468-668, HCoV-NL63 (NC005831) aa 648-848, HCoV-OC43/HCoV-HKU1 (NC006577) aa 660-860 and SARS-related CoV (MT019529) aa 585-785. Conservation of FCS motifs in further human sequences was examined in translation alignments of protein sequences as described in the materials and methods.

HCoV-229E (total n=144):	
Host orders:	104 Primates, 6 Chiroptera, 33 Artiodactyla
40 nucleotide sequences used for ProP analysis:	JQ410000, KT253269, KT253270, KT253271, KT253272, KT253324, KT253325, KT253326, KT253327, KT253328, KT368892, KT368893, KT368894, KT368895, KT368896, KT368897, KT368898, KT368899, KT368900, KT368901, KT368902, KT368903, KT368904, KT368905, KT368906, KT368908, KT368909, KT368910, KT368911, KT368912, KT368913, KT368914, KT368915, KT368916, KU291449, KY073747, KY073748, MF593473, NC002645, NC028752
104 human-derived protein sequences:	6IXA_A, 6U7H_A, AAK32188, ABB90506, ABB90507, ABB90508, ABB90509, ABB90510, ABB90513, ABB90514, ABB90515, ABB90516, ABB90519, ABB90520, ABB90522, ABB90523, ABB90526, ABB90527, ABB90528, ABB90529, AFR79250, AFR79257, AGT21338, AGT21345, AGT21353, AGT21367, AGW80677, AGW80932, AIW52728, AIW52729, AIW52731, AIW52732, AIW52735, AIW52740, AIW52741, AIW52745, AIW52746, AIW52747, AIW52748, AIW52749, AIW52750, AIW52752, AIW52753, AIW52754, AIW52755, AIW52756, AIW52757, AOG74783, APT69849, APT69856, APT69862, APT69876, APT69883, APT69890, ARA15429, ARB07425, ARK08642, ARU07601, AWH62679, BAL45637, BAL45638, BAL45639, BAL45640, BDB58082, BDB58089, CAA71146, CAA71147, NP_073551, QEG03785, QEO75985, QJY77946, QJY77970, QJY77978, QNT54756, QNT54802, QNT54803, QNT54804, QNT54805, QNT54806, QNT54807, QNT54808, QNT54809, QNT54811, QNT54812, QNT54813, QNT54814, QNT54815, QNT54817, QNT54822, QNT54825, QNT54826, QNT54828, QNT54833, QNT54834, QNT54838, QNT54839, QNT54840, QNT54841, QQY99248, QRK03775, QRK03782, QRK03789, QRK03803, UDL16997
MERS-CoV (total n=451):	
Host orders:	180 Primates, 241 Artiodactyla, 27 Chiroptera, 3 Eulipotyphla
274 nucleotide sequences used for ProP analysis:	DQ648794, EF065506, EF065507, EF065508, EF065510, EF065511, EF065512, KC545386, KC869678, KJ473820, KJ473821, KJ473822, KJ477103, KU740200, KY581695, KY581696, KY581697, KY581698, KY581699, KY581700, KY673149, MF593268, MF598594, MF598595, MF598596, MF598597, MF598598, MF598599, MF598600, MF598601, MF598602, MF598603, MF598604, MF598605, MF598606, MF598607, MF598608, MF598609, MF598610, MF598611, MF598612, MF598613, MF598614, MF598615, MF598616, MF598617, MF598618, MF598619, MF598620, MF598621, MF598622, MF598623, MF598624, MF598625, MF598626, MF598627, MF598628, MF598629, MF598630, MF598631, MF598632, MF598633, MF598634, MF598635, MF598636, MF598637, MF598638, MF598639, MF598640, MF598641, MF598642, MF598643, MF598644, MF598645, MF598646, MF598647, MF598648, MF598649, MF598650, MF598651, MF598652, MF598653, MF598654, MF598655,

	<p>MF598656, MF598657, MF598658, MF598659, MF598660, MF598661, MF598662, MF598663, MF598664, MF598665, MF598666, MF598667, MF598668, MF598669, MF598670, MF598671, MF598672, MF598673, MF598674, MF598675, MF598676, MF598677, MF598678, MF598679, MF598680, MF598681, MF598682, MF598683, MF598684, MF598685, MF598686, MF598687, MF598688, MF598689, MF598690, MF598691, MF598692, MF598693, MF598694, MF598695, MF598696, MF598697, MF598698, MF598699, MF598700, MF598701, MF598702, MF598703, MF598704, MF598705, MF598706, MF598707, MF598708, MF598709, MF598710, MF598711, MF598712, MF598713, MF598714, MF598715, MF598716, MF598717, MF598718, MF598719, MF598720, MF598721, MF598722, MG021451, MG021452, MG596802, MG596803, MG923466, MG923467, MG923468, MG923469, MG923470, MG923471, MG923472, MG923473, MG923474, MG923475, MG923476, MG923477, MG923478, MG923479, MG923480, MG923481, MG987420, MG987421, MH002337, MH002338, MH002339, MH002340, MH002341, MH002342, MH259485, MH259486, MH371127, MH734114, MH734115, MK357908, MK357909, MK564474, MK564475, MK679660, MK967708, MN654970, MN654971, MN654972, MN654973, MN654974, MN654975, MN654976, MN654977, MN654978, MN654979, MN654980, MN654981, MN654982, MN654983, MN654984, MN654985, MN654986, MN654987, MN654988, MN654989, MN654990, MN654991, MN654992, MN654993, MN654994, MN654995, MN654996, MN654997, MN654998, MN654999, MN655000, MN655001, MN655002, MN655003, MN655004, MN655005, MN655006, MN655007, MN655008, MN655009, MN655010, MN655011, MN655012, MN655013, MN655014, MN655015, MN655016, MN655017, MN758606, MN758607, MN758608, MN758609, MN758610, MN758611, MN758612, MN758613, MN758614, MT226600, MT226601, MT226602, MT226603, MT226604, MT226605, MT226606, MT226607, MW086527, MW086528, MW086529, MW086530, MW086531, MW086532, MW086533, MW086534, MW086535, MW086538, MW086539, MW218395, MW545527, MW545528, NC009019, NC009020, NC019843, NC038294, NC039207</p>
<p>180 human-derived protein sequences:</p>	<p>5X59_A, AGH58717, AGN52936, AGN70929, AGV08379, AGV08390, AGV08438, AGV08455, AGV08467, AGV08477, AGV08492, AGV08524, AGV08584, AHC74088, AHI48528, AHI48550, AHI48572, AHI48583, AHI48616, AHI48672, AHI48692, AHI48702, AHI48727, AHI48729, AHI48731, AHI48733, AHI48735, AHI48737, AHI48800, AHI48801, AHZ20790, AHZ58501, AHZ64057, AHZ90568, AID55066, AID55079, AID55084, AID55085, AID55087, AID55088, AID55090, AID55095, AID55096, AID55097, AID55099, AIY60518, AIY60528, AIY60578, AIZ48769, AIZ74405, AIZ74433, AIZ74439, AJD81440, AKI29255, AKL59401, AKM76239, AKN11072, AKN11074, AKN11076, AKN24803, AKN24812, AKN24830, ALB08246, ALB08311, ALB08322, ALD51904, ALJ54446, ALJ54448, ALJ54450, ALJ54451, ALJ54452, ALJ54453, ALJ54455, ALJ54456, ALJ54458, ALJ54461, ALJ54467, ALJ54471, ALJ54472, ALJ54474, ALJ54486, ALJ54488, ALJ54490, ALJ54491, ALJ54495, ALJ54496, ALJ54500, ALJ54501, ALJ54502, ALJ54508, ALJ54512, ALJ54516, ALJ54517, ALJ54518, ALJ54521, ALJ76277, ALJ76278, ALJ76286, ALK80251, ALK80261, ALK80291, ALK80311, ALM26400, ALW82691, ALW82742, ALX27228, ALX27232, AMW90836, AMW90843, AMW90852, AMW90853, ANC28634, ANC28678, ANF29173,</p>

	ANF29184, ANF29195, ANF29217, ANF29228, ANF29239, ANF29261, ANF29272, AQZ41282, AQZ41285, ASU45807, ASY99811, ASY99842, ATG84679, ATG84701, ATG84723, ATG84756, ATG84767, ATG84833, ATG84855, ATG84866, ATG84877, ATG84888, ATG84899, ATG84910, ATY74392, AVK87340, AVK87417, AVK87439, AVK87450, AVK87471, AWU59321, AXG21654, AXN73370, AXN73437, AXN73525, AXN92228, AXN92260, AZU90731, QBF44113, QBF80455, QBF80488, QBF80499, QBF80521, QBF80543, QBF80554, QBF80565, QBF80598, QBF80607, QBF80608, QBF80611, QCQ28829, QCQ28832, QCQ29075, QDP16195, QDP16206, QEU56412, QEU56412, QGW51938, QLD97934, QLD97977, QLD98042, QLD98075, QLD98105, UHI99922, YP_007188579, YP_009047204
HCoV-NL63 (total n=58):	
Host orders:	55 Primates, 3 Chiroptera,
4 nucleotide sequences used for ProP analysis:	KY073746, NC032107, NC048216, NC005831
55 human-derived protein sequences:	AAS89767, ABE97130, ABE97137, AFD64748, AFD64760, AFD98750, AFD98771, AFD98778, AFD98785, AFD98806, AFD98820, AFD98827, AFD98834, AFO70497, AFV53148, AGT51324, AGT51331, AGT51345, AGT51366, AGT51380, AIW52831, AIW52835, AIW52836, AIW52839, AIW52840, AIW52841, AIW52843, AKT07952, ALJ53431, APF29071, ARB07399, ARB07406, ARE29967, ARE29973, ARE29979, ARU07594, AVA26873, AWK59979, BBL54109, BBL54113, BBL54116, BDB58168, QDH43736, QED88012, QED88019, QED88026, QED88033, QED88040, QEG03731, QEG03748, QQY99321, QQY99388, UDL16979, UDL16985, YP_003767
HCoV-OC43 + HCoV-HKU1 (total n=466):	
Host orders:	264 Primates for HCoV-OC43, 58 Primates for HCoV-HKU1, 125 Artiodactyla, 12 Rodentia, 4 Perissodactyla, 2 Carnivora, 1 Lagomorpha
146 nucleotide sequences used for ProP analysis:	AB354579, AF220295, AF391542, DQ011855, DQ811784, DQ915164, EF424615, EF424617, EF424619, EF424620, EF424621, EF424623, FJ425184, FJ425185, FJ425186, FJ425187, FJ425188, FJ425189, FJ938063, FJ938064, FJ938065, FJ938066, JF792616, JF792617, JX860640, KF850449, KF906249, KF906250, KM349743, KM349744, KU558922, KU558923, KU886219, KX432213, KX982264, KY419103, KY419104, KY419105, KY419106, KY419107, KY419109, KY419110, KY419112, KY419113, KY994645, LC061272, LC061273, LC061274, LC494126, LC494127, LC494128, LC494129, LC494130, LC494131, LC494132, LC494133, LC494134, LC494135, LC494136, LC494137, LC494138, LC494139, LC494140, LC494141, LC494142, LC494143, LC494144, LC494145, LC494147, LC494148, LC494149, LC494150, LC494151, LC494152, LC494153, LC494154, LC494155, LC494156, LC494157, LC494158, LC494159, LC494160, LC494161, LC494162, LC494163, LC494164, LC494165, LC494166, LC494167, LC494168, LC494169, LC494170, LC494171, LC494172, LC494173, LC494174, LC494175, LC494176, LC494177, LC494178, LC494179, LC494180, LC494181, LC494182, LC494183, LC494184, LC494185, LC494186, LC494187, LC494188, LC494189, LC494190, LC494191, LC494192, LC592689, MF083115, MG518518, MG757138, MG757139, MG757140, MG757141, MG757142, MH043953, MH810163, MN514962, MN514963, MN514964, MN514965, MN514966, MN514967, MN982198, MN982199, MT820628, MT820629, MT820630, MT820631, MW165134, MW711287, MW773844,

		NC026011, NC0003045, NC006213, NC006577, NC012936, NC017083, U00735
264	human-derived protein sequences HCoV-OC43:	AAB27260, AAR01015, AAT84362, AAX84791, AAX84792, AAX84795, AAX85669, ABD43177, ABD43178, ABD43179, ABD43180, ABD43181, ABD43182, ABD43183, AEN19358, AEN19366, AFR45567, AFR45569, AFR45570, AFR45571, AFR45577, AGT51402, AGT51412, AGT51422, AGT51431, AGT51441, AGT51451, AGT51461, AGT51481, AGT51491, AGT51501, AGT51511, AGT51521, AGT51541, AGT51551, AGT51570, AGT51590, AGT51600, AGT51610, AGT51620, AGT51630, AGT51650, AGT51680, AGT51690, AGT51700, AGT51740, AGT51760, AGT51790, AGT51800, AIL49484, AIL49485, AIL49486, AIL49487, AIL49488, AIL49489, AIL49490, AIL49491, AIL49492, AIL49493, AIL49494, AIL49495, AIL49496, AIL49497, AIL49498, AIL49499, AIL49500, AIL49501, AIL49502, AIL49503, AIL49505, AIL49507, AIL49508, AIL49509, AIL49511, AIL49512, AIL49513, AIL49515, AIL49516, AIL49518, AIL49519, AIL49520, AIL49521, AIL49523, AIL49525, AIL49526, AIL49528, AIL49529, AIL49531, AIL49536, AIL49538, AIL49539, AIL49542, AIL49543, AIL49544, AIL49545, AIL49546, AIL49548, AIL49550, AIV41831, AIV41873, AIV41891, AIV41909, AIX09799, AIX09807, AIX10748, AIX10749, AIX10750, AIX10751, AIX10752, AIX10753, AIX10755, AIX10756, AIX10757, AIX10758, AIX10759, AIX10760, AIX10761, AIX10762, AIX10763, AMK59677, ANZ78834, ANZ78835, ANZ78836, ANZ78837, ANZ78838, ANZ78839, ANZ78840, ANZ78841, ANZ78843, ANZ78844, ANZ78845, ANZ78847, ANZ78849, AOL02453, APU51916, APU51936, AQN78656, AQN78664, AQN78672, AQN78680, AQN78704, AQN78736, AQN78744, AQN78768, AQN78776, ARA15421, ARB07414, ARB07433, ARB66060, ARB66061, ARB66062, ARB66063, ARB66064, ARB66065, ARB66066, ARB66067, ARB66068, ARB66069, ARB66070, ARB66071, ARE30008, ARK08635, ARK08651, ARK08661, ARK08670, ARK08679, ATN39872, ATP16757, AVQ05264, AVR40342, AVR40344, AWW13566, AXX83303, AXX83309, AXX83321, AXX83327, AXX83333, AXX83339, AXX83345, AXX83351, AXX83369, AXX83375, AXX83381, BBA20973, BBA20976, BBA20979, BBA20982, BDB58125, BDB58134, BDB58143, BDB58152, BDB58161, BDB58177, CAA79896, CAA83660, CAA83661, P36334, QBP84704, QBP84713, QBP84722, QBP84731, QBP84759, QBQ01839, QDH43719, QDH43726, QEG03740, QEG03757, QEG03767, QEG03776, QEG03794, QEG03803, QEG03814, QHB49097, QHB49098, QHB49099, QKS68383, QKS68384, QKS68385, QKS68390, QKS68398, QKS68406, QKS68411, QKS68414, QQY99171, QQY99180, QQY99202, QQY99219, QQY99257, QQY99373, QRK03812, QRK03822, QRK03832, QRK03842, QRK03852, QRK03862, QRK03872, QRK03882, QRK03892, QRK03912, QRK03922, QXL74886, UCC70448, UCC70458, UCC70468, UCC70478, UCC70488, UCC70498, UCC70518, UCC70528, UCC70538, UCC70548, UDL16993, UDM84825, UDM84835, UDM84845, UDM84855, UDM84865, UDM84875, UDM84885, UDM84895, UDM84905, UDM84911, UDM84912, YP_009555241
58	human-derived protein sequences HCoV-HK1:	ABD75497, ABD75505, ABD75513, ABD75529, ABD75545, ABD75553, ABD75593, ABD75601, ABD75609, ABD75617, ABD75625, ADN03339, AGT17758, AGT17769, AGT17777, AGW27836, AGW27863, AGW27872, AMN88686, AMN88694,

	ARB07438, ARU07577, AXT92528, AXT92529, AXT92530, AXT92531, AXT92533, AXT92534, AXT92537, AXT92538, AXT92539, AXT92540, AXT92541, AXT92542, AXT92543, AXT92544, AXT92545, AXT92546, AXT92547, AXT92548, AXT92549, AXT92550, AXT92551, AXT92552, AXT92555, AXT92556, AXT92557, AXT92559, AXT92560, AXT92561, AZS52618, BBA20983, BBA20986, BDB58097, Q0ZME7, Q14EB0, QHB49101, YP 173238
SARS-related CoV (total n=4,824,450):	
Host orders:	40 Primates SARS-CoV, 4,824,313 Primates SARS-CoV-2, 79 Chiroptera, 12 Carnivora, 6 Pholidota
99 nucleotide sequences for ProP analysis:	AY304486, AY304488, AY545914, AY545915, AY545916, AY545917, AY545918, AY545919, AY572034, AY572035, AY572038, DQ071615, DQ412042, DQ412043, DQ648856, DQ648857, EPIISL852604, EPIISL852605, FJ588686, FJ959407, GU190215, JX993987, JX993988, KC881005, KC881006, KF294457, KF367457, KF569996, KJ473811, KJ473812, KJ473813, KJ473814, KJ473815, KJ473816, KP886808, KP886809, KT444582, KU182964, KY352407, KY417142, KY417143, KY417144, KY417145, KY417146, KY417147, KY417148, KY417149, KY417150, KY417151, KY417152, KY770858, KY770859, KY770860, KY938558, LC556375, MG772933, MG772934, MK211374, MK211375, MK211376, MK211377, MK211378, MN996532, MT019529, MT040333, MT040334, MT040335, MT040336, MT072864, MT121216, MT726043, MT726045, MW251308, MW703458, MW719567, MZ081376, MZ081377, MZ081378, MZ081379, MZ081380, MZ081381, MZ081382, MZ190137, MZ190138, MZ937000, MZ937001, MZ937002, MZ937003, MZ937004, NC004718. Acc. No. of sequences generated in this work are written in the material and methods)
40 human-derived protein sequences SARS-CoV:	5WRG_A, 5X58_A, 6ACC_A, 6CRV_A, 6CRW_A, 6M3W_A, 6NB6_A, 7AKJ_A, 7SG4_A, AAP13567, AAP33697, AAP51227, AAP82968, AAR07624, AAR07625, AAR07626, AAR07627, AAR07628, AAR07629, AAR07630, AAR07631, AAR23250, AAR86775, AAR91586, AAS00003, AAS10463, AAT74874, AAT76147, AAU93320, AAX16192, ABD72969, ABD72970, ABD72972, ACB69860, ACB69883, ACB69894, ACB69905, ACQ82725, BAE93401, YP 009825051
304 human-derived sequences of SARS-CoV-2 showing complete or partial deletion of the SARS-CoV-2 FCS (682RRAR ₆₈₅) or had at least one amino acid substitution at position R682 or R685:	EPI_ISL_1076731, EPI_ISL_1078726, EPI_ISL_1109980, EPI_ISL_1110008, EPI_ISL_1190402, EPI_ISL_1237025, EPI_ISL_1254709, EPI_ISL_1347446, EPI_ISL_1358702, EPI_ISL_1369650, EPI_ISL_1420505, EPI_ISL_1430408, EPI_ISL_1443308, EPI_ISL_1568824, EPI_ISL_1639691, EPI_ISL_1639709, EPI_ISL_1653927, EPI_ISL_1653931, EPI_ISL_1669438, EPI_ISL_1686748, EPI_ISL_1708518, EPI_ISL_1727961, EPI_ISL_1727963, EPI_ISL_1785076, EPI_ISL_1785078, EPI_ISL_1807267, EPI_ISL_1843335, EPI_ISL_1843346, EPI_ISL_1863962, EPI_ISL_1885892, EPI_ISL_1886161, EPI_ISL_2036293, EPI_ISL_2094800, EPI_ISL_2095285, EPI_ISL_2110767, EPI_ISL_2110768, EPI_ISL_2112529, EPI_ISL_2122601, EPI_ISL_2190982, EPI_ISL_2256945, EPI_ISL_2299225, EPI_ISL_2315924, EPI_ISL_2316018, EPI_ISL_2321068, EPI_ISL_2331786, EPI_ISL_2360926, EPI_ISL_2360937, EPI_ISL_2360960, EPI_ISL_2365909, EPI_ISL_2388184, EPI_ISL_2443366, EPI_ISL_2452004, EPI_ISL_2454062, EPI_ISL_2454066,

EPI_ISL_2454092,	EPI_ISL_2454093,	EPI_ISL_2454099,
EPI_ISL_2469883,	EPI_ISL_2509960,	EPI_ISL_2611084,
EPI_ISL_2611842,	EPI_ISL_2613345,	EPI_ISL_2635448,
EPI_ISL_2635450,	EPI_ISL_2640526,	EPI_ISL_2645416,
EPI_ISL_2645418,	EPI_ISL_2659920,	EPI_ISL_2659923,
EPI_ISL_2659924,	EPI_ISL_2677318,	EPI_ISL_2718547,
EPI_ISL_2718549,	EPI_ISL_2749135,	EPI_ISL_2749136,
EPI_ISL_2749137,	EPI_ISL_2749138,	EPI_ISL_2749139,
EPI_ISL_2766891,	EPI_ISL_2772687,	EPI_ISL_2793811,
EPI_ISL_2820575,	EPI_ISL_2858011,	EPI_ISL_2886746,
EPI_ISL_2904295,	EPI_ISL_2904459,	EPI_ISL_2937365,
EPI_ISL_2958115,	EPI_ISL_2974182,	EPI_ISL_2974823,
EPI_ISL_2999299,	EPI_ISL_3050136,	EPI_ISL_3060276,
EPI_ISL_3102027,	EPI_ISL_3148593,	EPI_ISL_3185949,
EPI_ISL_3189869,	EPI_ISL_3203667,	EPI_ISL_3215752,
EPI_ISL_3255044,	EPI_ISL_3285991,	EPI_ISL_3305291,
EPI_ISL_3396879,	EPI_ISL_3447848,	EPI_ISL_3447849,
EPI_ISL_3447853,	EPI_ISL_3447868,	EPI_ISL_3514164,
EPI_ISL_3538952,	EPI_ISL_3563989,	EPI_ISL_3564263,
EPI_ISL_3631514,	EPI_ISL_3642575,	EPI_ISL_3675483,
EPI_ISL_3733692,	EPI_ISL_3766703,	EPI_ISL_3796518,
EPI_ISL_3823148,	EPI_ISL_3834624,	EPI_ISL_3863103,
EPI_ISL_3865686,	EPI_ISL_3884022,	EPI_ISL_3934401,
EPI_ISL_3934402,	EPI_ISL_3940178,	EPI_ISL_3959032,
EPI_ISL_3973828,	EPI_ISL_4092950,	EPI_ISL_4101925,
EPI_ISL_4154911,	EPI_ISL_415709,	EPI_ISL_4169885,
EPI_ISL_4169986,	EPI_ISL_4172121,	EPI_ISL_417443,
EPI_ISL_4180537,	EPI_ISL_4205108,	EPI_ISL_4227754,
EPI_ISL_4236860,	EPI_ISL_4279957,	EPI_ISL_4281523,
EPI_ISL_4340215,	EPI_ISL_4344416,	EPI_ISL_4397068,
EPI_ISL_4404022,	EPI_ISL_4439252,	EPI_ISL_4440355,
EPI_ISL_4440571,	EPI_ISL_444274,	EPI_ISL_447903,
EPI_ISL_4482057,	EPI_ISL_4516894,	EPI_ISL_4517077,
EPI_ISL_4522543,	EPI_ISL_4522912,	EPI_ISL_4550637,
EPI_ISL_455793,	EPI_ISL_4621066,	EPI_ISL_4628409,
EPI_ISL_4639243,	EPI_ISL_4677908,	EPI_ISL_471513,
EPI_ISL_4736248,	EPI_ISL_4754069,	EPI_ISL_4756596,
EPI_ISL_4778830,	EPI_ISL_4818989,	EPI_ISL_4891301,
EPI_ISL_493080,	EPI_ISL_4931280,	EPI_ISL_4931455,
EPI_ISL_4931484,	EPI_ISL_4992415,	EPI_ISL_5023562,
EPI_ISL_5050968,	EPI_ISL_510689,	EPI_ISL_5133657,
EPI_ISL_5169301,	EPI_ISL_5191888,	EPI_ISL_5195707,
EPI_ISL_5227881,	EPI_ISL_5305513,	EPI_ISL_5305556,
EPI_ISL_5384893,	EPI_ISL_541970,	EPI_ISL_5441707,
EPI_ISL_5441708,	EPI_ISL_5441709,	EPI_ISL_5441710,
EPI_ISL_5441713,	EPI_ISL_5441716,	EPI_ISL_5441717,
EPI_ISL_5441718,	EPI_ISL_5441720,	EPI_ISL_5441722,
EPI_ISL_5441723,	EPI_ISL_5441724,	EPI_ISL_5441726,
EPI_ISL_5441727,	EPI_ISL_5441729,	EPI_ISL_5441732,
EPI_ISL_5441733,	EPI_ISL_5441734,	EPI_ISL_5441735,
EPI_ISL_5441736,	EPI_ISL_5441746,	EPI_ISL_5441749,
EPI_ISL_5441750,	EPI_ISL_5441751,	EPI_ISL_5441752,
EPI_ISL_5441753,	EPI_ISL_5441757,	EPI_ISL_5441760,
EPI_ISL_5441764,	EPI_ISL_5441780,	EPI_ISL_5441781,
EPI_ISL_5441782,	EPI_ISL_5441783,	EPI_ISL_5515002,
EPI_ISL_5519972,	EPI_ISL_5534262,	EPI_ISL_5614578,

EPI_ISL_5697612,	EPI_ISL_5780076,	EPI_ISL_578177,
EPI_ISL_5906012,	EPI_ISL_5922571,	EPI_ISL_5931938,
EPI_ISL_5948487,	EPI_ISL_601791,	EPI_ISL_6035472,
EPI_ISL_603638,	EPI_ISL_6067310,	EPI_ISL_6073488,
EPI_ISL_6073647,	EPI_ISL_6085281,	EPI_ISL_6164232,
EPI_ISL_6167579,	EPI_ISL_6219748,	EPI_ISL_6252380,
EPI_ISL_6257598,	EPI_ISL_6262484,	EPI_ISL_6336576,
EPI_ISL_634863,	EPI_ISL_6382348,	EPI_ISL_641315,
EPI_ISL_6537121,	EPI_ISL_6673320,	EPI_ISL_6709715,
EPI_ISL_6729768,	EPI_ISL_6782421,	EPI_ISL_6785019,
EPI_ISL_6801038,	EPI_ISL_6917541,	EPI_ISL_6925434,
EPI_ISL_6955691,	EPI_ISL_7025895,	EPI_ISL_7043618,
EPI_ISL_7043718,	EPI_ISL_7091307,	EPI_ISL_7091320,
EPI_ISL_7091334,	EPI_ISL_7121427,	EPI_ISL_7155501,
EPI_ISL_7169349,	EPI_ISL_7264175,	EPI_ISL_7264177,
EPI_ISL_7264180,	EPI_ISL_7264576,	EPI_ISL_7265072,
EPI_ISL_7270157,	EPI_ISL_7409466,	EPI_ISL_7540782,
EPI_ISL_7575527,	EPI_ISL_7590941,	EPI_ISL_7637869,
EPI_ISL_7670242,	EPI_ISL_7697486,	EPI_ISL_7698630,
EPI_ISL_7748806,	EPI_ISL_7759536,	EPI_ISL_7759688,
EPI_ISL_7808991,	EPI_ISL_7842549,	EPI_ISL_7999577,
EPI_ISL_8034850,	EPI_ISL_8217093,	EPI_ISL_8225229,
EPI_ISL_8227160,	EPI_ISL_8254562,	EPI_ISL_8273887,
EPI_ISL_8316196,	EPI_ISL_8320593,	EPI_ISL_8323986,
EPI_ISL_8323987,	EPI_ISL_8335542,	EPI_ISL_8411573,
EPI_ISL_956130,	EPI_ISL_960197,	EPI_ISL_968173,
EPI_ISL_977058,	EPI_ISL_982899,	EPI_ISL_982903,
EPI_ISL_982930		

Supplementary Table S7. Predicted furin cleavage sites with ProP scores for analysis in Figure 19.

Related HCoV	Accession number	Prototype	Host category	Host order	Predicted FCS	ProP score
OC43/HKU1	JF792617		Reservoir	Rodentia	TAHRARR SV	0.881
OC43/HKU1	NC012936		Reservoir	Rodentia	TAHRARR SV	0.881
OC43/HKU1	JF792616		Reservoir	Rodentia	IAHRARR SV	0.867
OC43/HKU1	KF850449		Reservoir	Rodentia	IAHRARR SV	0.867
OC43/HKU1	MW773844		Reservoir	Rodentia	TSHRARR SI	0.865
HKU1	NC006577	HCoV-HKU1		Primates	SSRRKRR SI	0.877
HKU1	NC006577	HCoV-HKU1		Primates	SSSRKRR RS	0.654
MERS	MH002341		Reservoir	Chiroptera	TSSRVRR AT	0.853
MERS	NC_009020		Reservoir	Chiroptera	TSSRVRR AT	0.822
MERS	KJ473820		Reservoir	Chiroptera	TSSRLRR AT	0.811
MERS	MH002340		Reservoir	Chiroptera	TPSRVLR AA	0.746
MERS	MH002342		Reservoir	Chiroptera	PSARLAR SA	0.701
MERS	EF065510		Reservoir	Chiroptera	TSTRFRR AT	0.588
MERS	EF065511		Reservoir	Chiroptera	TSTRFRR AT	0.588
MERS	EF065512		Reservoir	Chiroptera	TSTRFRR AT	0.588
MERS	KC869678		Reservoir	Chiroptera	TNLRSGR ST	0.572
MERS	MF593268		Reservoir	Chiroptera	TNLRSGR ST	0.572
MERS	MH002342		Reservoir	Chiroptera	RLARSAR SG	0.509
MERS	MK967708		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.607
MERS	KJ477103		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	KU740200		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	KY581695		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	KY581696		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	KY581697		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	KY581698		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	KY581699		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	KY581700		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	KY673149		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598594		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598595		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598596		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563

MERS	MF598699		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598700		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598701		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598702		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598703		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598704		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598705		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598706		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598707		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598708		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598709		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598710		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598711		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598712		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598713		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598714		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598715		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598716		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598717		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598718		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598719		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598720		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598721		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MF598722		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MG923469		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MG923470		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MG923471		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MG923472		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MG923473		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MG923474		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MG923475		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MG923476		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MG923477		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MG923478		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563

MERS	MG923479		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MG923480		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MG923481		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MH259485		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MH259486		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MH734114		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MH734115		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MK357908		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MK357909		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MK564474		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MK564475		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654970		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654971		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654972		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654973		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654974		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654975		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654976		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654977		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654978		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654979		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654980		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654981		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654982		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654983		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654984		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654985		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654986		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654987		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654988		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654989		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654990		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654991		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654992		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563

MERS	MN654993		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654994		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654996		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654997		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654998		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN654999		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN655000		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN655001		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN655002		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN655003		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN655004		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN655005		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN655007		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN655008		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN655009		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN655010		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN655011		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN655012		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN655013		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN655014		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN655015		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN655016		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN655017		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN758606		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN758607		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN758608		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN758609		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN758610		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN758611		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN758612		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN758613		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MN758614		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MT226600		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MT226601		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563

MERS	MT226602		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MT226603		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MT226604		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MT226605		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MT226606		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MT226607		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MW086527		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MW086528		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MW086529		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MW086530		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MW086531		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MW086532		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MW086533		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MW086534		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MW086535		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MW086538		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MW086539		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MW545527		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MW545528		Conspecific CoV	Artiodactyla	LTPRSVR SV	0.563
MERS	MG923466		Conspecific CoV	Artiodactyla	FTPRSVR SV	0.536
MERS	MG923467		Conspecific CoV	Artiodactyla	FTPRSVR SV	0.536
MERS	MN654995		Conspecific CoV	Artiodactyla	FTPRSVR SV	0.536
MERS	MN655006		Conspecific CoV	Artiodactyla	FTPRSVR SV	0.536
MERS	MG923468		Conspecific CoV	Artiodactyla	FTPRSVR SV	0.522
MERS	MH371127	MERS-CoV		Primates	LTPRSVR SV	0.563
MERS	NC019843	MERS-CoV		Primates	LTPRSVR SV	0.563
MERS	NC038294	MERS-CoV		Primates	LTPRSVR SV	0.563
OC43/HKU1	KM349743		Reservoir	Rodentia	STWRAKR DL	0.749
OC43/HKU1	KM349744		Reservoir	Rodentia	STWRAKR DL	0.749
OC43/HKU1	MT820628		Reservoir	Rodentia	ATRAKR DL	0.761
OC43/HKU1	MT820629		Reservoir	Rodentia	ATRAKR DL	0.761
OC43/HKU1	MT820630		Reservoir	Rodentia	ATRSKR DL	0.793
OC43/HKU1	MT820631		Reservoir	Rodentia	ATRAKR DL	0.761
OC43/HKU1	NC026011		Reservoir	Rodentia	STWRAKR DL	0.749

OC43	AB354579		Conspecific CoV	Artiodactyla	TKRRSRR AI	0.776
OC43	AF220295		Conspecific CoV	Artiodactyla	TKRRSRR AI	0.776
OC43	AF391542		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	DQ011855		Conspecific CoV	Artiodactyla	TALRSRR SF	0.715
OC43	DQ811784		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	DQ915164		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	EF424615		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	EF424617		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	EF424619		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	EF424620		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	EF424621		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	EF424623		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	FJ425184		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	FJ425185		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	FJ425186		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	FJ425187		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	FJ425188		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	FJ425189		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	FJ938063		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	FJ938064		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	FJ938065		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	FJ938066		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	KF906249		Conspecific CoV	Artiodactyla	IDRRSRR AI	0.719
OC43	KF906250		Conspecific CoV	Artiodactyla	IDRRSRR AI	0.719
OC43	KU558922		Conspecific CoV	Artiodactyla	IKRRSRR SI	0.829
OC43	KU558923		Conspecific CoV	Artiodactyla	IKRRSRR SI	0.829
OC43	KU886219		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	KX982264		Conspecific CoV	Artiodactyla	TKRRSRR AI	0.776
OC43	KY419103		Conspecific CoV	Artiodactyla	TALRSRR SF	0.715
OC43	KY419104		Conspecific CoV	Artiodactyla	TALRSRR SF	0.715
OC43	KY419105		Conspecific CoV	Artiodactyla	TALRSRR SF	0.715
OC43	KY419106		Conspecific CoV	Artiodactyla	TSLRSRR SL	0.759
OC43	KY419107		Conspecific CoV	Artiodactyla	TSLRSRR SL	0.759
OC43	KY419109		Conspecific CoV	Artiodactyla	TSLRSRR SL	0.759

OC43	KY419110		Conspecific CoV	Artiodactyla	TALRSRR SF	0.715
OC43	KY419112		Conspecific CoV	Artiodactyla	TALRSRR SF	0.715
OC43	KY419113		Conspecific CoV	Artiodactyla	TSLRSRR SL	0.759
OC43	KY994645		Conspecific CoV	Artiodactyla	TALRSRR SF	0.715
OC43	LC494126		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494127		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494128		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494129		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494130		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494131		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494132		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494133		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494134		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494135		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494136		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494137		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494138		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494139		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494140		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494141		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494142		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494143		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494144		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494145		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494147		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494148		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494149		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494150		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494151		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494152		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494153		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494154		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494155		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494156		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851

OC43	LC494191		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	LC494192		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	MF083115		Conspecific CoV	Artiodactyla	TALRSRR SF	0.715
OC43	MG518518		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	MG757138		Conspecific CoV	Artiodactyla	TKRRSRR AI	0.776
OC43	MG757139		Conspecific CoV	Artiodactyla	TKRRSRR AI	0.776
OC43	MG757140		Conspecific CoV	Artiodactyla	TKRRSRR AI	0.776
OC43	MG757141		Conspecific CoV	Artiodactyla	TKRRSRR AI	0.776
OC43	MG757142		Conspecific CoV	Artiodactyla	TKRRSRR AI	0.776
OC43	MH043953		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	MH810163		Conspecific CoV	Artiodactyla	TKRRSRR AI	0.813
OC43	MN514962		Conspecific CoV	Artiodactyla	KERRSRR AI	0.728
OC43	MN514963		Conspecific CoV	Artiodactyla	TDRRSRR AV	0.812
OC43	MN514964		Conspecific CoV	Artiodactyla	TDRRSRR AI	0.752
OC43	MN514965		Conspecific CoV	Artiodactyla	TDRRXRR AI	0.738
OC43	MN514966		Conspecific CoV	Artiodactyla	TDRRSRR AV	0.814
OC43	MN514967		Conspecific CoV	Artiodactyla	TDRRSRR AI	0.752
OC43	MN982198		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	MN982199		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	MW165134		Conspecific CoV	Artiodactyla	TALRSRR SF	0.715
OC43	MW711287		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	NC003045		Conspecific CoV	Artiodactyla	TKRRSRR SI	0.851
OC43	U00735		Conspecific CoV	Artiodactyla	TKRRSRR AI	0.776
OC43	JX860640		Conspecific CoV	Carnivora	TKRRSRR SI	0.851
OC43	KX432213		Conspecific CoV	Carnivora	TKRRSRR SI	0.851
OC43	NC017083		Conspecific CoV	Lagomorpha	TQLRSRR AI	0.627
OC43	LC061272		Conspecific CoV	Perissodactyla	IARRQRR SP	0.795
OC43	LC061273		Conspecific CoV	Perissodactyla	TARRQRR SP	0.819
OC43	LC061274		Conspecific CoV	Perissodactyla	TARRQRR SP	0.819
OC43	LC592689		Conspecific CoV	Perissodactyla	TARRQRR SP	0.819
OC43	NC006213	HCoV-OC43		Primates	SKNRRSR GA	0.529
SARS-related	MT019529	SARS-CoV-2		Primates	NSPRRAR SV	0.620

Materials

Enzymes

Name	Application	Source
Superscript III Reverse Transcriptase	q-RT-PCR	Thermo Fisher Scientific
SuperScript [®] III Platinum [®] Taq High Fidelity Enzyme Mix	1 st round (hemi-) nested RT-PCR assays and quantitative RT-PCR assays	Thermo Fisher Scientific
Platinum [®] Taq DNA Polymerase	2 nd round (hemi-) nested RT- PCR assays	Thermo Fisher Scientific
Proteinase K	RNA extraction	Thermo Fisher Scientific

Chemicals

Name	Application	Source
RNAlater RNA stabilization solution	Storage bat fecal pellets	Qiagen
Viral transport medium	Storage oro-nasopharyngeal swabs	VWR Chemicals
dNTPs 10 mM	PCR	Life Technologies
MgSO ₄ 50 mM	PCR	Sigma
BSA 1 mg/ml	PCR	Roche

Primer

Some of the primers were designed by other members of the institute, including Prof. Dr. Jan Felix Drexler and Dr. Andres Moreira-Soto. Primer sequences that were obtained from published research articles by other research groups are referenced in the following table and in the text. Self-designed primers were either manually designed in Geneious, or by using the primer design tools of Geneious, or IDT (PrimerQuest Tool). Primers were ordered from Integrated DNA Technologies, Inc. (IDT). All sequences in the following table are provided in the 5' to 3' direction.

Project	Sequence Name	Sequence	Direction
MAVS gene sequencing (chapter 2.3.3)			
Assay 1	MAVS as1 F438	GTTACCCYWWGCCTGTCCAGGA	forward
	MAVS as1 R727	AGGCAAGCGGCTTGCCCTGGG	reverse
	MAVS as1 R686nest	TGGAAGGAGACAGWTGGAGACAC	reverse
Assay 2	MAVS as2 F181	CTBMRGCCBMGRMVHGGCTGGGT	forward
	MAVS as2 R707	GGYTKGAMRGA VRCDGWHGGAG	reverse
	MAVS as2 R697nest	AGACAGCTGGAGABAVHGGYCC	reverse
Assay 3	MAVS as3 R1562	GCATCCACCGACACGTACT	reverse
	MAVS SH as3 F513	GTCTGGTGGCTCACTGATGCC	forward
	MAVS SH as3 F473nest	AGCAAGCCCCATCGTCCAGC	forward
	MAVS MG as3 F541	AGTCCTCAGCCCTCTAGAGAGC	forward
	MAVS MG as3 F515nest	CAACCTCACTAAGCTMCCTGC	forward
	MAVS EH as3 F599	AACCAGGCAGTACAGACACAGC	forward
	MAVS EH as3 F529nest	CTGGAGCCCTCTTCTAACATGG	forward
	MAVS HM as3 F585	TCAGAAGAAGGACACAGAACC	forward

	MAVS HM as3 F559nest	CTCAGCCCTATGACCTCCAGTG	forward
	MAVS RF as3 F567	TCTGACCTCCAGTGGGCCTTCG	forward
	MAVS RF as3 F536nest	AGCCCTCCTCTGACATGGCAG	forward
	MAVS PP as3 F551	TGGCGTCCCTCAGCCCTCTG	forward
	MAVS PP as3 F505nest	GCTGACCTGAAGAGACCCACTGG	forward
5'RACE	MAVS 5'R F1 bat	ASCTGGASAGTGATGACG	forward
	MAVS 5'R F2 bat nest	AGTGATGACGTTTGCCGAGG	forward
	MAVS 5'R F01 rod	CTYCAGAARCTGAGCAGCCATG	forward
	MAVS 5'R F02 rod nest	AARCTGAGCAGCCATGAC	forward
	MAVS 5'R bat R103	GATGAGGTATTCCACCCAGCC	reverse
	MAVS 5'R rod R103	GATGAAGACCTCTACCCAGCC	reverse
	MAVS 5'R batrod R13	GCCCKCAGTWGRTCCTGGTCACT	reverse
3' RACE	MAVS 3'R batrod F1214	GCTGCTCYGCRGACCTTGCCAT	forward
	MAVS 3'R batrod F1218	CTCYGCRGACCTTGCCATCAGC	forward
	MAVS EH 3'R F	TTGGGCTTGGAGCCAGAGC	forward
	MAVS EH 3'R Fnest	TCCTGCATGAGTAGCCAGC	forward
	MAVS RF 3'R F	AGTGTCTCCGGCGTCCACAG	forward
	MAVS RF 3'R Fnest	TGCAGACAGCAGCGTTGAGAGC	forward
	MAVS RF 3'R R1852	GAGGGCTGAGGGAGTTCAC	reverse
SARS-CoV-2 full genome sequencing (chapter 3.3.3) (Arctic Network, 2020)			
Pool 1	nCoV-2019 1 LEFT	ACCAACCAACTTTCGATCTCTTGT	forward
	nCoV-2019 1 RIGHT	CATCTTTAAGATGTTGACGTGCCTC	reverse
	nCoV-2019 3 LEFT	CGGTAATAAAGGAGCTGGTGGC	forward
	nCoV-2019 3 RIGHT	AAGGTGTCTGCAATTCATAGCTCT	reverse
	nCoV-2019 5 LEFT	TGGTGAAACTTCATGGCAGACG	forward
	nCoV-2019 5 RIGHT	ATTGATGTTGACTTTCTCTTTTTGGAGT	reverse
	nCoV-2019 7 LEFT	ATCAGAGGCTGCTCGTGTGTA	forward
	nCoV-2019 7 RIGHT	TGCACAGGTGACAATTTGTCCA	reverse
	nCoV-2019 9 LEFT	TCCCACAGAAGTGTTAACAGAGGA	forward
	nCoV-2019 9 RIGHT	ATGACAGCATCTGCCACAACAC	reverse
	nCoV-2019 11 LEFT	GGAATTTGGTGCCACTTCTGCT	forward
	nCoV-2019 11 RIGHT	TCATCAGATTCAACTTGCATGGCA	reverse
	nCoV-2019 13 LEFT	TCGCACAAATGTCTACTTAGCTGT	forward
	nCoV-2019 13 RIGHT	ACCACAGCAGTTAAAACACCCT	reverse
	nCoV-2019 15 LEFT	ACAGTGCTTAAAAAGTGTAAGTGCC	forward
	nCoV-2019 15 RIGHT	AACAGAACTGTAGCTGGCACT	reverse
	nCoV-2019 17 LEFT	CTTCTTTCTTTGAGAGAAGTGAGGACT	forward
	nCoV-2019 17 RIGHT	TTTGTTGGAGTGTTAACAATGCAGT	reverse
	nCoV-2019 19 LEFT	GCTGTTATGTACATGGGCACACT	forward
	nCoV-2019 19 RIGHT	TGTCCAACCTAGGGTCAATTTCTGT	reverse
	nCoV-2019 21 LEFT	TGGCTATTGATTATAAACACTACACACC	forward
	nCoV-2019 21 RIGHT	TAGATCTGTGTGGCCAACCTCT	reverse
	nCoV-2019 23 LEFT	ACAACACTAACATAGTTACACGGTGT	forward
	nCoV-2019 23 RIGHT	ACCAGTACAGTAGGTTGCAATAGTG	reverse
	nCoV-2019 25 LEFT	GCAATTGTTTTTCAGCTATTTTGCAGT	forward
	nCoV-2019 25 RIGHT	ACTGTAGTGACAAGTCTCTCGCA	reverse
	nCoV-2019 27 LEFT	ACTACAGTCAGCTTATGTGTCAACC	forward
	nCoV-2019 27 RIGHT	AATACAAGCACCAAGGTCACGG	reverse
	nCoV-2019 29 LEFT	ACTTGTGTTCTTTTTGTTGCTGC	forward
	nCoV-2019 29 RIGHT	AGTGTACTCTATAAGTTTTGATGGTGTGT	reverse
	nCoV-2019 31 LEFT	TTCTGAGTACTGTAGGCACGGC	forward
	nCoV-2019 31 RIGHT	ACAGAATAAACACCAGGTAAGAATGAGT	reverse
	nCoV-2019 33 LEFT	ACTTTTGAAGAAGCTGCGCTGT	forward
	nCoV-2019 33 RIGHT	TGGACAGTAAACTACGTCATCAAGC	reverse

nCoV-2019 35 LEFT	TGTTTCGCATTCAACCAGGACAG	forward
nCoV-2019 35 RIGHT	ACTTCATAGCCACAAGGTAAAGTCA	reverse
nCoV-2019 37 LEFT	ACACACCACTGGTTGTTACTCAC	forward
nCoV-2019 37 RIGHT	GTCCACACTCTCCTAGCACCAT	reverse
nCoV-2019 39 LEFT	AGTATTGCCCTATTTTCTTCATAACTGGT	forward
nCoV-2019 39 RIGHT	TGTAACCTGGACACATTGAGCCC	reverse
nCoV-2019 41 LEFT	GTTCCCTTCCATCATATGCAGCT	forward
nCoV-2019 41 RIGHT	TGGTATGACAACCATTAGTTTGGCT	reverse
nCoV-2019 43 LEFT	TACGACAGATGTCTTGTGCTGC	forward
nCoV-2019 43 RIGHT	AGCAGCATCTACAGCAAAGCA	reverse
nCoV-2019 45 LEFT	TACCTACAACCTGTGCTAATGACCC	forward
nCoV-2019 45 RIGHT	AAATTGTTTCTTCATGTTGGTAGTTAGAGA	reverse
nCoV-2019 47 LEFT	AGGACTGGTATGATTTTGTAGAAAACCC	forward
nCoV-2019 47 RIGHT	AATAACGGTCAAAGAGTTTTAACCTCTC	reverse
nCoV-2019 49 LEFT	AGGAATTACTIONTGTGTATGCTGCTGA	forward
nCoV-2019 49 RIGHT	TGACGATGACTTGGTTAGCATTAAATACA	reverse
nCoV-2019 51 LEFT	TCAATAGCCGCCACTAGAGGAG	forward
nCoV-2019 51 RIGHT	AGTGCATTAACATTGGCCGTGA	reverse
nCoV-2019 53 LEFT	AGCAAAATGTTGGACTGAGACTGA	forward
nCoV-2019 53 RIGHT	AGCCTCATAAACTCAGGTTCCC	reverse
nCoV-2019 55 LEFT	ACTCAACTTTACTTAGGAGGTATGAGCT	forward
nCoV-2019 55 RIGHT	GGTGTACTCTCCTATTTGTACTTTACTGT	reverse
nCoV-2019 57 LEFT	ATTCTACACTCCAGGGACCACC	forward
nCoV-2019 57 RIGHT	GTAATTGAGCAGGGTCGCCAAT	reverse
nCoV-2019 59 LEFT	TCACGCATGATGTTTCATCTGCA	forward
nCoV-2019 59 RIGHT	AAGAGTCCTGTTACATTTTCAGCTTG	reverse
nCoV-2019 61 LEFT	TGTTTATCACCCGCGAAGAAGC	forward
nCoV-2019 61 RIGHT	ATCACATAGACAACAGGTGCGC	reverse
nCoV-2019 63 LEFT	TGTTAAGCGTGTGACTGGACT	forward
nCoV-2019 63 RIGHT	ACAACTGCCACCATCACAAACC	reverse
nCoV-2019 65 LEFT	GCTGGCTTTAGCTTGTGGGTTT	forward
nCoV-2019 65 RIGHT	TGTCAGTCATAGAACAACACCAATAGT	reverse
nCoV-2019 67 LEFT	GTTGTCCAACAATTACCTGAACTTACT	forward
nCoV-2019 67 RIGHT	CAACCTTAGAACTACAGATAAATCTTGGG	reverse
nCoV-2019 69 LEFT	TGTCGCAAAATATACTCAACTGTGTCA	forward
nCoV-2019 69 RIGHT	TCTTTATAGCCACGGAACCTCCA	reverse
nCoV-2019 71 LEFT	ACAAATCCAATTACAGTTGTCTTCCTATTC	forward
nCoV-2019 71 RIGHT	TGGAAAAGAAAGGTAAGAACAAGTCCT	reverse
nCoV-2019 73 LEFT	CAATTTTGTAAATGATCCATTTTGGGTGT	forward
nCoV-2019 73 RIGHT	CACCAGCTGTCCAACCTGAAGA	reverse
nCoV-2019 75 LEFT	AGAGTCCAACCAACAGAATCTATTGT	forward
nCoV-2019 75 RIGHT	ACCACCAACCTTAGAATCAAGATTGT	reverse
nCoV-2019 77 LEFT	CCAGCAACTGTTTGTGGACCTA	forward
nCoV-2019 77 RIGHT	CAGCCCCTATTAACAGCCTGC	reverse
nCoV-2019 79 LEFT	GTGGTGATTCAACTGAATGCAGC	forward
nCoV-2019 79 RIGHT	CATTTTCATCTGTGAGCAAAGGTGG	reverse
nCoV-2019 81 LEFT	GCACTTGGAACCTTCAAGATGTGG	forward
nCoV-2019 81 RIGHT	GTGAAGTTCTTTTCTTGTGCAGGG	reverse
nCoV-2019 83 LEFT	TCCTTTGCAACCTGAATTAGACTCA	forward
nCoV-2019 83 RIGHT	TTTGACTCCTTTGAGCACTGGC	reverse
nCoV-2019 85 LEFT	ACTAGCACTCTCCAAGGGTGT	forward
nCoV-2019 85 RIGHT	ACACAGTCTTTTACTCCAGATTCCC	reverse
nCoV-2019 87 LEFT	CGACTACTAGCGTGCCTTTGTA	forward
nCoV-2019 87 RIGHT	ACTAGGTTCCATTGTTCAAGGAGC	reverse

	nCoV-2019 89 LEFT	GTACGCGTTCCATGTGGTCATT	forward
	nCoV-2019 89 RIGHT	ACCTGAAAGTCAACGAGATGAAACA	reverse
	nCoV-2019 91 LEFT	TCACTACCAAGAGTGTGTTAGAGGT	forward
	nCoV-2019 91 RIGHT	TTCAAGTGAGAACCAAAAAGATAATAAGCA	reverse
	nCoV-2019 93 LEFT	TGAGGCTGGTTCTAAATCACCCA	forward
	nCoV-2019 93 RIGHT	AGGTCTTCCTTGCCATGTTGAG	reverse
	nCoV-2019 95 LEFT	TGAGGGAGCCTTGAATACACCA	forward
	nCoV-2019 95 RIGHT	CAGTACGTTTTTGCCGAGGCTT	reverse
	nCoV-2019 97 LEFT	TGGATGACAAAGATCCAAATTTCAAAGA	forward
	nCoV-2019 97 RIGHT	ACACACTGATTAAGATTGCTATGTGAG	reverse
Pool 2	nCoV-2019 2 LEFT	CTGTTTTACAGGTTTCGCGACGT	forward
	nCoV-2019 2 RIGHT	TAAGGATCAGTGCCAAGCTCGT	reverse
	nCoV-2019 4 LEFT	GGTGTATACTGCTGCCGTGAAC	forward
	nCoV-2019 4 RIGHT	CACAAGTAGTGGCACCTTCTTTAGT	reverse
	nCoV-2019 6 LEFT	GGTGTGTTGGAGAAGGTTCCG	forward
	nCoV-2019 6 RIGHT	TAGCGGCCTTCTGTAAAACACG	reverse
	nCoV-2019 8 LEFT	AGAGTTTCTTAGAGACGGTTGGGA	forward
	nCoV-2019 8 RIGHT	GCTTCAACAGCTTCACTAGTAGGT	reverse
	nCoV-2019 10 LEFT	TGAGAAGTGCTCTGCCTATACAGT	forward
	nCoV-2019 10 RIGHT	TCATCTAACCAATCTTCTTTGCTCT	reverse
	nCoV-2019 12 LEFT	AAACATGGAGGAGGTGTTGCAG	forward
	nCoV-2019 12 RIGHT	TTCACTCTTCATTTCCAAAAGCTTGA	reverse
	nCoV-2019 14 LEFT	CATCCAGATTCTGCCACTCTTGT	forward
	nCoV-2019 14 RIGHT	AGTTTCCACACAGACAGGCATT	reverse
	nCoV-2019 16 LEFT	AATTTGGAAGAAGCTGCTCGGT	forward
	nCoV-2019 16 RIGHT	CACAACCTTGCCTGTGGAGGTTA	reverse
	nCoV-2019 18 LEFT alt2	ACTTCTATTAATGGGCAGATAACAACCTGT	forward
	nCoV-2019 18 RIGHT	AGCTTGTTTACCACACGTACAAGG	reverse
	nCoV-2019 20 LEFT	ACAAAGAAAACAGTTACACAACAACCA	forward
	nCoV-2019 20 RIGHT	ACGTGGCTTTATTAGTTGCATTGTT	reverse
	nCoV-2019 22 LEFT	ACTACCGAAGTTGTAGGAGACATTATACT	forward
	nCoV-2019 22 RIGHT	ACAGTATTCTTTGCTATAGTAGTCGGC	reverse
	nCoV-2019 24 LEFT	AGGCATGCCTTCTTACTGTACTG	forward
	nCoV-2019 24 RIGHT	ACATTCTAACCATAGCTGAAATCGGG	reverse
	nCoV-2019 26 LEFT	TTGTGATACATTCTGTGCTGGTAGT	forward
	nCoV-2019 26 RIGHT	TCCGCACTATCACCAACATCAG	reverse
	nCoV-2019 28 LEFT	ACATAGAAGTTACTGGCGATAGTTGT	forward
	nCoV-2019 28 RIGHT	TGTTTAGACATGACATGAACAGGTGT	reverse
	nCoV-2019 30 LEFT	GCACAACATAATGGTGACTTTTTGCA	forward
	nCoV-2019 30 RIGHT	ACCACTAGTAGATACAAAACACCAG	reverse
	nCoV-2019 32 LEFT	TGGTGAATACAGTCATGTAGTTGCC	forward
	nCoV-2019 32 RIGHT	AGCACATCACTACGCAACTTTAGA	reverse
	nCoV-2019 34 LEFT	TCCCATCTGGTAAAGTTGAGGGT	forward
	nCoV-2019 34 RIGHT	AGTGAAATTGGGCCTCATAGCA	reverse
	nCoV-2019 36 LEFT	TTAGCTTGGTTGTACGCTGCTG	forward
	nCoV-2019 36 RIGHT	GAACAAAGACCATTGAGTACTCTGGA	reverse
	nCoV-2019 38 LEFT	ACTGTGTTATGTATGCATCAGCTGT	forward
	nCoV-2019 38 RIGHT	CACCAAGAGTCAGTCTAAAGTAGCG	reverse
	nCoV-2019 40 LEFT	TGCACATCAGTAGTCTTACTCTCAGT	forward
	nCoV-2019 40 RIGHT	CATGGCTGCATCACGGTCAAAT	reverse
	nCoV-2019 42 LEFT	TGCAAGAGATGGTTGTGTTCCC	forward
	nCoV-2019 42 RIGHT	CCTACCTCCCTTTGTTGTGTTGT	reverse
	nCoV-2019 44 LEFT	TGCCACAGTACGTCTACAAGCT	forward
	nCoV-2019 44 RIGHT	AACCTTCCACATACCGCAGAC	reverse

nCoV-2019 46 LEFT	TGTCGCTTCCAAGAAAAGGACG	forward
nCoV-2019 46 RIGHT	CACGTTACCTAAGTTGGCGTA	reverse
nCoV-2019 48 LEFT	TGTTGACACTGACTTAACAAAGCCT	forward
nCoV-2019 48 RIGHT	TAGATTACCAGAAGCAGCGTGC	reverse
nCoV-2019 50 LEFT	GTTGATAAGTACTTTGATTGTTACGATGGT	forward
nCoV-2019 50 RIGHT	TAACATGTTGTGCCAACCACCA	reverse
nCoV-2019 52 LEFT	CATCAGGAGATGCCACAACCTGC	forward
nCoV-2019 52 RIGHT	GTTGAGAGCAAATTCATGAGGTCC	reverse
nCoV-2019 54 LEFT	TGAGTTAACAGGACACATGTTAGACA	forward
nCoV-2019 54 RIGHT	AACCAAAAACCTTGCCATTAGCACA	reverse
nCoV-2019 56 LEFT	ACCTAGACCACCACTTAACCGA	forward
nCoV-2019 56 RIGHT	ACACTATGCGAGCAGAAGGGTA	reverse
nCoV-2019 58 LEFT	TGATTTGAGTGTGTCAATGCCAGA	forward
nCoV-2019 58 RIGHT	CTTTTCTCCAAGCAGGGTTACGT	reverse
nCoV-2019 60 LEFT	TGATAGAGACCTTTATGACAAGTTGCA	forward
nCoV-2019 60 RIGHT	GGTACCAACAGCTTCTCTAGTAGC	reverse
nCoV-2019 62 LEFT	GGCACATGGCTTTGAGTTGACA	forward
nCoV-2019 62 RIGHT	GTTGAACCTTTCTACAAGCCGC	reverse
nCoV-2019 64 LEFT	TCGATAGATATCCTGCTAATTCCATTGT	forward
nCoV-2019 64 RIGHT	AGTCTTGTAAGTGTTCAGAGGT	reverse
nCoV-2019 66 LEFT	GGGTGTGGACATTGCTGCTAAT	forward
nCoV-2019 66 RIGHT	TCAATTTCCATTTGACTCCTGGGT	reverse
nCoV-2019 68 LEFT	ACAGGTTTACTAAGTGTGTGTGT	forward
nCoV-2019 68 RIGHT	CTCCTTTATCAGAACCAGCACCA	reverse
nCoV-2019 70 LEFT	ACAAAAGAAAATGACTCTAAAGAGGGTTT	forward
nCoV-2019 70 RIGHT	TGACCTTCTTTTAAAGACATAACAGCAG	reverse
nCoV-2019 72 LEFT	ACACGTGGTGTTTATTACCCTGAC	forward
nCoV-2019 72 RIGHT	ACTCTGAACACTTTCCATCCAAC	reverse
nCoV-2019 74 LEFT	ACATCACTAGGTTTCAAACCTTACTTGC	forward
nCoV-2019 74 RIGHT	GCAACACAGTTGCTGATTCTCTTC	reverse
nCoV-2019 76 LEFT	AGGGCAAACCTGGAAAGATTGCT	forward
nCoV-2019 76 RIGHT	ACACCTGTGCCTGTAAACCAT	reverse
nCoV-2019 78 LEFT	CAACTTACTCCTACTTGGCGTGT	forward
nCoV-2019 78 RIGHT	TGTGTACAAAACCTGCCATATTGCA	reverse
nCoV-2019 80 LEFT	TTGCCTTGGTGATATTGCTGCT	forward
nCoV-2019 80 RIGHT	TGGAGCTAAGTTGTTTAAACAAGCG	reverse
nCoV-2019 82 LEFT	GGGCTATCATCTTATGTCCTTCCCT	forward
nCoV-2019 82 RIGHT	TGCCAGAGATGTCACCTAAATCAA	reverse
nCoV-2019 84 LEFT	TGCTGTAGTTGTCTCAAGGGCT	forward
nCoV-2019 84 RIGHT	AGGTGTGAGTAAACTGTTACAAACAAC	reverse
nCoV-2019 86 LEFT	TCAGGTGATGGCACAACAAGTC	forward
nCoV-2019 86 RIGHT	ACGAAAGCAAGAAAAAGAAGTACGC	reverse
nCoV-2019 88 LEFT	CCATGGCAGATTCCAACGGTAC	forward
nCoV-2019 88 RIGHT	TGGTCAGAATAGTGCCATGGAGT	reverse
nCoV-2019 90 LEFT	ACACAGACCATTCCAGTAGCAGT	forward
nCoV-2019 90 RIGHT	TGAAATGGTGAATTGCCCTCGT	reverse
nCoV-2019 92 LEFT	TTTGTGCTTTTAGCCTTTCTGCT	forward
nCoV-2019 92 RIGHT	AGGTTCTGGCAATTAATTGTAAAAGG	reverse
nCoV-2019 94 LEFT	GGCCCAAGGTTTACCCAATAA	forward
nCoV-2019 94 RIGHT	TTTGGCAATGTTGTTCTTGAGG	reverse
nCoV-2019 96 LEFT	GCCAACAACAACAAGGCCAAAC	forward
nCoV-2019 96 RIGHT	TAGGCTCTGTTGGTGGGAATGT	reverse
nCoV-2019 98 LEFT	AACAATTGCAACAATCCATGAGCA	forward
nCoV-2019 98 RIGHT	TTCTCCTAAGAAGCTATTAAAATCACATGG	reverse

Coronavirus screening; RdRp gene (chapter 4.3.2)			
	PC2S2	equimolar mixture of TTATGGGTTGGGATTATC and TGATGGGATGGGACTATC	forward
	PC2As1	equimolar mixture of TCATCACTCAGAATCATCA, TCATCAGAAAGAATCATCA and TCGTCGGACAAGATCATCA	reverse
Extension SrC partial-RdRp fragment (chapter 4.3.2)			
	SP3080	CTTCTTCTTTGCTCAGGATGGCAATGCTGC	forward
	SP3195	ATACTTTGATTGTTACGATGGTGGCTG	forward
	SP3374	CTATAACTCAAATGAATCTTAAGTATGC	forward
	GrISP1	TTCTTTGCACAGAAGGGTGATGC	reverse
	GrISP2	CTTTGCACAAAAAGGTGATGCWGC	reverse
Amplification SrC S1/S2 genomic region (chapter 4.3.2)			
	panSARS-S1S2-F1	TDGCTGTTGHTAYCARGATGT	forward
	panSARS-S1S2-F2	CARGATGTWAAAYTGYACWGATGT	forward
	panSARS-S1S2-R	CAGATGTACATDKTACAATCBAC	reverse
Extension S2 region (chapter 4.3.2)			
	panSARS-S1S2-R2	AGDCCATTRAACCTTYTGHGCACA	reverse
Amplification SrC full genome (chapter 4.3.2)			
	SrCrhino assay1 F1	TAGAGTTGTTGTCTCYAGTGA	forward
	SrCrhino assay1 R1	CATAATTGTGGTAGACATATAG	reverse
	SrCrhino assay1 R2	CCATCTACATTTCTGAACAC	reverse
	SrCrhino assay2 F1	GATTTATACTAATGCTGCTAAC	forward
	SrCrhino assay2 R1	TAATCAGCAATAACACCHGTCTG	reverse
	SrCrhino assay2 R2	CAATAACACCHGTCTGAGCAGG	reverse
	SrCrhino assay3 F1	CCATCCGTTTATGCGTGGGA	forward
	SrCrhino assay3 F2	GATTGTGTKGCTGATTACTCAGTGC	forward
	SrCrhino assay3 R3	GCACATCTGTACAATTAACATCCTG	reverse
	SrCrhino assay4 F1	TTGTTACATACGGCGGTGTCA	forward
	SrCrhino assay4 R1	CTGGGTCAGCAAGTGTGACT	reverse
	SrCrhino assay4 R2	GATCTACTAGACGGTTTAGCAGG	reverse
	SrCrhino assay5 F1	AGGCCATGTACAAGACTCCTTC	forward
	SrCrhino assay5 R1	CAATTTGAACTTCGGCCTCTACT	reverse
	SrCrhino assay5 R2	TGAACTTCGGCCTCTACTTTGTC	reverse
	SrCrhino assay6 F1	GCCATAGCCCTCAACACACT	forward
	SrCrhino assay6 F2	CACTAGTTAAACAGCTTAGCTC	forward
	SrCrhino assay6 R1	AAGCCAAGCCACACATACCA	reverse
	SrCrhino assay7 F1	TCCTCGTGAAGGTGTCTTCG	forward
	SrCrhino assay7 F2	AATGGCACACACTGGTTTATCAC	forward
	SrCrhino assay7 R1	GTGGCCTGTAAAGGGATGGA	reverse

Commercial Kits

Name	Application	Source
SuperScript® One-Cycle cDNA Kit	cDNA synthesis	Invitrogen
Viral RNA mini kit	Viral RNA extraction	Qiagen
VirSNip Assays	SNP detection	TIB Molbiol
MiSeq reagent kit v2	NGS	Illumina
KAPA Frag Kit and KAPA Hyper Prep kit	Library preparation for Illumina sequencing	Roche Molecular Diagnostics
LunaScript® RT SuperMix Kit	cDNA synthesis	New England Biolabs Inc.
EasySeq™ SARS-CoV-2 WGS Library Prep Kit	NGS	NimaGen B.V.

Detection Kit for 2019 Novel Coronavirus (2019-nCoV) RNA (PCR-Fluorescence Probing)	SARS-CoV-2 RT-PCR	Da An Gene
RealSart SARS-CoV-2 RT-PCR kit 1.0	SARS-CoV-2 RT-PCR	Altona Diagnostics
SarbecoV E-gene kit	SARS-CoV-2 RT-PCR	TIB Molbiol
SARS-CoV-2 RdRp kit	SARS-CoV-2 RT-PCR	TIB Molbiol
MagNA Pure 96 DNA and Viral NA SV	Viral RNA extraction	Roche Diagnostics
MagNA Pure 96 DNA and Viral NA LV	Viral RNA extraction	Roche Diagnostics
OneStep RT-PCR Kit	PCR	Qiagen
SSIII One-Step Kit	PCR	Thermo Fisher
RNeasy Kit	Total RNA extraction from cell cultured cells	Qiagen

Software

Software	Source	Identifier
PRISM 6	GraphPad	https://www.graphpad.com/
MrBayes V3	MrBayes	(Ronquist and Huelsenbeck, 2003)
TreeAnnotator	BEAST	Drummond and Rambaut, 2007
MEGA7	Mega	https://www.megasoftware.net/
SSE V1.2	SSE	(Simmonds, 2012)
Geneious 6.1.8	Geneious	https://www.geneious.com
FigTree V1.4.3	FigTree	http://tree.bio.ed.ac.uk/software/
Datamonkey	N/A	http://classic.datamonkey.org/
CodeML	pamlX	http://abacus.gene.ucl.ac.uk/software/paml.html
CorelDRAW 2017	Corel	https://www.coreldraw.com
EndNote 20	Clarivate Analytics	https://www.endnote.de/
R		https://www.r-project.org/
BEAST2	BEAST2	https://www.beast2.org/
Pangolin COVID-19 Lineage Assigner version 3.0.2	Centre for Genomic Pathogen Surveillance	https://pangolin.cog-uk.io

5.2.1.1 Technical equipment

Equipment	Model	Source
Real-time PCR Cycler	LightCycler 480 II	Roche Diagnostics
Vortexer	VV2	VWR
Pipettes	Research Plus, 100-1000 µl, 10-100 µl, 0.5-10 µl, 0.1-2 µl	Eppendorf
PCR cycler	Mastercycler Nexus Gradient and EP Gradient S	Eppendorf
NGS sequencing platform	MiSeq System	Illumina
Multichannel Pipette	Xplorer	Eppendorf
Freezer	-20er	Liebherr
Freezer	-80er, Model U725	New Brunswick
Centrifuges	Eppendorf 5424; 5430R; 5810R	Eppendorf
Nucleic acid extractor	MagNAPure 96	Roche Diagnostics
Laminar flow	Herasafe™ KS12/KS18	Thermo Fisher
Fluorometer	Qubit® 4	Thermo Fisher

REFERENCES

- Al-Aly, Z., Xie, Y., & Bowe, B. (2021). High-dimensional characterization of post-acute sequelae of COVID-19. *Nature*, *594*(7862), 259-264. doi:10.1038/s41586-021-03553-9
- Ali, I. K., McKendrick, L., Morley, S. J., & Jackson, R. J. (2001). Activity of the hepatitis A virus IRES requires association between the cap-binding translation initiation factor (eIF4E) and eIF4G. *J Virol*, *75*(17), 7854-7863. doi:10.1128/jvi.75.17.7854-7863.2001
- Anthony, S. J., St Leger, J. A., Liang, E., Hicks, A. L., Sanchez-Leon, M. D., Jain, K., Lefkowitz, J. H., Navarrete-Macias, I., Knowles, N., Goldstein, T., Pugliares, K., Ip, H. S., Rowles, T., & Lipkin, W. I. (2015). Discovery of a Novel Hepatovirus (Phopivirus of Seals) Related to Human Hepatitis A Virus. *MBio*, *6*(4). doi:10.1128/mBio.01180-15
- Anti, P., Owusu, M., Agbenyega, O., Annan, A., Badu, E. K., Nkrumah, E. E., Tschapka, M., Oppong, S., Adu-Sarkodie, Y., & Drosten, C. (2015). Human-Bat Interactions in Rural West Africa. *Emerg Infect Dis*, *21*(8), 1418-1421. doi:10.3201/eid2108.142015
- Antia, R., Regoes, R. R., Koella, J. C., & Bergstrom, C. T. (2003). The role of evolution in the emergence of infectious diseases. *Nature*, *426*(6967), 658-661. doi:10.1038/nature02104
- Aragones, L., Guix, S., Ribes, E., Bosch, A., & Pinto, R. M. (2010). Fine-tuning translation kinetics selection as the driving force of codon usage bias in the hepatitis A virus capsid. *PLoS Pathog*, *6*(3), e1000797. doi:10.1371/journal.ppat.1000797
- Armstrong, G. L., & Bell, B. P. (2002). Hepatitis A virus infections in the United States: model-based estimates and implications for childhood immunization. *Pediatrics*, *109*(5), 839-845. doi:10.1542/peds.109.5.839
- Athey, J., Alexaki, A., Osipova, E., Rostovtsev, A., Santana-Quintero, L. V., Katneni, U., Simonyan, V., & Kimchi-Sarfaty, C. (2017). A new and updated resource for codon usage tables. *BMC Bioinformatics*, *18*(1), 391. doi:10.1186/s12859-017-1793-7
- Atkinson, N. J., Witteveldt, J., Evans, D. J., & Simmonds, P. (2014). The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. *Nucleic Acids Res*, *42*(7), 4527-4545. doi:10.1093/nar/gku075
- Augustin Anoh, E., Schubert, G., Wayoro, O., Pacôme, M., Belarbi, E., Sachse, A., Calvignac-Spencer, S., Leendertz, F., Diané, B., & Akoua-Koffi, C. (2021). SARS-CoV-2 variants of concern, variants of interest and lineage A.27 are on the rise in Côte d'Ivoire. *medRxiv*.
- Bachmann, M. E., Nielsen, M. R., Cohen, H., Haase, D., Kouassi, J. A. K., Mundry, R., & Kuehl, H. S. (2020). Saving rodents, losing primates—Why we need tailored bushmeat management strategies. *People and Nature*, *2*(4), 889-902. doi:<https://doi.org/10.1002/pan3.10119>
- Baj, A., Novazzi, F., Pasciuta, R., Genoni, A., Ferrante, F. D., Valli, M., Partenope, M., Tripiciano, R., Ciserchia, A., Catanoso, G., Focosi, D., & Maggi, F. (2021). Breakthrough Infections of E484K-Harboring SARS-CoV-2 Delta Variant, Lombardy, Italy. *Emerg Infect Dis*, *27*(12), 3180-3182. doi:10.3201/eid2712.211792
- Baker, R. E., Mahmud, A. S., Miller, I. F., Rajeev, M., Rasambainarivo, F., Rice, B. L., Takahashi, S., Tatem, A. J., Wagner, C. E., Wang, L. F., Wesolowski, A., & Metcalf, C. J. E. (2022). Infectious disease in an era of global change. *Nat Rev Microbiol*, *20*(4), 193-205. doi:10.1038/s41579-021-00639-z
- Balboni, A., Palladini, A., Bogliani, G., & Battilani, M. (2011). Detection of a virus related to betacoronaviruses in Italian greater horseshoe bats. *Epidemiol Infect*, *139*(2), 216-219. doi:10.1017/S0950268810001147

- Bashor, L., Gagne, R. B., Bosco-Lauth, A. M., Bowen, R. A., Stenglein, M., & VandeWoude, S. (2021). SARS-CoV-2 evolution in animals suggests mechanisms for rapid variant selection. *Proc Natl Acad Sci U S A*, *118*(44). doi:10.1073/pnas.2105253118
- Becker, D. J., Albery, G. F., Sjodin, A. R., Poisot, T., Bergner, L. M., Chen, B., Cohen, L. E., Dallas, T. A., Eskew, E. A., Fagre, A. C., Farrell, M. J., Guth, S., Han, B. A., Simmons, N. B., Stock, M., Teeling, E. C., & Carlson, C. J. (2022). Optimising predictive models to prioritise viral discovery in zoonotic reservoirs. *Lancet Microbe*. doi:10.1016/S2666-5247(21)00245-7
- Beerens, N., Heutink, R., Harders, F., Bossers, A., Koch, G., & Peeters, B. (2020). Emergence and Selection of a Highly Pathogenic Avian Influenza H7N3 Virus. *J Virol*, *94*(8). doi:10.1128/JVI.01818-19
- Beerenwinkel, N., Gunthard, H. F., Roth, V., & Metzner, K. J. (2012). Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. *Front Microbiol*, *3*, 329. doi:10.3389/fmicb.2012.00329
- Belalov, I. S., & Lukashev, A. N. (2013). Causes and implications of codon usage bias in RNA viruses. *PLoS One*, *8*(2), e56642. doi:10.1371/journal.pone.0056642
- Bell, M. A. (2009). I.17 Microevolution. In A. L. Simon, R. C. Stephen, H. C. J. Godfray, P. K. Ann, L. Michel, B. L. Jonathan, W. Brian, & S. W. David (Eds.), *The Princeton Guide to Ecology* (pp. 126-133). Princeton: Princeton University Press.
- Belouzard, S., Chu, V. C., & Whittaker, G. R. (2009). Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc Natl Acad Sci U S A*, *106*(14), 5871-5876. doi:10.1073/pnas.0809524106
- Bertram, S., Glowacka, I., Muller, M. A., Lavender, H., Gnirss, K., Nehlmeier, I., Niemeyer, D., He, Y., Simmons, G., Drosten, C., Soilleux, E. J., Jahn, O., Steffen, I., & Pohlmann, S. (2011). Cleavage and activation of the severe acute respiratory syndrome coronavirus spike protein by human airway trypsin-like protease. *J Virol*, *85*(24), 13363-13372. doi:10.1128/JVI.05300-11
- Bhatt, P. R., Scaiola, A., Loughran, G., Leibundgut, M., Kratzel, A., Meurs, R., Dreos, R., O'Connor, K. M., McMillan, A., Bode, J. W., Thiel, V., Gatfield, D., Atkins, J. F., & Ban, N. (2021). Structural basis of ribosomal frameshifting during translation of the SARS-CoV-2 RNA genome. *Science*, *372*(6548), 1306-1313. doi:10.1126/science.abf3546
- Bloom, J. D., Chan, Y. A., Baric, R. S., Bjorkman, P. J., Cobey, S., Deverman, B. E., Fisman, D. N., Gupta, R., Iwasaki, A., Lipsitch, M., Medzhitov, R., Neher, R. A., Nielsen, R., Patterson, N., Stearns, T., van Nimwegen, E., Worobey, M., & Relman, D. A. (2021). Investigate the origins of COVID-19. *Science*, *372*(6543), 694. doi:10.1126/science.abj0016
- Boehm, T. (2012). Evolution of Vertebrate Immunity. *Current Biology*, *22*(17), R722-R732. doi:<https://doi.org/10.1016/j.cub.2012.07.003>
- Boni, M. F., Zhou, Y., Taubenberger, J. K., & Holmes, E. C. (2008). Homologous recombination is very rare or absent in human influenza A virus. *J Virol*, *82*(10), 4807-4811. doi:10.1128/JVI.02683-07
- Bosch, B. J., Bartelink, W., & Rottier, P. J. (2008). Cathepsin L functionally cleaves the severe acute respiratory syndrome coronavirus class I fusion protein upstream of rather than adjacent to the fusion peptide. *J Virol*, *82*(17), 8887-8890. doi:10.1128/JVI.00415-08
- Bowie, A. G., & Unterholzner, L. (2008). Viral evasion and subversion of pattern-recognition receptor signalling. *Nat Rev Immunol*, *8*(12), 911-922. doi:10.1038/nri2436
- Braun, K. M., Moreno, G. K., Wagner, C., Accola, M. A., Rehauer, W. M., Baker, D. A., Koelle, K., O'Connor, D. H., Bedford, T., Friedrich, T. C., & Moncla, L. H. (2021). Acute SARS-CoV-2 infections harbor limited within-host diversity and transmit via tight transmission bottlenecks. *PLoS Pathog*, *17*(8), e1009849. doi:10.1371/journal.ppat.1009849

- Brito, A. F., Semenova, E., Dudas, G., Hassler, G. W., Kalinich, C. C., Kraemer, M. U. G., Ho, J., Tegally, H., Githinji, G., Agoti, C. N., Matkin, L. E., Whittaker, C., Bulgarian, S.-C.-s. g., Communicable Diseases Genomics, N., Project, C.-I., Danish Covid-19 Genome, C., Fiocruz, C.-G. S. N., team, G. c. c., Network for Genomic Surveillance in South, A., Swiss, S.-C.-S. C., Howden, B. P., Sintchenko, V., Zuckerman, N. S., Mor, O., Blankenship, H. M., de Oliveira, T., Lin, R. T. P., Siqueira, M. M., Resende, P. C., Vasconcelos, A. T. R., Spilki, F. R., Aguiar, R. S., Alexiev, I., Ivanov, I. N., Philipova, I., Carrington, C. V. F., Sahadeo, N. S. D., Branda, B., Gurry, C., Maurer-Stroh, S., Naidoo, D., von Eije, K. J., Perkins, M. D., van Kerkhove, M., Hill, S. C., Sabino, E. C., Pybus, O. G., Dye, C., Bhatt, S., Flaxman, S., Suchard, M. A., Grubaugh, N. D., Baele, G., & Faria, N. R. (2022). Global disparities in SARS-CoV-2 genomic surveillance. *Nat Commun*, *13*(1), 7003. doi:10.1038/s41467-022-33713-y
- Brown, E. A., Day, S. P., Jansen, R. W., & Lemon, S. M. (1991). The 5' nontranslated region of hepatitis A virus RNA: secondary structure and elements required for translation in vitro. *J Virol*, *65*(11), 5828-5838.
- Bugembe, D. L., Phan, M. V. T., Ssewanyana, I., Semanda, P., Nansumba, H., Dhaala, B., Nabadda, S., O'Toole, A. N., Rambaut, A., Kaleebu, P., & Cotten, M. (2021). Emergence and spread of a SARS-CoV-2 lineage A variant (A.23.1) with altered spike protein in Uganda. *Nat Microbiol*, *6*(8), 1094-1101. doi:10.1038/s41564-021-00933-9
- Bulmer, M. (1987). Coevolution of codon usage and transfer RNA abundance. *Nature*, *325*(6106), 728-730. doi:10.1038/325728a0
- Burgin, C. J., Colella, J. P., Kahn, P. L., & Upham, N. S. (2018). How many species of mammals are there? *Journal of Mammalogy*, *99*(1), 1-14. doi:10.1093/jmammal/gyx147
- Calisher, C. H., Carroll, D., Colwell, R., Corley, R. B., Daszak, P., Drosten, C., Enjuanes, L., Farrar, J., Field, H., Golding, J., Gorbalenya, A. E., Haagmans, B., Hughes, J. M., Keusch, G. T., Lam, S. K., Lubroth, J., Mackenzie, J. S., Madoff, L., Mazet, J. K., Perlman, S. M., Poon, L., Saif, L., Subbarao, K., & Turner, M. (2021). Science, not speculation, is essential to determine how SARS-CoV-2 reached humans. *Lancet*, *398*(10296), 209-211. doi:10.1016/S0140-6736(21)01419-7
- Calisher, C. H., Childs, J. E., Field, H. E., Holmes, K. V., & Schountz, T. (2006). Bats: important reservoir hosts of emerging viruses. *Clin Microbiol Rev*, *19*(3), 531-545. doi:10.1128/CMR.00017-06
- Calvignac-Spencer, S., Budt, M., Huska, M., Richard, H., Leipold, L., Grabenhenrich, L., Semmler, T., von Kleist, M., Kroger, S., Wolff, T., & Holzer, M. (2021). Rise and Fall of SARS-CoV-2 Lineage A.27 in Germany. *Viruses*, *13*(8). doi:10.3390/v13081491
- Carlson, C. J., Albery, G. F., Merow, C., Trisos, C. H., Zipfel, C. M., Eskew, E. A., Olival, K. J., Ross, N., & Bansal, S. (2022). Climate change increases cross-species viral transmission risk. *Nature*. doi:10.1038/s41586-022-04788-w
- CDC, A. (2021). *Outbreak Brief #84: Coronavirus Disease 2019 (COVID-19) Pandemic*. 24 August 2021. Retrieved from <https://africacdc.org/download/outbreak-brief-84-coronavirus-disease-2019-covid-19-pandemic/>.
- Chausson, A. M., Rowcliffe, J. M., Escoufflaire, L., Wieland, M., & Wright, J. H. (2019). Understanding the Sociocultural Drivers of Urban Bushmeat Consumption for Behavior Change Interventions in Pointe Noire, Republic of Congo. *Human Ecology*, *47*(2), 179-191. doi:10.1007/s10745-019-0061-z
- Cherian, S., Potdar, V., Jadhav, S., Yadav, P., Gupta, N., Das, M., Rakshit, P., Singh, S., Abraham, P., Panda, S., & Team, N. (2021). SARS-CoV-2 Spike Mutations, L452R, T478K, E484Q and P681R, in the Second Wave of COVID-19 in Maharashtra, India. *Microorganisms*, *9*(7). doi:10.3390/microorganisms9071542
- Chiara, M., D'Erchia, A. M., Gissi, C., Manzari, C., Parisi, A., Resta, N., Zambelli, F., Picardi, E., Pavesi, G., Horner, D. S., & Pesole, G. (2021). Next generation sequencing of SARS-CoV-2 genomes: challenges, applications and opportunities. *Brief Bioinform*, *22*(2), 616-630. doi:10.1093/bib/bbaa297
- Cholankeril, G., Podboy, A., Aivaliotis, V. I., Tarlow, B., Pham, E. A., Spencer, S. P., Kim, D., Hsing, A., & Ahmed, A. (2020). High Prevalence of Concurrent Gastrointestinal Manifestations in Patients With

- Severe Acute Respiratory Syndrome Coronavirus 2: Early Experience From California. *Gastroenterology*, 159(2), 775-777. doi:10.1053/j.gastro.2020.04.008
- Chouikha, A., Lagare, A., Ghedira, K., Diallo, A., Njouom, R., Sankhe, S., Derrar, F., Victoir, K., Dellagi, K., Triki, H., Diagne, M. M., & Consortium, R. (2022). SARS-CoV-2 Lineage A.27: New Data from African Countries and Dynamics in the Context of the COVID-19 Pandemic. *Viruses*, 14(5). doi:10.3390/v14051007
- Cohen, J. I., Rosenblum, B., Feinstone, S. M., Ticehurst, J., & Purcell, R. H. (1989). Attenuation and cell culture adaptation of hepatitis A virus (HAV): a genetic analysis with HAV cDNA. *J Virol*, 63(12), 5364-5370.
- Cohen, L., Benichou, D., & Martin, A. (2002). Analysis of deletion mutants indicates that the 2A polypeptide of hepatitis A virus participates in virion morphogenesis. *J Virol*, 76(15), 7495-7505. doi:10.1128/jvi.76.15.7495-7505.2002
- Corman, V. M., Kallies, R., Philipps, H., Gopner, G., Muller, M. A., Eckerle, I., Brunink, S., Drosten, C., & Drexler, J. F. (2014). Characterization of a novel betacoronavirus related to middle East respiratory syndrome coronavirus in European hedgehogs. *J Virol*, 88(1), 717-724. doi:10.1128/JVI.01600-13
- Corman, V. M., Muth, D., Niemeyer, D., & Drosten, C. (2018). Chapter Eight - Hosts and Sources of Endemic Human Coronaviruses. In M. Kielian, T. C. Mettenleiter, & M. J. Roossinck (Eds.), *Adv Virus Res* (Vol. 100, pp. 163-188): Academic Press.
- Coutard, B., Valle, C., de Lamballerie, X., Canard, B., Seidah, N. G., & Decroly, E. (2020). The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res*, 176, 104742. doi:10.1016/j.antiviral.2020.104742
- Cristina, J., & Costa-Mattioli, M. (2007). Genetic variability and molecular evolution of hepatitis A virus. *Virus Res*, 127(2), 151-157. doi:10.1016/j.virusres.2007.01.005
- Das, A., Hirai-Yuki, A., Gonzalez-Lopez, O., Rhein, B., Moller-Tank, S., Brouillette, R., Hensley, L., Misumi, I., Lovell, W., Cullen, J. M., Whitmire, J. K., Maury, W., & Lemon, S. M. (2017). TIM1 (HAVCR1) Is Not Essential for Cellular Entry of Either Quasi-enveloped or Naked Hepatitis A Virions. *MBio*, 8(5). doi:10.1128/mBio.00969-17
- de Carvalho Dominguez Souza, B. F., Konig, A., Rasche, A., de Oliveira Carneiro, I., Stephan, N., Corman, V. M., Roppert, P. L., Goldmann, N., Kepper, R., Muller, S. F., Volker, C., de Souza, A. J. S., Gomes-Gouvea, M. S., Moreira-Soto, A., Stocker, A., Nassal, M., Franke, C. R., Rebello Pinho, J. R., Soares, M., Geyer, J., Lemey, P., Drosten, C., Netto, E. M., Glebe, D., & Drexler, J. F. (2018). A novel hepatitis B virus species discovered in capuchin monkeys sheds new light on the evolution of primate hepadnaviruses. *J Hepatol*. doi:10.1016/j.jhep.2018.01.029
- de Oliveira Carneiro, I., Sander, A. L., Silva, N., Moreira-Soto, A., Normann, A., Flehmig, B., Lukashev, A. N., Dotzauer, A., Wieseke, N., Franke, C. R., & Drexler, J. F. (2018). A Novel Marsupial Hepatitis A Virus Corroborates Complex Evolutionary Patterns Shaping the Genus Hepatovirus. *J Virol*, 92(13). doi:10.1128/JVI.00082-18
- de Souza Luna, L. K., Heiser, V., Regamey, N., Panning, M., Drexler, J. F., Mulangu, S., Poon, L., Baumgarte, S., Haijema, B. J., Kaiser, L., & Drosten, C. (2007). Generic detection of coronaviruses and differentiation at the prototype strain level by reverse transcription-PCR and nonfluorescent low-density microarray. *J Clin Microbiol*, 45(3), 1049-1052. doi:10.1128/JCM.02426-06
- Delaune, D., Hul, V., Karlsson, E. A., Hassanin, A., Ou, T. P., Baidaliuk, A., Gambaro, F., Prot, M., Tu, V. T., Chea, S., Keatts, L., Mazet, J., Johnson, C. K., Buchy, P., Dussart, P., Goldstein, T., Simon-Loriere, E., & Duong, V. (2021). A novel SARS-CoV-2 related coronavirus in bats from Cambodia. *Nat Commun*, 12(1), 6563. doi:10.1038/s41467-021-26809-4
- Dietzel, E., Schudt, G., Kraehling, V., Matrosovich, M., & Becker, S. (2017). Functional Characterization of Adaptive Mutations during the West African Ebola Virus Outbreak. *J Virol*, 91(2). doi:10.1128/JVI.01913-16

- Drexler, J. F., Corman, V. M., & Drosten, C. (2014). Ecology, evolution and classification of bat coronaviruses in the aftermath of SARS. *Antiviral Res*, *101*, 45-56. doi:10.1016/j.antiviral.2013.10.013
- Drexler, J. F., Corman, V. M., Lukashev, A. N., van den Brand, J. M., Gmyl, A. P., Brunink, S., Rasche, A., Seggewibeta, N., Feng, H., Leijten, L. M., Vallo, P., Kuiken, T., Dotzauer, A., Ulrich, R. G., Lemon, S. M., Drosten, C., & Hepatovirus Ecology, C. (2015). Evolutionary origins of hepatitis A virus in small mammals. *Proc Natl Acad Sci U S A*, *112*(49), 15190-15195. doi:10.1073/pnas.1516992112
- Drexler, J. F., Corman, V. M., Muller, M. A., Lukashev, A. N., Gmyl, A., Coutard, B., Adam, A., Ritz, D., Leijten, L. M., van Riel, D., Kallies, R., Klose, S. M., Gloza-Rausch, F., Binger, T., Annan, A., Adu-Sarkodie, Y., Oppong, S., Bourgarel, M., Rupp, D., Hoffmann, B., Schlegel, M., Kummerer, B. M., Kruger, D. H., Schmidt-Chanasit, J., Setien, A. A., Cottontail, V. M., Hemachudha, T., Wacharapluesadee, S., Osterrieder, K., Bartenschlager, R., Mathee, S., Beer, M., Kuiken, T., Reusken, C., Leroy, E. M., Ulrich, R. G., & Drosten, C. (2013). Evidence for novel hepaciviruses in rodents. *PLoS Pathog*, *9*(6), e1003438. doi:10.1371/journal.ppat.1003438
- Drexler, J. F., Corman, V. M., Muller, M. A., Maganga, G. D., Vallo, P., Binger, T., Gloza-Rausch, F., Cottontail, V. M., Rasche, A., Yordanov, S., Seebens, A., Knornschild, M., Oppong, S., Adu Sarkodie, Y., Pongombo, C., Lukashev, A. N., Schmidt-Chanasit, J., Stocker, A., Carneiro, A. J., Erbar, S., Maisner, A., Fronhoffs, F., Buettner, R., Kalko, E. K., Kruppa, T., Franke, C. R., Kallies, R., Yandoko, E. R., Herrler, G., Reusken, C., Hassanin, A., Kruger, D. H., Mathee, S., Ulrich, R. G., Leroy, E. M., & Drosten, C. (2012). Bats host major mammalian paramyxoviruses. *Nat Commun*, *3*, 796. doi:10.1038/ncomms1796
- Drexler, J. F., Gloza-Rausch, F., Glende, J., Corman, V. M., Muth, D., Goettsche, M., Seebens, A., Niedrig, M., Pfefferle, S., Yordanov, S., Zhelyazkov, L., Hermanns, U., Vallo, P., Lukashev, A., Muller, M. A., Deng, H., Herrler, G., & Drosten, C. (2010). Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. *J Virol*, *84*(21), 11336-11349. doi:10.1128/JVI.00650-10
- Duchene, S., Holt, K. E., Weill, F. X., Le Hello, S., Hawkey, J., Edwards, D. J., Fourment, M., & Holmes, E. C. (2016). Genome-scale rates of evolutionary change in bacteria. *Microb Genom*, *2*(11), e000094. doi:10.1099/mgen.0.000094
- Duckert, P., Brunak, S., & Blom, N. (2004). Prediction of proprotein convertase cleavage sites. *Protein Eng Des Sel*, *17*(1), 107-112. doi:10.1093/protein/gzh013
- Duffy, S. (2018). Why are RNA virus mutation rates so damn high? *PLoS Biol*, *16*(8), e3000003. doi:10.1371/journal.pbio.3000003
- Duffy, S., Shackelton, L. A., & Holmes, E. C. (2008). Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet*, *9*(4), 267-276. doi:10.1038/nrg2323
- Eaton, B. T., Broder, C. C., Middleton, D., & Wang, L. F. (2006). Hendra and Nipah viruses: different and dangerous. *Nat Rev Microbiol*, *4*(1), 23-35. doi:10.1038/nrmicro1323
- El Najjar, F., Lampe, L., Baker, M. L., Wang, L. F., & Dutch, R. E. (2015). Analysis of cathepsin and furin proteolytic enzymes involved in viral fusion protein activation in cells of the bat reservoir host. *PLoS One*, *10*(2), e0115736. doi:10.1371/journal.pone.0115736
- Emerson, S. U., Huang, Y. K., McRill, C., Lewis, M., & Purcell, R. H. (1992). Mutations in both the 2B and 2C genes of hepatitis A virus are involved in adaptation to growth in cell culture. *J Virol*, *66*(2), 650-654.
- Emerson, S. U., Huang, Y. K., Nguyen, H., Brockington, A., Govindarajan, S., St Claire, M., Shapiro, M., & Purcell, R. H. (2002). Identification of VP1/2A and 2C as virulence genes of hepatitis A virus and demonstration of genetic instability of 2C. *J Virol*, *76*(17), 8551-8559. doi:10.1128/jvi.76.17.8551-8559.2002
- Erwin, D. H. (2000). Macroevolution is more than repeated rounds of microevolution. *Evolution & Development*, *2*(2), 78-84. doi:<https://doi.org/10.1046/j.1525-142x.2000.00045.x>

- Fauci, A. S., & Morens, D. M. (2012). The perpetual challenge of infectious diseases. *N Engl J Med*, *366*(5), 454-461. doi:10.1056/NEJMra1108296
- Feng, H., Sander, A. L., Moreira-Soto, A., Yamane, D., Drexler, J. F., & Lemon, S. M. (2019). Hepatovirus 3ABC proteases and evolution of mitochondrial antiviral signaling protein (MAVS). *J Hepatol*, *71*(1), 25-34. doi:10.1016/j.jhep.2019.02.020
- Feng, Z., Hensley, L., McKnight, K. L., Hu, F., Madden, V., Ping, L., Jeong, S. H., Walker, C., Lanford, R. E., & Lemon, S. M. (2013). A pathogenic picornavirus acquires an envelope by hijacking cellular membranes. *Nature*, *496*(7445), 367-371. doi:10.1038/nature12029
- Feng, Z., Hirai-Yuki, A., McKnight, K. L., & Lemon, S. M. (2014). Naked Viruses That Aren't Always Naked: Quasi-Enveloped Agents of Acute Hepatitis. *Annu Rev Virol*, *1*(1), 539-560. doi:10.1146/annurev-virology-031413-085359
- Feng, Z., & Lemon, S. M. (2019). Innate Immunity to Enteric Hepatitis Viruses. *Cold Spring Harb Perspect Med*, *9*(3). doi:10.1101/cshperspect.a033464
- Fensterl, V., Grotheer, D., Berk, I., Schlemminger, S., Vallbracht, A., & Dotzauer, A. (2005). Hepatitis A virus suppresses RIG-I-mediated IRF-3 activation to block induction of beta interferon. *J Virol*, *79*(17), 10968-10977. doi:10.1128/JVI.79.17.10968-10977.2005
- Finkel, Y., Mizrahi, O., Nachshon, A., Weingarten-Gabbay, S., Morgenstern, D., Yahalom-Ronen, Y., Tamir, H., Achdout, H., Stein, D., Israeli, O., Beth-Din, A., Melamed, S., Weiss, S., Israely, T., Paran, N., Schwartz, M., & Stern-Ginossar, N. (2020). The coding capacity of SARS-CoV-2. *Nature*. doi:10.1038/s41586-020-2739-1
- Foley, N. M., Springer, M. S., & Teeling, E. C. (2016). Mammal madness: is the mammal tree of life not yet resolved? *Philos Trans R Soc Lond B Biol Sci*, *371*(1699). doi:10.1098/rstb.2015.0140
- Freuling, C. M., Breithaupt, A., Muller, T., Sehl, J., Balkema-Buschmann, A., Rissmann, M., Klein, A., Wylezich, C., Hoper, D., Wernike, K., Aebischer, A., Hoffmann, D., Friedrichs, V., Dorhoi, A., Groschup, M. H., Beer, M., & Mettenleiter, T. C. (2020). Susceptibility of Raccoon Dogs for Experimental SARS-CoV-2 Infection. *Emerg Infect Dis*, *26*(12), 2982-2985. doi:10.3201/eid2612.203733
- Gallaher, W. R. (2020). A palindromic RNA sequence as a common breakpoint contributor to copy-choice recombination in SARS-COV-2. *Arch Virol*, *165*(10), 2341-2348. doi:10.1007/s00705-020-04750-z
- Ge, X. Y., Li, J. L., Yang, X. L., Chmura, A. A., Zhu, G., Epstein, J. H., Mazet, J. K., Hu, B., Zhang, W., Peng, C., Zhang, Y. J., Luo, C. M., Tan, B., Wang, N., Zhu, Y., Crameri, G., Zhang, S. Y., Wang, L. F., Daszak, P., & Shi, Z. L. (2013). Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature*, *503*(7477), 535-538. doi:nature12711 [pii]
- 10.1038/nature12711
- Geoghegan, J. L., Duchene, S., & Holmes, E. C. (2017). Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. *PLoS Pathog*, *13*(2), e1006215. doi:10.1371/journal.ppat.1006215
- Geoghegan, J. L., Senior, A. M., & Holmes, E. C. (2016). Pathogen population bottlenecks and adaptive landscapes: overcoming the barriers to disease emergence. *Proc Biol Sci*, *283*(1837). doi:10.1098/rspb.2016.0727
- Giandhari, J., Pillay, S., Wilkinson, E., Tegally, H., Sinayskiy, I., Schuld, M., Lourenco, J., Chimukangara, B., Lessells, R., Moosa, Y., Gazy, I., Fish, M., Singh, L., Sedwell Khanyile, K., Fonseca, V., Giovanetti, M., Carlos Junior Alcantara, L., Petruccione, F., & de Oliveira, T. (2021). Early transmission of SARS-CoV-2 in South Africa: An epidemiological and phylogenetic report. *Int J Infect Dis*, *103*, 234-241. doi:10.1016/j.ijid.2020.11.128

- Gibb, R., Redding, D. W., Chin, K. Q., Donnelly, C. A., Blackburn, T. M., Newbold, T., & Jones, K. E. (2020). Zoonotic host diversity increases in human-dominated ecosystems. *Nature*, *584*(7821), 398-402. doi:10.1038/s41586-020-2562-8
- Githinji, G. (2020). Introduction and local transmission of SARS-CoV-2 cases in Kenya. <https://virological.org/t/introduction-and-local-transmission-of-sars-cov-2-cases-in-kenya/497>.
- Glowacka, I., Bertram, S., Muller, M. A., Allen, P., Soilleux, E., Pfefferle, S., Steffen, I., Tsegaye, T. S., He, Y., Gnirss, K., Niemeyer, D., Schneider, H., Drosten, C., & Pohlmann, S. (2011). Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. *J Virol*, *85*(9), 4122-4134. doi:10.1128/JVI.02232-10
- Graham, R. L., & Baric, R. S. (2010). Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *J Virol*, *84*(7), 3134-3146. doi:10.1128/JVI.01394-09
- Grange, Z. L., Goldstein, T., Johnson, C. K., Anthony, S., Gilardi, K., Daszak, P., Olival, K. J., O'Rourke, T., Murray, S., Olson, S. H., Togami, E., Vidal, G., Expert, P., Consortium, P., Mazet, J. A. K., & University of Edinburgh Epigroup members those who wish to remain, a. (2021). Ranking the risk of animal-to-human spillover for newly discovered viruses. *Proc Natl Acad Sci U S A*, *118*(15). doi:10.1073/pnas.2002324118
- Grantham, R., Gautier, C., & Gouy, M. (1980). Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. *Nucleic Acids Res*, *8*(9), 1893-1912. doi:10.1093/nar/8.9.1893
- Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L., Daly, J. M., Mumford, J. A., & Holmes, E. C. (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, *303*(5656), 327-332. doi:10.1126/science.1090727
- Gulyaev, A. P., Spronken, M. I., Funk, M., Fouchier, R. A. M., & Richard, M. (2021). Insertions of codons encoding basic amino acids in H7 hemagglutinins of influenza A viruses occur by recombination with RNA at hotspots near snoRNA binding sites. *RNA*, *27*(2), 123-132. doi:10.1261/rna.077495.120
- Hahn, M. B., Gurley, E. S., Epstein, J. H., Islam, M. S., Patz, J. A., Daszak, P., & Luby, S. P. (2014). The role of landscape composition and configuration on Pteropus giganteus roosting ecology and Nipah virus spillover risk in Bangladesh. *Am J Trop Med Hyg*, *90*(2), 247-255. doi:10.4269/ajtmh.13-0256
- Hale, V. L., Dennis, P. M., McBride, D. S., Nolting, J. M., Madden, C., Huey, D., Ehrlich, M., Grieser, J., Winston, J., Lombardi, D., Gibson, S., Saif, L., Killian, M. L., Lantz, K., Tell, R., Torchetti, M., Robbe-Austerman, S., Nelson, M. I., Faith, S. A., & Bowman, A. S. (2021). SARS-CoV-2 infection in free-ranging white-tailed deer. *Nature*. doi:10.1038/s41586-021-04353-x
- Han, B. A., Schmidt, J. P., Bowden, S. E., & Drake, J. M. (2015). Rodent reservoirs of future zoonotic diseases. *Proc Natl Acad Sci U S A*, *112*(22), 7039-7044. doi:10.1073/pnas.1501598112
- Happi, C., Ihekweazu, C., Oluniyi, P., & Olawoye, I. (2020). SARS-CoV-2 genomes from Nigeria reveal community transmission, multiple virus lineages and spike protein mutation associated with higher transmission and pathogenicity. <https://virological.org/t/sars-cov-2-genomes-from-nigeria-reveal-community-transmission-multiple-virus-lineages-and-spike-protein-mutation-associated-with-higher-transmission-and-pathogenicity/494>.
- Harmon, S. A., Emerson, S. U., Huang, Y. K., Summers, D. F., & Ehrenfeld, E. (1995). Hepatitis A viruses with deletions in the 2A gene are infectious in cultured cells and marmosets. *J Virol*, *69*(9), 5576-5581. doi:10.1128/JVI.69.9.5576-5581.1995
- Harvey, W. T., Carabelli, A. M., Jackson, B., Gupta, R. K., Thomson, E. C., Harrison, E. M., Ludden, C., Reeve, R., Rambaut, A., Consortium, C.-G. U., Peacock, S. J., & Robertson, D. L. (2021). SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol*, *19*(7), 409-424. doi:10.1038/s41579-021-00573-0

- Hirai-Yuki, A., Hensley, L., McGivern, D. R., Gonzalez-Lopez, O., Das, A., Feng, H., Sun, L., Wilson, J. E., Hu, F., Feng, Z., Lovell, W., Misumi, I., Ting, J. P., Montgomery, S., Cullen, J., Whitmire, J. K., & Lemon, S. M. (2016). MAVS-dependent host species range and pathogenicity of human hepatitis A virus. *Science*, 353(6307), 1541-1545. doi:10.1126/science.aaf8325
- Hoffmann, M., Kleine-Weber, H., & Pohlmann, S. (2020a). A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Mol Cell*, 78(4), 779-784 e775. doi:10.1016/j.molcel.2020.04.022
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N. H., Nitsche, A., Muller, M. A., Drosten, C., & Pohlmann, S. (2020b). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*. doi:10.1016/j.cell.2020.02.052
- Holmes, E. C., Zhang, L. Q., Simmonds, P., Ludlam, C. A., & Brown, A. J. (1992). Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. *Proc Natl Acad Sci U S A*, 89(11), 4835-4839.
- Hu, B., Zeng, L. P., Yang, X. L., Ge, X. Y., Zhang, W., Li, B., Xie, J. Z., Shen, X. R., Zhang, Y. Z., Wang, N., Luo, D. S., Zheng, X. S., Wang, M. N., Daszak, P., Wang, L. F., Cui, J., & Shi, Z. L. (2017). Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathog*, 13(11), e1006698. doi:10.1371/journal.ppat.1006698
- Hughes, A. L., Irausquin, S., & Friedman, R. (2010). The evolutionary biology of poxviruses. *Infect Genet Evol*, 10(1), 50-59. doi:10.1016/j.meegid.2009.10.001
- Hughes, A. L., & Yeager, M. (1998). Natural selection at major histocompatibility complex loci of vertebrates. *Annu Rev Genet*, 32, 415-435. doi:10.1146/annurev.genet.32.1.415
- Inzaule, S. C., Tessema, S. K., Kebede, Y., Ogwel Ouma, A. E., & Nkengasong, J. N. (2021). Genomic-informed pathogen surveillance in Africa: opportunities and challenges. *Lancet Infect Dis*, 21(9), e281-e289. doi:10.1016/S1473-3099(20)30939-7
- Isabel, S., Grana-Miraglia, L., Gutierrez, J. M., Bundalovic-Torma, C., Groves, H. E., Isabel, M. R., Eshaghi, A., Patel, S. N., Gubbay, J. B., Poutanen, T., Guttman, D. S., & Poutanen, S. M. (2020). Evolutionary and structural analyses of SARS-CoV-2 D614G spike protein mutation now documented worldwide. *Sci Rep*, 10(1), 14031. doi:10.1038/s41598-020-70827-z
- Isidro, J., Borges, V., Pinto, M., Sobral, D., Santos, J. D., Nunes, A., Mixão, V., Ferreira, R., Santos, D., Duarte, S., Vieira, L., Borrego, M. J., Nuncio, S., de Carvalho, I. L., Pelerito, A., Cordeiro, R., & Gomes, J. P. (2022). Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus. *Nature Medicine*, 28(8), 1569-1572. doi:10.1038/s41591-022-01907-y
- Iwasaki, A., & Medzhitov, R. (2015). Control of adaptive immunity by the innate immune system. *Nature Immunology*, 16(4), 343-353. doi:10.1038/ni.3123
- Jacobsen, K. H. (2018). Globalization and the Changing Epidemiology of Hepatitis A Virus. *Cold Spring Harb Perspect Med*, 8(10). doi:10.1101/cshperspect.a031716
- Jayasundara, D., Saeed, I., Maheswararajah, S., Chang, B. C., Tang, S. L., & Halgamuge, S. K. (2015). ViQuaS: an improved reconstruction pipeline for viral quasispecies spectra generated by next-generation sequencing. *Bioinformatics*, 31(6), 886-896. doi:10.1093/bioinformatics/btu754
- Jenkins, G. M., & Holmes, E. C. (2003). The extent of codon usage bias in human RNA viruses and its evolutionary origin. *Virus Res*, 92(1), 1-7.
- Jiang, Y., Deng, F., Wang, H., & Hu, Z. (2008). An extensive analysis on the global codon usage pattern of baculoviruses. *Arch Virol*, 153(12), 2273-2282. doi:10.1007/s00705-008-0260-1
- Jo, W. K., de Oliveira-Filho, E. F., Rasche, A., Greenwood, A. D., Osterrieder, K., & Drexler, J. F. (2020). Potential zoonotic sources of SARS-CoV-2 infections. *Transbound Emerg Dis*. doi:10.1111/tbed.13872

- Jo, W. K., Drosten, C., & Drexler, J. F. (2021). The evolutionary dynamics of endemic human coronaviruses. *Virus Evol*, 7(1), veab020. doi:10.1093/ve/veab020
- Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., & Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, 451(7181), 990-993. doi:10.1038/nature06536
- Jones, T. C., Biele, G., Muhlemann, B., Veith, T., Schneider, J., Beheim-Schwarzbach, J., Bleicker, T., Tesch, J., Schmidt, M. L., Sander, L. E., Kurth, F., Menzel, P., Schwarzer, R., Zuchowski, M., Hofmann, J., Krumbholz, A., Stein, A., Edelmann, A., Corman, V. M., & Drosten, C. (2021). Estimating infectiousness throughout SARS-CoV-2 infection course. *Science*. doi:10.1126/science.abi5273
- Jonsson, C. B., Figueiredo, L. T., & Vapalahti, O. (2010). A global perspective on hantavirus ecology, epidemiology, and disease. *Clin Microbiol Rev*, 23(2), 412-441. doi:10.1128/CMR.00062-09
- Kaleta, T., Kern, L., Hong, S. L., Holzer, M., Kochs, G., Beer, J., Schnepf, D., Schwemmler, M., Bollen, N., Kolb, P., Huber, M., Ulferts, S., Weigang, S., Dudas, G., Wittig, A., Jaki, L., Padane, A., Lagare, A., Salou, M., Ozer, E. A., Nnaemeka, N., Odoom, J. K., Rutayisire, R., Benkahla, A., Akoua-Koffi, C., Ouedraogo, A. S., Simon-Loriere, E., Enouf, V., Kroger, S., Calvignac-Spencer, S., Baele, G., Panning, M., & Fuchs, J. (2022). Antibody escape and global spread of SARS-CoV-2 lineage A.27. *Nat Commun*, 13(1), 1152. doi:10.1038/s41467-022-28766-y
- Kannan, S. R., Spratt, A. N., Cohen, A. R., Naqvi, S. H., Chand, H. S., Quinn, T. P., Lorson, C. L., Byrareddy, S. N., & Singh, K. (2021). Evolutionary analysis of the Delta and Delta Plus variants of the SARS-CoV-2 viruses. *J Autoimmun*, 124, 102715. doi:10.1016/j.jaut.2021.102715
- Kanteh, A., Manneh, J., Sanyang, B., Kujabi, M. A., Jallow, H. S., K, D. D., Ndure, S. L., & Sesay, A. K. (2022). Simple and structured model to build sequencing capacity in west Africa. *Lancet Glob Health*, 10(9), e1240-e1241. doi:10.1016/S2214-109X(22)00319-9
- Kaplan, G., Totsuka, A., Thompson, P., Akatsuka, T., Moritsugu, Y., & Feinstone, S. M. (1996). Identification of a surface glycoprotein on African green monkey kidney cells as a receptor for hepatitis A virus. *EMBO J*, 15(16), 4282-4296.
- Karlin, S., Doerfler, W., & Cardon, L. R. (1994). Why is CpG suppressed in the genomes of virtually all small eukaryotic viruses but not in those of large eukaryotic viruses? *J Virol*, 68(5), 2889-2897.
- Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*, 30(14), 3059-3066. doi:10.1093/nar/gkf436
- Keusch, G. T., Amuasi, J. H., Anderson, D. E., Daszak, P., Eckerle, I., Field, H., Koopmans, M., Lam, S. K., Das Neves, C. G., Peiris, M., Perlman, S., Wacharapluesadee, S., Yadana, S., & Saif, L. (2022). Pandemic origins and a One Health approach to preparedness and prevention: Solutions based on SARS-CoV-2 and other RNA viruses. *Proc Natl Acad Sci U S A*, 119(42), e2202871119. doi:10.1073/pnas.2202871119
- Kitchen, A., Shackelton, L. A., & Holmes, E. C. (2011). Family level phylogenies reveal modes of macroevolution in RNA viruses. *Proc Natl Acad Sci U S A*, 108(1), 238-243. doi:10.1073/pnas.1011090108
- Konduru, K., & Kaplan, G. G. (2006). Stable growth of wild-type hepatitis A virus in cell culture. *J Virol*, 80(3), 1352-1360. doi:10.1128/JVI.80.3.1352-1360.2006
- Kosakovsky Pond, S. L., & Frost, S. D. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol*, 22(5), 1208-1222. doi:10.1093/molbev/msi105
- Kraemer, M. U. G., Reiner, R. C., Jr., Brady, O. J., Messina, J. P., Gilbert, M., Pigott, D. M., Yi, D., Johnson, K., Earl, L., Marczak, L. B., Shirude, S., Davis Weaver, N., Bisanzio, D., Perkins, T. A., Lai, S., Lu, X., Jones, P., Coelho, G. E., Carvalho, R. G., Van Bortel, W., Marsboom, C., Hendrickx, G., Schaffner, F., Moore, C. G., Nax, H. H., Bengtsson, L., Wetter, E., Tatem, A. J., Brownstein, J. S., Smith, D. L., Lambrechts, L., Cauchemez, S., Linard, C., Faria, N. R., Pybus, O. G., Scott, T. W., Liu, Q., Yu, H., Wint, G. R. W., Hay, S. I., & Golding, N. (2019). Past and future spread of the arbovirus vectors *Aedes aegypti* and *Aedes albopictus*. *Nat Microbiol*, 4(5), 854-863. doi:10.1038/s41564-019-0376-y

- Kryazhimskiy, S., & Plotkin, J. B. (2008). The population genetics of dN/dS. *PLoS Genet*, *4*(12), e1000304. doi:10.1371/journal.pgen.1000304
- Kulkarni, M. A., Walimbe, A. M., Cherian, S., & Arankalle, V. A. (2009). Full length genomes of genotype IIIA Hepatitis A Virus strains (1995-2008) from India and estimates of the evolutionary rates and ages. *Infect Genet Evol*, *9*(6), 1287-1294. doi:10.1016/j.meegid.2009.08.009
- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol*, *33*(7), 1870-1874. doi:10.1093/molbev/msw054
- Lam, T. T., Jia, N., Zhang, Y. W., Shum, M. H., Jiang, J. F., Zhu, H. C., Tong, Y. G., Shi, Y. X., Ni, X. B., Liao, Y. S., Li, W. J., Jiang, B. G., Wei, W., Yuan, T. T., Zheng, K., Cui, X. M., Li, J., Pei, G. Q., Qiang, X., Cheung, W. Y., Li, L. F., Sun, F. F., Qin, S., Huang, J. C., Leung, G. M., Holmes, E. C., Hu, Y. L., Guan, Y., & Cao, W. C. (2020). Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature*, *583*(7815), 282-285. doi:10.1038/s41586-020-2169-0
- Lamers, M. M., Beumer, J., van der Vaart, J., Knoops, K., Puschhof, J., Breugem, T. I., Ravelli, R. B. G., Paul van Schayck, J., Mykytyn, A. Z., Duimel, H. Q., van Donselaar, E., Riesebosch, S., Kuijpers, H. J. H., Schipper, D., van de Wetering, W. J., de Graaf, M., Koopmans, M., Cuppen, E., Peters, P. J., Haagmans, B. L., & Clevers, H. (2020). SARS-CoV-2 productively infects human gut enterocytes. *Science*, *369*(6499), 50-54. doi:10.1126/science.abc1669
- Lanford, R. E., Feng, Z., Chavez, D., Guerra, B., Brasky, K. M., Zhou, Y., Yamane, D., Perelson, A. S., Walker, C. M., & Lemon, S. M. (2011). Acute hepatitis A virus infection is associated with a limited type I interferon response and persistence of intrahepatic viral RNA. *Proc Natl Acad Sci U S A*, *108*(27), 11223-11228. doi:10.1073/pnas.1101939108
- Latinne, A., Hu, B., Olival, K. J., Zhu, G., Zhang, L., Li, H., Chmura, A. A., Field, H. E., Zambrana-Torrel, C., Epstein, J. H., Li, B., Zhang, W., Wang, L. F., Shi, Z. L., & Daszak, P. (2020). Origin and cross-species transmission of bat coronaviruses in China. *Nat Commun*, *11*(1), 4235. doi:10.1038/s41467-020-17687-3
- Lau, S. K., Woo, P. C., Li, K. S., Huang, Y., Tsoi, H. W., Wong, B. H., Wong, S. S., Leung, S. Y., Chan, K. H., & Yuen, K. Y. (2005). Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc Natl Acad Sci U S A*, *102*(39), 14040-14045. doi:10.1073/pnas.0506735102
- Lednicky, J. A., Tagliamonte, M. S., White, S. K., Elbadry, M. A., Alam, M. M., Stephenson, C. J., Bonny, T. S., Loeb, J. C., Telisma, T., Chavannes, S., Ostrov, D. A., Mavian, C., Beau De Rochars, V. M., Salemi, M., & Morris, J. G., Jr. (2021). Independent infections of porcine deltacoronavirus among Haitian children. *Nature*, *600*(7887), 133-137. doi:10.1038/s41586-021-04111-z
- Lee, D. H., Torchetti, M. K., Killian, M. L., Berhane, Y., & Swayne, D. E. (2017a). Highly Pathogenic Avian Influenza A(H7N9) Virus, Tennessee, USA, March 2017. *Emerg Infect Dis*, *23*(11). doi:10.3201/eid2311.171013
- Lee, D. H., Torchetti, M. K., Killian, M. L., & Swayne, D. E. (2017b). Deep sequencing of H7N8 avian influenza viruses from surveillance zone supports H7N8 high pathogenicity avian influenza was limited to a single outbreak farm in Indiana during 2016. *Virology*, *507*, 216-219. doi:10.1016/j.virol.2017.04.025
- Lemon, S. M., LeDuc, J. W., Binn, L. N., Escajadillo, A., & Ishak, K. G. (1982). Transmission of hepatitis A virus among recently captured Panamanian owl monkeys. *J Med Virol*, *10*(1), 25-36. doi:10.1002/jmv.1890100105
- Lester, S. N., & Li, K. (2014). Toll-like receptors in antiviral innate immunity. *J Mol Biol*, *426*(6), 1246-1264. doi:10.1016/j.jmb.2013.11.024
- Li, B., Deng, A., Li, K., Hu, Y., Li, Z., Shi, Y., Xiong, Q., Liu, Z., Guo, Q., Zou, L., Zhang, H., Zhang, M., Ouyang, F., Su, J., Su, W., Xu, J., Lin, H., Sun, J., Peng, J., Jiang, H., Zhou, P., Hu, T., Luo, M., Zhang, Y., Zheng, H., Xiao, J., Liu, T., Tan, M., Che, R., Zeng, H., Zheng, Z., Huang, Y., Yu, J., Yi, L., Wu, J., Chen, J., Zhong, H., Deng, X., Kang, M., Pybus, O. G., Hall, M., Lythgoe, K. A., Li, Y., Yuan, J., He, J., & Lu, J.

- (2022). Viral infection and transmission in a large, well-traced outbreak caused by the SARS-CoV-2 Delta variant. *Nature Communications*, 13(1), 460. doi:10.1038/s41467-022-28089-y
- Li, K., Chen, Z., Kato, N., Gale, M., Jr., & Lemon, S. M. (2005a). Distinct poly(I-C) and virus-activated signaling pathways leading to interferon-beta production in hepatocytes. *J Biol Chem*, 280(17), 16739-16747. doi:10.1074/jbc.M414139200
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S. M., Lau, E. H. Y., Wong, J. Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., Tu, W., Chen, C., Jin, L., Yang, R., Wang, Q., Zhou, S., Wang, R., Liu, H., Luo, Y., Liu, Y., Shao, G., Li, H., Tao, Z., Yang, Y., Deng, Z., Liu, B., Ma, Z., Zhang, Y., Shi, G., Lam, T. T. Y., Wu, J. T., Gao, G. F., Cowling, B. J., Yang, B., Leung, G. M., & Feng, Z. (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med*, 382(13), 1199-1207. doi:10.1056/NEJMoa2001316
- Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., Somasundaran, M., Sullivan, J. L., Luzuriaga, K., Greenough, T. C., Choe, H., & Farzan, M. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*, 426(6965), 450-454. doi:10.1038/nature02145
- Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J. H., Wang, H., Cramer, G., Hu, Z., Zhang, H., Zhang, J., McEachern, J., Field, H., Daszak, P., Eaton, B. T., Zhang, S., & Wang, L. F. (2005b). Bats are natural reservoirs of SARS-like coronaviruses. *Science*, 310(5748), 676-679. doi:10.1126/science.1118391
- Li, X., Wang, L., Liu, P., Li, H., Huo, S., Zong, K., Zhu, S., Guo, Y., Zhang, L., Hu, B., Lan, Y., Chmura, A., Wu, G., Daszak, P., Liu, W. J., & Gao, G. F. (2021). A Novel Potentially Recombinant Rodent Coronavirus with a Polybasic Cleavage Site in the Spike Protein. *J Virol*, 95(22), e0117321. doi:10.1128/JVI.01173-21
- Li, X. D., Sun, L., Seth, R. B., Pineda, G., & Chen, Z. J. (2005c). Hepatitis C virus protease NS3/4A cleaves mitochondrial antiviral signaling protein off the mitochondria to evade innate immunity. *Proc Natl Acad Sci U S A*, 102(49), 17717-17722. doi:10.1073/pnas.0508531102
- Littlejohn, M., Locarnini, S., & Yuen, L. (2016). Origins and Evolution of Hepatitis B Virus and Hepatitis D Virus. *Cold Spring Harb Perspect Med*, 6(1), a021360. doi:10.1101/cshperspect.a021360
- Liu-Helmersson, J., Rocklöv, J., Sewe, M., & Brannström, A. (2019). Climate change may enable *Aedes aegypti* infestation in major European cities by 2100. *Environ Res*, 172, 693-699. doi:10.1016/j.envres.2019.02.026
- Liu, P., Jiang, J. Z., Wan, X. F., Hua, Y., Li, L., Zhou, J., Wang, X., Hou, F., Chen, J., Zou, J., & Chen, J. (2020). Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? *PLoS Pathog*, 16(5), e1008421. doi:10.1371/journal.ppat.1008421
- Liu, X., Wu, Q., & Zhang, Z. (2021). Global Diversification and Distribution of Coronaviruses With Furin Cleavage Sites. *Front Microbiol*, 12, 649314. doi:10.3389/fmicb.2021.649314
- Liu, Y., Liu, J., Plante, K. S., Plante, J. A., Xie, X., Zhang, X., Ku, Z., An, Z., Scharton, D., Schindewolf, C., Widen, S. G., Menachery, V. D., Shi, P. Y., & Weaver, S. C. (2022). The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. *Nature*, 602(7896), 294-299. doi:10.1038/s41586-021-04245-0
- Lu, L., Sikkema, R. S., Velkers, F. C., Nieuwenhuijse, D. F., Fischer, E. A. J., Meijer, P. A., Bouwmeester-Vincken, N., Rietveld, A., Wegdam-Blans, M. C. A., Tolsma, P., Koppelman, M., Smit, L. A. M., Hakze-van der Honing, R. W., van der Poel, W. H. M., van der Spek, A. N., Spierenburg, M. A. H., Molenaar, R. J., Rond, J., Augustijn, M., Woolhouse, M., Stegeman, J. A., Lycett, S., Oude Munnink, B. B., & Koopmans, M. P. G. (2021). Adaptation, spread and transmission of SARS-CoV-2 in farmed minks and associated humans in the Netherlands. *Nat Commun*, 12(1), 6802. doi:10.1038/s41467-021-27096-9
- Luis, A. D., Hayman, D. T., O'Shea, T. J., Cryan, P. M., Gilbert, A. T., Pulliam, J. R., Mills, J. N., Timonin, M. E., Willis, C. K., Cunningham, A. A., Fooks, A. R., Rupprecht, C. E., Wood, J. L., & Webb, C. T. (2013). A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? *Proc Biol Sci*, 280(1756), 20122753. doi:10.1098/rspb.2012.2753

- Lythgoe, K. A., Hall, M., Ferretti, L., de Cesare, M., MacIntyre-Cockett, G., Trebes, A., Andersson, M., Otecko, N., Wise, E. L., Moore, N., Lynch, J., Kidd, S., Cortes, N., Mori, M., Williams, R., Vernet, G., Justice, A., Green, A., Nicholls, S. M., Ansari, M. A., Abeler-Dorner, L., Moore, C. E., Peto, T. E. A., Eyre, D. W., Shaw, R., Simmonds, P., Buck, D., Todd, J. A., Oxford Virus Sequencing Analysis, G., Connor, T. R., Ashraf, S., da Silva Filipe, A., Shepherd, J., Thomson, E. C., Consortium, C.-G. U., Bonsall, D., Fraser, C., & Golubchik, T. (2021). SARS-CoV-2 within-host diversity and transmission. *Science*, 372(6539). doi:10.1126/science.abg0821
- Lytras, S., Xia, W., Hughes, J., Jiang, X., & Robertson, D. L. (2021). The animal origin of SARS-CoV-2. *Science*, 373(6558), 968-970. doi:10.1126/science.abh0117
- Martin, D. P., Weaver, S., Tegally, H., San, J. E., Shank, S. D., Wilkinson, E., Lucaci, A. G., Giandhari, J., Naidoo, S., Pillay, Y., Singh, L., Lessells, R. J., Ngs, S. A., UK, C.-G., Gupta, R. K., Wertheim, J. O., Nekturenko, A., Murrell, B., Harkins, G. W., Lemey, P., MacLean, O. A., Robertson, D. L., de Oliveira, T., & Kosakovsky Pond, S. L. (2021). The emergence and ongoing convergent evolution of the SARS-CoV-2 N501Y lineages. *Cell*, 184(20), 5189-5200 e5187. doi:10.1016/j.cell.2021.09.003
- Matsuyama, S., Nagata, N., Shirato, K., Kawase, M., Takeda, M., & Taguchi, F. (2010). Efficient activation of the severe acute respiratory syndrome coronavirus spike protein by the transmembrane protease TMPRSS2. *J Virol*, 84(24), 12658-12664. doi:10.1128/JVI.01542-10
- McKnight, K. L., & Lemon, S. M. (2018). Hepatitis A Virus Genome Organization and Replication Strategy. *Cold Spring Harb Perspect Med*, 8(12). doi:10.1101/cshperspect.a033480
- Meylan, E., Curran, J., Hofmann, K., Moradpour, D., Binder, M., Bartenschlager, R., & Tschopp, J. (2005). Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature*, 437(7062), 1167-1172. doi:10.1038/nature04193
- Mildenstein, T., Tanshi, I., & Racey, P. A. (2016). Exploitation of Bats for Bushmeat and Medicine. In C. C. Voigt & T. Kingston (Eds.), *Bats in the Anthropocene: Conservation of Bats in a Changing World* (pp. 325-375). Cham: Springer International Publishing.
- Millet, J. K., & Whittaker, G. R. (2014). Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. *Proc Natl Acad Sci USA*, 111(42), 15214-15219. doi:10.1073/pnas.1407087111
- Mlcochova, P., Kemp, S., Dhar, M. S., Papa, G., Meng, B., Ferreira, I., Datir, R., Collier, D. A., Albecka, A., Singh, S., Pandey, R., Brown, J., Zhou, J., Goonawardane, N., Mishra, S., Whittaker, C., Mellan, T., Marwal, R., Datta, M., Sengupta, S., Ponnusamy, K., Radhakrishnan, V. S., Abdullahi, A., Charles, O., Chattopadhyay, P., Devi, P., Caputo, D., Peacock, T., Wattal, D. C., Goel, N., Satwik, A., Vaishya, R., Agarwal, M., Indian, S.-C.-G. C., Genotype to Phenotype Japan, C., Collaboration, C.-N. B. C.-., Mavousian, A., Lee, J. H., Bassi, J., Silacci-Fegni, C., Saliba, C., Pinto, D., Irie, T., Yoshida, I., Hamilton, W. L., Sato, K., Bhatt, S., Flaxman, S., James, L. C., Corti, D., Piccoli, L., Barclay, W. S., Rakshit, P., Agrawal, A., & Gupta, R. K. (2021). SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature*. doi:10.1038/s41586-021-03944-y
- Mollentze, N., Babayan, S. A., & Streicker, D. G. (2021). Identifying and prioritizing potential human-infecting viruses from their genome sequences. *PLoS Biol*, 19(9), e3001390. doi:10.1371/journal.pbio.3001390
- Mollentze, N., & Streicker, D. G. (2020). Viral zoonotic risk is homogenous among taxonomic orders of mammalian and avian reservoir hosts. *Proc Natl Acad Sci U S A*, 117(17), 9423-9430. doi:10.1073/pnas.1919176117
- Monne, I., Fusaro, A., Nelson, M. I., Bonfanti, L., Mulatti, P., Hughes, J., Murcia, P. R., Schivo, A., Valastro, V., Moreno, A., Holmes, E. C., & Cattoli, G. (2014). Emergence of a highly pathogenic avian influenza virus from a low-pathogenic progenitor. *J Virol*, 88(8), 4375-4388. doi:10.1128/JVI.03181-13
- Montagutelli, X., Decaudin, B., Beretta, M., Mouquet, H., & Simon-Lorière, E. (2022). SARS-CoV-2 infection in domestic rats after transmission from their infected owner. *bioRxiv*, 2022.2010.2013.512053. doi:10.1101/2022.10.13.512053

- Mora, C., McKenzie, T., Gaw, I. M., Dean, J. M., von Hammerstein, H., Knudson, T. A., Setter, R. O., Smith, C. Z., Webster, K. M., Patz, J. A., & Franklin, E. C. (2022). Over half of known human pathogenic diseases can be aggravated by climate change. *Nat Clim Chang*, 1-7. doi:10.1038/s41558-022-01426-1
- Moreira-Soto, A., Torres, M. C., Lima de Mendonca, M. C., Mares-Guia, M. A., Dos Santos Rodrigues, C. D., Fabri, A. A., Dos Santos, C. C., Machado Araujo, E. S., Fischer, C., Ribeiro Nogueira, R. M., Drosten, C., Sequeira, P. C., Drexler, J. F., & Bispo de Filippis, A. M. (2018). Evidence for multiple sylvatic transmission cycles during the 2016-2017 yellow fever virus outbreak, Brazil. *Clin Microbiol Infect*, 24(9), 1019 e1011-1019 e1014. doi:10.1016/j.cmi.2018.01.026
- Morse, S. S. (1995). Factors in the emergence of infectious diseases. *Emerg Infect Dis*, 1(1), 7-15. doi:10.3201/eid0101.950102
- Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L., & Scheffler, K. (2013). FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol Biol Evol*, 30(5), 1196-1205. doi:10.1093/molbev/mst030
- Murrell, B., Weaver, S., Smith, M. D., Wertheim, J. O., Murrell, S., Aylward, A., Eren, K., Pollner, T., Martin, D. P., Smith, D. M., Scheffler, K., & Kosakovsky Pond, S. L. (2015). Gene-wide identification of episodic selection. *Mol Biol Evol*, 32(5), 1365-1371. doi:10.1093/molbev/msv035
- Muth, D., Corman, V. M., Roth, H., Binger, T., Dijkman, R., Gottula, L. T., Gloza-Rausch, F., Balboni, A., Battilani, M., Rihtaric, D., Toplak, I., Ameneiros, R. S., Pfeifer, A., Thiel, V., Drexler, J. F., Muller, M. A., & Drosten, C. (2018). Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. *Sci Rep*, 8(1), 15177. doi:10.1038/s41598-018-33487-8
- Nainan, O. V., Margolis, H. S., Robertson, B. H., Balayan, M., & Brinton, M. A. (1991). Sequence analysis of a new hepatitis A virus naturally infecting cynomolgus macaques (*Macaca fascicularis*). *J Gen Virol*, 72 (Pt 7), 1685-1689. doi:10.1099/0022-1317-72-7-1685
- Nao, N., Yamagishi, J., Miyamoto, H., Igarashi, M., Manzoor, R., Ohnuma, A., Tsuda, Y., Furuyama, W., Shigeno, A., Kajihara, M., Kishida, N., Yoshida, R., & Takada, A. (2017). Genetic Predisposition To Acquire a Polybasic Cleavage Site for Highly Pathogenic Avian Influenza Virus Hemagglutinin. *MBio*, 8(1). doi:10.1128/mBio.02298-16
- Network, A. (2020, 24 March 2020). SARS-CoV-2. Retrieved from <https://artic.network/ncov-2019>
- Nguyen, V. K., Parra-Rojas, C., & Hernandez-Vargas, E. A. (2018). The 2017 plague outbreak in Madagascar: Data descriptions and epidemic modelling. *Epidemics*. doi:10.1016/j.epidem.2018.05.001
- Normann, A., Jung, C., Vallbracht, A., & Flehmig, B. (2004). Time course of hepatitis A viremia and viral load in the blood of human hepatitis A patients. *J Med Virol*, 72(1), 10-16. doi:10.1002/jmv.10532
- Nunez, J. J., Fritz, C. L., Knust, B., Buttke, D., Enge, B., Novak, M. G., Kramer, V., Osadebe, L., Messenger, S., Albarino, C. G., Stroher, U., Niemela, M., Amman, B. R., Wong, D., Manning, C. R., Nichol, S. T., Rollin, P. E., Xia, D., Watt, J. P., Vugia, D. J., & Yosemite Hantavirus Outbreak Investigation, T. (2014). Hantavirus infections among overnight visitors to Yosemite National Park, California, USA, 2012. *Emerg Infect Dis*, 20(3), 386-393. doi:10.3201/eid2003.131581
- O'Donnell, J. S., & Chappell, K. J. (2021). Chronic SARS-CoV-2, a Cause of Post-acute COVID-19 Sequelae (Long-COVID)? *Front Microbiol*, 12, 724654. doi:10.3389/fmicb.2021.724654
- Obermeyer, F., Jankowiak, M., Barkas, N., Schaffner, S. F., Pyle, J. D., Yurkovetskiy, L., Bosso, M., Park, D. J., Babadi, M., MacInnis, B. L., Luban, J., Sabeti, P. C., & Lemieux, J. E. (2022). Analysis of 6.4 million SARS-CoV-2 genomes identifies mutations associated with fitness. *Science*, 376(6599), 1327-1332. doi:10.1126/science.abm1208
- Olival, K. J., Hosseini, P. R., Zambrana-Torrel, C., Ross, N., Bogich, T. L., & Daszak, P. (2017). Host and viral traits predict zoonotic spillover from mammals. *Nature*, 546(7660), 646-650. doi:10.1038/nature22975

- One Health High-Level Expert, P., Adisasmito, W. B., Almuhairi, S., Behraves, C. B., Bilivogui, P., Bukachi, S. A., Casas, N., Cediel Becerra, N., Charron, D. F., Chaudhary, A., Ciacci Zanella, J. R., Cunningham, A. A., Dar, O., Debnath, N., Dungu, B., Farag, E., Gao, G. F., Hayman, D. T. S., Khaitsa, M., Koopmans, M. P. G., Machalaba, C., Mackenzie, J. S., Markotter, W., Mettenleiter, T. C., Morand, S., Smolenskiy, V., & Zhou, L. (2022). One Health: A new definition for a sustainable and healthy future. *PLoS Pathog*, *18*(6), e1010537. doi:10.1371/journal.ppat.1010537
- Orton, R. J., Wright, C. F., Morelli, M. J., King, D. J., Paton, D. J., King, D. P., & Haydon, D. T. (2015). Distinguishing low frequency mutations from RT-PCR and sequence errors in viral deep sequencing data. *BMC Genomics*, *16*, 229. doi:10.1186/s12864-015-1456-x
- Oude Munnink, B. B., Sikkema, R. S., Nieuwenhuijse, D. F., Molenaar, R. J., Munger, E., Molenkamp, R., van der Spek, A., Tolsma, P., Rietveld, A., Brouwer, M., Bouwmeester-Vincken, N., Harders, F., Hakze-van der Honing, R., Wegdam-Blans, M. C. A., Bouwstra, R. J., GeurtsvanKessel, C., van der Eijk, A. A., Velkers, F. C., Smit, L. A. M., Stegeman, A., van der Poel, W. H. M., & Koopmans, M. P. G. (2021). Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science*, *371*(6525), 172-177. doi:10.1126/science.abe5901
- Ozer, E. A., Simons, L. M., Adewumi, O. M., Fowotade, A. A., Omoruyi, E. C., Adeniji, J. A., Olayinka, O. A., Dean, T. J., Zayas, J., Bhimalli, P. P., Ash, M. K., Maiga, A. I., Somboro, A. M., Maiga, M., Godzik, A., Schneider, J. R., Mamede, J. I., Taiwo, B. O., Hultquist, J. F., & Lorenzo-Redondo, R. (2022). Multiple expansions of globally uncommon SARS-CoV-2 lineages in Nigeria. *Nat Commun*, *13*(1), 688. doi:10.1038/s41467-022-28317-5
- Pandey, S., Kawai, T., & Akira, S. (2014). Microbial sensing by Toll-like receptors and intracellular nucleic acid sensors. *Cold Spring Harb Perspect Biol*, *7*(1), a016246. doi:10.1101/cshperspect.a016246
- Parrish, C. R., Holmes, E. C., Morens, D. M., Park, E. C., Burke, D. S., Calisher, C. H., Laughlin, C. A., Saif, L. J., & Daszak, P. (2008). Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiol Mol Biol Rev*, *72*(3), 457-470. doi:10.1128/MMBR.00004-08
- Patel, M. R., Loo, Y. M., Horner, S. M., Gale, M., Jr., & Malik, H. S. (2012). Convergent evolution of escape from hepaciviral antagonism in primates. *PLoS Biol*, *10*(3), e1001282. doi:10.1371/journal.pbio.1001282
- Peacock, T. P., Brown, J. C., Zhou, J., Thakur, N., Newman, J., Kugathasan, R., Sukhova, K., Kafrou, M., Bailey, D., & Barclay, W. S. (2022). The SARS-CoV-2 variant, Omicron, shows rapid replication in human primary nasal epithelial cultures and efficiently uses the endosomal route of entry. *bioRxiv*, 2021.2012.2031.474653. doi:10.1101/2021.12.31.474653
- Pekar, J. E., Magee, A., Parker, E., Moshiri, N., Izhikevich, K., Havens, J. L., Gangavarapu, K., Malpica Serrano, L. M., Crits-Christoph, A., Matteson, N. L., Zeller, M., Levy, J. I., Wang, J. C., Hughes, S., Lee, J., Park, H., Park, M. S., Ching Zi Yan, K., Lin, R. T. P., Mat Isa, M. N., Noor, Y. M., Vasylyeva, T. I., Garry, R. F., Holmes, E. C., Rambaut, A., Suchard, M. A., Andersen, K. G., Worobey, M., & Wertheim, J. O. (2022). The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2. *Science*, *377*(6609), 960-966. doi:10.1126/science.abp8337
- Pickering, B., Lung, O., Maguire, F., Kruczkiewicz, P., Kotwa, J. D., Buchanan, T., Gagnier, M., Guthrie, J. L., Jardine, C. M., Marchand-Austin, A., Masse, A., McClinchey, H., Nirmalarajah, K., Aftanas, P., Blais-Savoie, J., Chee, H. Y., Chien, E., Yim, W., Banete, A., Griffin, B. D., Yip, L., Goolia, M., Suderman, M., Pinette, M., Smith, G., Sullivan, D., Rudar, J., Vernygora, O., Adey, E., Nebroski, M., Goyette, G., Finzi, A., Laroche, G., Ariana, A., Vahkal, B., Cote, M., McGeer, A. J., Nituch, L., Mubareka, S., & Bowman, J. (2022). Divergent SARS-CoV-2 variant emerges in white-tailed deer with deer-to-human transmission. *Nat Microbiol*. doi:10.1038/s41564-022-01268-9
- Pinto, R. M., Aragonés, L., Costafreda, M. I., Ribes, E., & Bosch, A. (2007). Codon usage and replicative strategies of hepatitis A virus. *Virus Res*, *127*(2), 158-163. doi:10.1016/j.virusres.2007.04.010
- Pinto, R. M., Perez-Rodriguez, F. J., L, D. A., de Castellarnau, M., Guix, S., & Bosch, A. (2018). Hepatitis A Virus Codon Usage: Implications for Translation Kinetics and Capsid Folding. *Cold Spring Harb Perspect Med*. doi:10.1101/cshperspect.a031781

- Planas, D., Veyer, D., Baidaliuk, A., Staropoli, I., Guivel-Benhassine, F., Rajah, M. M., Planchais, C., Porrot, F., Robillard, N., Puech, J., Prot, M., Gallais, F., Gantner, P., Velay, A., Le Guen, J., Kassis-Chikhani, N., Edriss, D., Belec, L., Seve, A., Courtellemont, L., Pere, H., Hocqueloux, L., Fafi-Kremer, S., Prazuck, T., Mouquet, H., Bruel, T., Simon-Loriere, E., Rey, F. A., & Schwartz, O. (2021). Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature*, *596*(7871), 276-280. doi:10.1038/s41586-021-03777-9
- Plante, J. A., Liu, Y., Liu, J., Xia, H., Johnson, B. A., Lokugamage, K. G., Zhang, X., Muruato, A. E., Zou, J., Fontes-Garfias, C. R., Mirchandani, D., Scharton, D., Bilello, J. P., Ku, Z., An, Z., Kalveram, B., Freiberg, A. N., Menachery, V. D., Xie, X., Plante, K. S., Weaver, S. C., & Shi, P. Y. (2021). Spike mutation D614G alters SARS-CoV-2 fitness. *Nature*, *592*(7852), 116-121. doi:10.1038/s41586-020-2895-3
- Player, M. R., & Torrence, P. F. (1998). The 2-5A system: modulation of viral and cellular processes through acceleration of RNA degradation. *Pharmacol Ther*, *78*(2), 55-113.
- Posada-Cespedes, S., Seifert, D., Topolsky, I., Jablonski, K. P., Metzner, K. J., & Beerenwinkel, N. (2021). V-pipe: a computational pipeline for assessing viral genetic diversity from high-throughput data. *Bioinformatics*. doi:10.1093/bioinformatics/btab015
- Prentis, P. J., Wilson, J. R., Dormontt, E. E., Richardson, D. M., & Lowe, A. J. (2008). Adaptive evolution in invasive species. *Trends Plant Sci*, *13*(6), 288-294. doi:10.1016/j.tplants.2008.03.004
- Price, M. N., Dehal, P. S., & Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol*, *26*(7), 1641-1650. doi:10.1093/molbev/msp077
- Qu, L., Feng, Z., Yamane, D., Liang, Y., Lanford, R. E., Li, K., & Lemon, S. M. (2011). Disruption of TLR3 signaling due to cleavage of TRIF by the hepatitis A virus protease-polymerase processing intermediate, 3CD. *PLoS Pathog*, *7*(9), e1002169. doi:10.1371/journal.ppat.1002169
- Quan, P. L., Firth, C., Conte, J. M., Williams, S. H., Zambrana-Torrel, C. M., Anthony, S. J., Ellison, J. A., Gilbert, A. T., Kuzmin, I. V., Niezgod, M., Osinubi, M. O., Recuenco, S., Markotter, W., Breiman, R. F., Kalemba, L., Malekani, J., Lindblade, K. A., Rostal, M. K., Ojeda-Flores, R., Suzan, G., Davis, L. B., Blau, D. M., Ogunkoya, A. B., Alvarez Castillo, D. A., Moran, D., Ngam, S., Akaibe, D., Agwanda, B., Briese, T., Epstein, J. H., Daszak, P., Rupprecht, C. E., Holmes, E. C., & Lipkin, W. I. (2013). Bats are a major natural reservoir for hepaciviruses and pegiviruses. *Proc Natl Acad Sci U S A*, *110*(20), 8194-8199. doi:10.1073/pnas.1303037110
- Rambaut, A., Holmes, E. C., O'Toole, A., Hill, V., McCrone, J. T., Ruis, C., du Plessis, L., & Pybus, O. G. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol*. doi:10.1038/s41564-020-0770-5
- Rasche, A., Lehmann, F., Konig, A., Goldmann, N., Corman, V. M., Moreira-Soto, A., Geipel, A., van Riel, D., Vakulenko, Y. A., Sander, A. L., Niekamp, H., Kepper, R., Schlegel, M., Akoua-Koffi, C., Souza, B., Sahr, F., Olayemi, A., Schulze, V., Petraityte-Burneikiene, R., Kazaks, A., Lowjaga, K., Geyer, J., Kuiken, T., Drosten, C., Lukashev, A. N., Fichet-Calvet, E., Ulrich, R. G., Glebe, D., & Drexler, J. F. (2019a). Highly diversified shrew hepatitis B viruses corroborate ancient origins and divergent infection patterns of mammalian hepadnaviruses. *Proc Natl Acad Sci U S A*, *116*(34), 17007-17012. doi:10.1073/pnas.1908072116
- Rasche, A., Sander, A. L., Corman, V. M., & Drexler, J. F. (2019b). Evolutionary biology of human hepatitis viruses. *J Hepatol*, *70*(3), 501-520. doi:10.1016/j.jhep.2018.11.010
- Redondo, N., Zaldivar-Lopez, S., Garrido, J. J., & Montoya, M. (2021). SARS-CoV-2 Accessory Proteins in Viral Pathogenesis: Knowns and Unknowns. *Front Immunol*, *12*, 708264. doi:10.3389/fimmu.2021.708264
- Rennert, L., Ma, Z., McMahan, C. S., & Dean, D. (2022). Effectiveness and protection duration of Covid-19 vaccines and previous infection against any SARS-CoV-2 infection in young adults. *Nat Commun*, *13*(1), 3946. doi:10.1038/s41467-022-31469-z
- Richard, M., Kok, A., de Meulder, D., Bestebroer, T. M., Lamers, M. M., Okba, N. M. A., Fentener van Vlissingen, M., Rockx, B., Haagmans, B. L., Koopmans, M. P. G., Fouchier, R. A. M., & Herfst, S. (2020). SARS-

- CoV-2 is transmitted via contact and via the air between ferrets. *Nat Commun*, 11(1), 3496. doi:10.1038/s41467-020-17367-2
- Rihtaric, D., Hostnik, P., Steyer, A., Grom, J., & Toplak, I. (2010). Identification of SARS-like coronaviruses in horseshoe bats (*Rhinolophus hipposideros*) in Slovenia. *Arch Virol*, 155(4), 507-514. doi:10.1007/s00705-010-0612-5
- Rima, B. K., & McFerran, N. V. (1997). Dinucleotide and stop codon frequencies in single-stranded RNA viruses. *J Gen Virol*, 78 (Pt 11), 2859-2870. doi:10.1099/0022-1317-78-11-2859
- Robertson, B. H. (2001). Viral hepatitis and primates: historical and molecular analysis of human and nonhuman primate hepatitis A, B, and the GB-related viruses. *J Viral Hepat*, 8(4), 233-242.
- Robertson, B. H., Jansen, R. W., Khanna, B., Totsuka, A., Nainan, O. V., Siegl, G., Widell, A., Margolis, H. S., Isomura, S., Ito, K., & et al. (1992). Genetic relatedness of hepatitis A virus strains recovered from different geographical regions. *J Gen Virol*, 73 (Pt 6), 1365-1377. doi:10.1099/0022-1317-73-6-1365
- Rohaim, M. A., El Naggari, R. F., Clayton, E., & Munir, M. (2021). Structural and functional insights into non-structural proteins of coronaviruses. *Microb Pathog*, 150, 104641. doi:10.1016/j.micpath.2020.104641
- Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19(12), 1572-1574.
- 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., Peel, A. J., Restif, O., Schountz, T., Parrish, C. R., Gurley, E. S., Lloyd-Smith, J. O., Hudson, P. J., Munster, V. J., & Plowright, R. K. (2022). Ecology, evolution and spillover of coronaviruses from bats. *Nat Rev Microbiol*, 20(5), 299-314. doi:10.1038/s41579-021-00652-2
- Russell, W. M. S., & Burch, R. L. (1959). *The Principles of Humane Experimental Technique*: Methuen.
- Sah, P., Fitzpatrick, M. C., Zimmer, C. F., Abdollahi, E., Juden-Kelly, L., Moghadas, S. M., Singer, B. H., & Galvani, A. P. (2021). Asymptomatic SARS-CoV-2 infection: A systematic review and meta-analysis. *Proc Natl Acad Sci U S A*, 118(34). doi:10.1073/pnas.2109229118
- Sander, A. L., Corman, V. M., Lukashev, A. N., & Drexler, J. F. (2018). Evolutionary Origins of Enteric Hepatitis Viruses. *Cold Spring Harb Perspect Med*, 8(12). doi:10.1101/cshperspect.a031690
- Sander, A. L., Moreira-Soto, A., Yordanov, S., Toplak, I., Balboni, A., Ameneiros, R. S., Corman, V., Drosten, C., & Drexler, J. F. (2022). Genomic determinants of Furin cleavage in diverse European SARS-related bat coronaviruses. *Commun Biol*, 5(1), 491. doi:10.1038/s42003-022-03421-w
- Sander, A. L., Yadouleton, A., de Oliveira Filho, E. F., Tchibozo, C., Hounkanrin, G., Badou, Y., Adewumi, P., Rene, K. K., Ange, D., Sourakatou, S., Sedjro, E., Aissi, M. A. J., Fidelia, H., Djingarey, M. H., Nagel, M., Jo, W. K., Moreira-Soto, A., Drosten, C., Landt, O., Corman, V. M., Hounkpatin, B., & Drexler, J. F. (2021a). Mutations Associated with SARS-CoV-2 Variants of Concern, Benin, Early 2021. *Emerg Infect Dis*, 27(11), 2889-2903. doi:10.3201/eid2711.211353
- Sander, A. L., Yadouleton, A., Moreira-Soto, A., Tchibozo, C., Hounkanrin, G., Badou, Y., Fischer, C., Krause, N., Akogbeto, P., E, F. d. O. F., Dossou, A., Brunink, S., Drosten, C., Aissi, M. A. J., Harouna Djingarey, M., Hounkpatin, B., Nagel, M., & Drexler, J. F. (2021b). An Observational Laboratory-Based Assessment of SARS-CoV-2 Molecular Diagnostics in Benin, Western Africa. *mSphere*, 6(1). doi:10.1128/mSphere.00979-20
- Sanjuan, R., Nebot, M. R., Chirico, N., Mansky, L. M., & Belshaw, R. (2010). Viral mutation rates. *J Virol*, 84(19), 9733-9748. doi:10.1128/JVI.00694-10
- Schroeder, S., Pott, F., Niemeyer, D., Veith, T., Richter, A., Muth, D., Goffinet, C., Muller, M. A., & Drosten, C. (2021). Interferon antagonism by SARS-CoV-2: a functional study using reverse genetics. *Lancet Microbe*, 2(5), e210-e218. doi:10.1016/S2666-5247(21)00027-6

- Schulte-Herbruggen, B., Cowlshaw, G., Homewood, K., & Rowcliffe, J. M. (2013). The importance of bushmeat in the livelihoods of West African cash-crop farmers living in a faunally-depleted landscape. *PLoS One*, *8*(8), e72807. doi:10.1371/journal.pone.0072807
- Seidah, N. G., Mayer, G., Zaid, A., Rousset, E., Nassoury, N., Poirier, S., Essalmani, R., & Prat, A. (2008). The activation and physiological functions of the proprotein convertases. *Int J Biochem Cell Biol*, *40*(6-7), 1111-1125. doi:10.1016/j.biocel.2008.01.030
- Shi, M., Lin, X. D., Chen, X., Tian, J. H., Chen, L. J., Li, K., Wang, W., Eden, J. S., Shen, J. J., Liu, L., Holmes, E. C., & Zhang, Y. Z. (2018). The evolutionary history of vertebrate RNA viruses. *Nature*, *556*(7700), 197-202. doi:10.1038/s41586-018-0012-7
- Shi, Z., & Hu, Z. (2008). A review of studies on animal reservoirs of the SARS coronavirus. *Virus Res*, *133*(1), 74-87. doi:10.1016/j.virusres.2007.03.012
- Shimodaira, H., & Hasegawa, M. (1999). Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. *Molecular Biology and Evolution*, *16*(8), 1114-1114. doi:10.1093/oxfordjournals.molbev.a026201
- Shin, E. C., & Jeong, S. H. (2018). Natural History, Clinical Manifestations, and Pathogenesis of Hepatitis A. *Cold Spring Harb Perspect Med*, *8*(9). doi:10.1101/cshperspect.a031708
- Shirasaki, T., Feng, H., Duyvesteyn, H. M. E., Fusco, W. G., McKnight, K. L., Xie, L., Boyce, M., Kumar, S., Barouch-Bentov, R., Gonzalez-Lopez, O., McNamara, R., Wang, L., Hertel-Wulff, A., Chen, X., Einav, S., Duncan, J. A., Kapustina, M., Fry, E. E., Stuart, D. I., & Lemon, S. M. (2022). Nonlytic cellular release of hepatitis A virus requires dual capsid recruitment of the ESCRT-associated Bro1 domain proteins HD-PTP and ALIX. *PLoS Pathog*, *18*(8), e1010543. doi:10.1371/journal.ppat.1010543
- Sila, T., Sunghan, J., Laochareonsuk, W., Surasombatpattana, S., Kongkamol, C., Ingviya, T., Siripaitoon, P., Kositpantawong, N., Kanchanasuwan, S., Hortiwakul, T., Charernmak, B., Nwabor, O. F., Silpapojakul, K., & Chusri, S. (2022). Suspected Cat-to-Human Transmission of SARS-CoV-2, Thailand, July-September 2021. *Emerg Infect Dis*, *28*(7), 1485-1488. doi:10.3201/eid2807.212605
- Simmonds, P. (2012). SSE: a nucleotide and amino acid sequence analysis platform. *BMC Res Notes*, *5*, 50. doi:10.1186/1756-0500-5-50
- Simmonds, P. (2020). Pervasive RNA Secondary Structure in the Genomes of SARS-CoV-2 and Other Coronaviruses. *MBio*, *11*(6). doi:10.1128/mBio.01661-20
- Song, H. D., Tu, C. C., Zhang, G. W., Wang, S. Y., Zheng, K., Lei, L. C., Chen, Q. X., Gao, Y. W., Zhou, H. Q., Xiang, H., Zheng, H. J., Chern, S. W., Cheng, F., Pan, C. M., Xuan, H., Chen, S. J., Luo, H. M., Zhou, D. H., Liu, Y. F., He, J. F., Qin, P. Z., Li, L. H., Ren, Y. Q., Liang, W. J., Yu, Y. D., Anderson, L., Wang, M., Xu, R. H., Wu, X. W., Zheng, H. Y., Chen, J. D., Liang, G., Gao, Y., Liao, M., Fang, L., Jiang, L. Y., Li, H., Chen, F., Di, B., He, L. J., Lin, J. Y., Tong, S., Kong, X., Du, L., Hao, P., Tang, H., Bernini, A., Yu, X. J., Spiga, O., Guo, Z. M., Pan, H. Y., He, W. Z., Manuguerra, J. C., Fontanet, A., Danchin, A., Niccolai, N., Li, Y. X., Wu, C. I., & Zhao, G. P. (2005). Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc Natl Acad Sci U S A*, *102*(7), 2430-2435. doi:10.1073/pnas.0409608102
- Starr, T. N., Zepeda, S. K., Walls, A. C., Greaney, A. J., Alkhovsky, S., Veelsler, D., & Bloom, J. D. (2022). ACE2 binding is an ancestral and evolvable trait of sarbecoviruses. *Nature*. doi:10.1038/s41586-022-04464-z
- Stout, A. E., Millet, J. K., Stanhope, M. J., & Whittaker, G. R. (2021). Furin cleavage sites in the spike proteins of bat and rodent coronaviruses: Implications for virus evolution and zoonotic transfer from rodent species. *One Health*, *13*, 100282. doi:10.1016/j.onehlt.2021.100282
- Suarez, D. L., Senne, D. A., Banks, J., Brown, I. H., Essen, S. C., Lee, C. W., Manvell, R. J., Mathieu-Benson, C., Moreno, V., Pedersen, J. C., Panigrahy, B., Rojas, H., Spackman, E., & Alexander, D. J. (2004). Recombination resulting in virulence shift in avian influenza outbreak, Chile. *Emerg Infect Dis*, *10*(4), 693-699. doi:10.3201/eid1004.030396

- Sundararaman, S. A., Plenderleith, L. J., Liu, W., Loy, D. E., Learn, G. H., Li, Y., Shaw, K. S., Ayoub, A., Peeters, M., Speede, S., Shaw, G. M., Bushman, F. D., Brisson, D., Rayner, J. C., Sharp, P. M., & Hahn, B. H. (2016). Genomes of cryptic chimpanzee *Plasmodium* species reveal key evolutionary events leading to human malaria. *Nat Commun*, 7, 11078. doi:10.1038/ncomms11078
- Swanepoel, R., Leman, P. A., Burt, F. J., Zachariades, N. A., Braack, L. E., Ksiazek, T. G., Rollin, P. E., Zaki, S. R., & Peters, C. J. (1996). Experimental inoculation of plants and animals with Ebola virus. *Emerg Infect Dis*, 2(4), 321-325. doi:10.3201/eid0204.960407
- Takata, M. A., Goncalves-Carneiro, D., Zang, T. M., Soll, S. J., York, A., Blanco-Melo, D., & Bieniasz, P. D. (2017). CG dinucleotide suppression enables antiviral defence targeting non-self RNA. *Nature*, 550(7674), 124-127. doi:10.1038/nature24039
- Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol*, 38(7), 3022-3027. doi:10.1093/molbev/msab120
- Tegally, H., San, J. E., Cotten, M., Moir, M., Tegomoh, B., Mboowa, G., Martin, D. P., Baxter, C., Lambisia, A. W., Diallo, A., Amoako, D. G., Diagne, M. M., Sisay, A., Zekri, A. N., Gueye, A. S., Sangare, A. K., Ouedraogo, A. S., Sow, A., Musa, A. O., Sesay, A. K., Abias, A. G., Elzagheid, A. I., Lagare, A., Kemi, A. S., Abar, A. E., Johnson, A. A., Fowotade, A., Oluwapelumi, A. O., Amuri, A. A., Juru, A., Kandeil, A., Mostafa, A., Rebai, A., Sayed, A., Kazeem, A., Balde, A., Christoffels, A., Trotter, A. J., Campbell, A., Keita, A. K., Kone, A., Bouzid, A., Souissi, A., Agweyu, A., Naguib, A., Gutierrez, A. V., Nkeshimana, A., Page, A. J., Yadouleton, A., Vinze, A., Happi, A. N., Chouikha, A., Iranzadeh, A., Maharaj, A., Batchi-Bouyou, A. L., Ismail, A., Sylverken, A. A., Goba, A., Femi, A., Sijuwola, A. E., Marycelin, B., Salako, B. L., Oderinde, B. S., Bolajoko, B., Diarra, B., Herring, B. L., Tsofa, B., Lekana-Douki, B., Mvula, B., Njanpop-Lafourcade, B. M., Marondera, B. T., Khaireh, B. A., Kouriba, B., Adu, B., Pool, B., McInnis, B., Brook, C., Williamson, C., Nduwimana, C., Ancombe, C., Pratt, C. B., Scheepers, C., Akoua-Koffi, C. G., Agoti, C. N., Mapanguy, C. M., Loucoubar, C., Onwuamah, C. K., Ihekweazu, C., Malaka, C. N., Peyrefitte, C., Grace, C., Omoruyi, C. E., Rafai, C. D., Morang'a, C. M., Eramah, C., Lule, D. B., Bridges, D. J., Mukadi-Bamuleka, D., Park, D., Rasmussen, D. A., Baker, D., Nokes, D. J., Ssemwanga, D., Tshiabuila, D., Amuzu, D. S. Y., Goedhals, D., Grant, D. S., Omuoyo, D. O., Maruapula, D., Wanjohi, D. W., Foster-Nyarko, E., Lusamaki, E. K., Simulundu, E., Ong'era, E. M., Ngabana, E. N., Abworo, E. O., Otieno, E., Shumba, E., Barasa, E., Ahmed, E. B., Ahmed, E. A., Lokilo, E., Mukantwari, E., Philomena, E., Belarbi, E., Simon-Loriere, E., Anoh, E. A., Manuel, E., Leendertz, F., Taweh, F. M., Wasfi, F., Abdelmoula, F., Takawira, F. T., Derrar, F., Ajogbasile, F. V., Treurnicht, F., Onikepe, F., Ntouni, F., Muyembe, F. M., Ragomzingba, F. E. Z., Dratibi, F. A., Iyanu, F. A., Mbunsu, G. K., Thilliez, G., Kay, G. L., Akpede, G. O., van Zyl, G. U., Awandare, G. A., Kpeli, G. S., Schubert, G., Maphalala, G. P., Ranaivoson, H. C., Omunakwe, H. E., Onywera, H., Abe, H., Karray, H., Nansumba, H., Triki, H., Kadjo, H. A. A., Elgahzaly, H., Gumbo, H., Mathieu, H., Kavunga-Membo, H., Smeti, I., Olawoye, I. B., Adetifa, I. M. O., Odia, I., Ben Boubaker, I. B., Mohammad, I. A., Ssewanyana, I., Wurie, I., Konstantinus, I. S., Halatoko, J. W. A., Ayei, J., Sonoo, J., Makangara, J. C., Tamfum, J. M., Heraud, J. M., Shaffer, J. G., Giandhari, J., Musyoki, J., Nkurunziza, J., Uwanibe, J. N., Bhiman, J. N., Yasuda, J., Morais, J., Kiconco, J., Sandi, J. D., Huddleston, J., Odoom, J. K., Morobe, J. M., Gyapong, J. O., Kayiwa, J. T., Okolie, J. C., Xavier, J. S., Gyamfi, J., Wamala, J. F., Bonney, J. H. K., Nyandwi, J., Everatt, J., Nakaseegu, J., Ngoi, J. M., Namulondo, J., Oguzie, J. U., Andeko, J. C., Lutwama, J. J., Mogga, J. J. H., O'Grady, J., Siddle, K. J., Victoir, K., Adeyemi, K. T., Tumedi, K. A., Carvalho, K. S., Mohammed, K. S., Dellagi, K., Musonda, K. G., Duedu, K. O., Fki-Berrajah, L., Singh, L., Kepler, L. M., Biscornet, L., de Oliveira Martins, L., Chabuka, L., Olubayo, L., Ojok, L. D., Deng, L. L., Ochola-Oyier, L. I., Tyers, L., Mine, M., Ramuth, M., Mastouri, M., ElHefnawi, M., Mbanne, M., Matsheka, M. I., Kebabonye, M., Diop, M., Momoh, M., Lima Mendonca, M. D. L., Venter, M., Paye, M. F., Faye, M., Nyaga, M. M., Mareka, M., Damaris, M. M., Mburu, M. W., Mpina, M. G., Owusu, M., Wiley, M. R., Tatteng, M. Y., Ayekaba, M. O., Abouelhoda, M., Beloufa, M. A., Seadawy, M. G., Khalifa, M. K., Matobo, M. M., Kane, M., Salou, M., Mbulawa, M. B., Mwenda, M., Allam, M., Phan, M. V. T., Abid, N., Rujeni, N., Abuzaid, N., Ismael, N., Elguindy, N., Top, N. M., Dia, N., Mabunda, N., Hsiao, N. Y., Silochi, N. B., Francisco, N. M., Saasa, N., Bbosa, N., Murunga, N., Gumede, N., Wolter, N., Sitharam, N., Ndodo, N., Ajayi, N. A., Tordo, N., Mbhele, N., Razanajatovo, N. H., Iguosadolo, N., Mba, N., Kingsley, O. C., Sylvanus, O., Femi, O., Adewumi, O. M., Testimony, O., Ogunsanya, O. A., Fakayode, O., Ogah, O. E., Oludayo, O. E., Faye, O., Smith-Lawrence, P., Ondoa, P., Combe, P., Nabisubi, P., Semanda, P., Oluniyi, P. E., Arnaldo, P., Quashie, P. K., Okokhere, P. O., Bejon, P., Dussart, P., Bester, P. A., Mbala, P. K., Kaleebu, P., Abechi, P., El-Shesheny, R., Joseph, R., Aziz, R. K., Essomba, R. G., Ayivor-Djanie, R., Njouom, R., Phillips, R. O., Gorman, R., Kingsley, R. A., Neto Rodrigues, R.,

- Audu, R. A., Carr, R. A. A., Gargouri, S., Masmoudi, S., Bootsma, S., Sankhe, S., Mohamed, S. I., Femi, S., Mhalla, S., Hosch, S., Kassim, S. K., Metha, S., Trabelsi, S., Agwa, S. H., Mwangi, S. W., Doumbia, S., Makiala-Mandanda, S., Aryeetey, S., Ahmed, S. S., Ahmed, S. M., Elhamoumi, S., Moyo, S., Lutucuta, S., Gaseitsiwe, S., Jalloh, S., Andriamandimby, S. F., Oguntope, S., Grayo, S., Lekana-Douki, S., Prosolek, S., Ouangraoua, S., van Wyk, S., Schaffner, S. F., Kanyerezi, S., Ahuka-Mundeki, S., Rudder, S., Pillay, S., Nabadda, S., Behillil, S., Budiaki, S. L., van der Werf, S., Mashe, T., Mohale, T., Le-Viet, T., Velavan, T. P., Schindler, T., Maponga, T. G., Bedford, T., Anyaneji, U. J., Chinedu, U., Ramphal, U., George, U. E., Enouf, V., Nene, V., Gorova, V., Roshdy, W. H., Karim, W. A., Ampofo, W. K., Preiser, W., Choga, W. T., Ahmed, Y. A., Ramphal, Y., Bediako, Y., Naidoo, Y., Butera, Y., de Laurent, Z. R., Africa Pathogen Genomics, I., Ouma, A. E. O., von Gottberg, A., Githinji, G., Moeti, M., Tomori, O., Sabeti, P. C., Sall, A. A., Oyola, S. O., Tebeje, Y. K., Tessema, S. K., de Oliveira, T., Happi, C., Lessells, R., Nkengasong, J., & Wilkinson, E. (2022). The evolving SARS-CoV-2 epidemic in Africa: Insights from rapidly expanding genomic surveillance. *Science*, eabq5358. doi:10.1126/science.abq5358
- Temmam, S., Vongphayloth, K., Baquero, E., Munier, S., Bonomi, M., Regnault, B., Douangboubpha, B., Karami, Y., Chretien, D., Sanamxay, D., Xayaphet, V., Paphaphanh, P., Lacoste, V., Somlor, S., Lakeomany, K., Phommavanh, N., Perot, P., Dehan, O., Amara, F., Donati, F., Bigot, T., Nilges, M., Rey, F. A., van der Werf, S., Brey, P. T., & Eloit, M. (2022a). Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature*, 604(7905), 330-336. doi:10.1038/s41586-022-04532-4
- Temmam, S., Vongphayloth, K., Salazar, E.-B., Munier, S., Bonomi, M., Régnault, B., Douangboubpha, B., Karami, Y., Chretien, D., Sanamxay, D., Xayaphet, V., Paphaphanh, P., Lacoste, V., Somlor, S., Lakeomany, K., Phommavanh, N., Pérot, P., Donati, F., Bigot, T., Nilges, M., Rey, F. A., van der Werf, S., Brey, P., & Eloit, M. (2022b). Coronaviruses with a SARS-CoV-2-like receptor-binding domain allowing ACE2-mediated entry into human cells isolated from bats of Indochinese peninsula. *Research Square*. doi:10.21203/rs.3.rs-871965/v1
- Thomas, G. (2002). Furin at the cutting edge: from protein traffic to embryogenesis and disease. *Nat Rev Mol Cell Biol*, 3(10), 753-766. doi:10.1038/nrm934
- Tian, S., Huajun, W., & Wu, J. (2012). Computational prediction of furin cleavage sites by a hybrid method and understanding mechanism underlying diseases. *Sci Rep*, 2, 261. doi:10.1038/srep00261
- Towner, J. S., Amman, B. R., Sealy, T. K., Carroll, S. A., Comer, J. A., Kemp, A., Swanepoel, R., Paddock, C. D., Balinandi, S., Khristova, M. L., Formenty, P. B., Albarino, C. G., Miller, D. M., Reed, Z. D., Kayiwa, J. T., Mills, J. N., Cannon, D. L., Greer, P. W., Byaruhanga, E., Farnon, E. C., Atimnedi, P., Okware, S., Katongole-Mbidde, E., Downing, R., Tappero, J. W., Zaki, S. R., Ksiazek, T. G., Nichol, S. T., & Rollin, P. E. (2009). Isolation of genetically diverse Marburg viruses from Egyptian fruit bats. *PLoS Pathog*, 5(7), e1000536. doi:10.1371/journal.ppat.1000536
- Tsarev, S. A., Emerson, S. U., Balayan, M. S., Ticehurst, J., & Purcell, R. H. (1991). Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome structure and growth in cell culture with other HAV strains. *J Gen Virol*, 72 (Pt 7), 1677-1683. doi:10.1099/0022-1317-72-7-1677
- Vakulenko, Y., Deviatkin, A., Drexler, J. F., & Lukashev, A. (2021a). Modular Evolution of Coronavirus Genomes. *Viruses*, 13(7). doi:10.3390/v13071270
- Vakulenko, Y., Deviatkin, A., Drexler, J. F., & Lukashev, A. (2021b). Modular Evolution of Coronavirus Genomes. *Viruses*, 13(7), 1270.
- Van Valen, L. (1973). A new evolutionary law. *Evolutionary theory*, 1, 1--30.
- Vijaykrishna, D., Mukerji, R., & Smith, G. J. (2015). RNA Virus Reassortment: An Evolutionary Mechanism for Host Jumps and Immune Evasion. *PLoS Pathog*, 11(7), e1004902. doi:10.1371/journal.ppat.1004902
- Voloch, C. M., da Silva Francisco, R., Jr., de Almeida, L. G. P., Cardoso, C. C., Brustolini, O. J., Gerber, A. L., Guimaraes, A. P. C., Mariani, D., da Costa, R. M., Ferreira, O. C., Jr., Covid19-Ufrj Workgroup, L. W. A. C. C., Frauches, T. S., de Mello, C. M. B., Leitao, I. C., Galliez, R. M., Faffe, D. S., Castineiras, T., Tanuri, A., & de Vasconcelos, A. T. R. (2021). Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil. *J Virol*. doi:10.1128/JVI.00119-21

- Wacharapluesadee, S., Tan, C. W., Maneeorn, P., Duengkae, P., Zhu, F., Joyjinda, Y., Kaewpom, T., Chia, W. N., Ampoot, W., Lim, B. L., Worachotsueptrakun, K., Chen, V. C., Sirichan, N., Ruchisrisarod, C., Rodpan, A., Noradechanon, K., Phaichana, T., Jantararat, N., Thongnumchaima, B., Tu, C., Cramer, G., Stokes, M. M., Hemachudha, T., & Wang, L. F. (2021). Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. *Nat Commun*, *12*(1), 972. doi:10.1038/s41467-021-21240-1
- Wang, P., Nair, M. S., Liu, L., Iketani, S., Luo, Y., Guo, Y., Wang, M., Yu, J., Zhang, B., Kwong, P. D., Graham, B. S., Mascola, J. R., Chang, J. Y., Yin, M. T., Sobieszczyk, M., Kyratsous, C. A., Shapiro, L., Sheng, Z., Huang, Y., & Ho, D. D. (2021). Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature*. doi:10.1038/s41586-021-03398-2
- Wang, S., Xu, X., Wei, C., Li, S., Zhao, J., Zheng, Y., Liu, X., Zeng, X., Yuan, W., & Peng, S. (2022). Molecular evolutionary characteristics of SARS-CoV-2 emerging in the United States. *J Med Virol*, *94*(1), 310-317. doi:10.1002/jmv.27331
- Watanabe, S., Masangkay, J. S., Nagata, N., Morikawa, S., Mizutani, T., Fukushima, S., Alviola, P., Omatsu, T., Ueda, N., Iha, K., Taniguchi, S., Fujii, H., Tsuda, S., Endoh, M., Kato, K., Tohya, Y., Kyuwa, S., Yoshikawa, Y., & Akashi, H. (2010). Bat coronaviruses and experimental infection of bats, the Philippines. *Emerg Infect Dis*, *16*(8), 1217-1223. doi:10.3201/eid1608.100208
- Wells, H. L., Letko, M., Lasso, G., Ssebidde, B., Nziza, J., Byarugaba, D. K., Navarrete-Macias, I., Liang, E., Cranfield, M., Han, B. A., Tingley, M. W., Diuk-Wasser, M., Goldstein, T., Johnson, C. K., Mazet, J. A. K., Chandran, K., Munster, V. J., Gilardi, K., & Anthony, S. J. (2021). The evolutionary history of ACE2 usage within the coronavirus subgenus Sarbecovirus. *Virus Evol*, *7*(1), veab007. doi:10.1093/ve/veab007
- WHO. (2017). *Global hepatitis report, 2017*. <https://www.who.int/publications/i/item/9789241565455>: ISBN 978-92-4-156545-5
- WHO. (2021a). *COVID-19 weekly epidemiological update*. 2 May 2021. Retrieved from <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---4-may-2021>,
- WHO. (2021b). Six in seven COVID-19 infections go undetected in Africa. Retrieved from <https://www.afro.who.int/news/six-seven-covid-19-infections-go-undetected-africa>
- WHO. (2022a). *COVID-19 weekly Epidemiological Update*. 25 January 2022. Retrieved from <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---25-january-2022>,
- WHO. (2022b). *Global genomic surveillance strategy for pathogens with pandemic and epidemic potential, 2022–2032*. Geneva.
- WHO. (2022c). WHO releases 10-year strategy for genomic surveillance of pathogens. Retrieved from <https://www.who.int/news/item/30-03-2022-who-releases-10-year-strategy-for-genomic-surveillance-of-pathogens>
- Wilkinson, E., Giovanetti, M., Tegally, H., San, J. E., Lessells, R., Cuadros, D., Martin, D. P., Rasmussen, D. A., Zekri, A. N., Sangare, A. K., Ouedraogo, A. S., Sesay, A. K., Priscilla, A., Kemi, A. S., Olubusuyi, A. M., Oluwapelumi, A. O. O., Hammami, A., Amuri, A. A., Sayed, A., Ouma, A. E. O., Elargoubi, A., Ajayi, N. A., Victoria, A. F., Kazeem, A., George, A., Trotter, A. J., Yahaya, A. A., Keita, A. K., Diallo, A., Kone, A., Souissi, A., Chtourou, A., Gutierrez, A. V., Page, A. J., Vinze, A., Iranzadeh, A., Lambisia, A., Ismail, A., Rosemary, A., Sylverken, A., Femi, A., Ibrahimi, A., Marycelin, B., Oderinde, B. S., Bolajoko, B., Dhaala, B., Herring, B. L., Njanpop-Lafourcade, B. M., Kleinhans, B., McInnis, B., Tegomoh, B., Brook, C., Pratt, C. B., Scheepers, C., Akoua-Koffi, C. G., Agoti, C. N., Peyrefitte, C., Daubenberger, C., Morang'a, C. M., Nokes, D. J., Amoako, D. G., Bugembe, D. L., Park, D., Baker, D., Doolabh, D., Ssemwanga, D., Tshiabuila, D., Bassirou, D., Amuzu, D. S. Y., Goedhals, D., Omuoyo, D. O., Maruapula, D., Foster-Nyarko, E., Lusamaki, E. K., Simulundu, E., Ong'era, E. M., Ngabana, E. N., Shumba, E., El Fahime, E., Lokilo, E., Mukantwari, E., Philomena, E., Belarbi, E., Simon-Loriere, E., Anoh, E. A., Leendertz, F., Ajili, F., Enoch, F. O., Wasfi, F., Abdelmoula, F., Mosha, F. S., Takawira, F. T., Derrar, F., Bouzid, F., Onikepe, F., Adeola, F., Muyembe, F. M., Tanser, F., Dratibi, F. A., Mbunzu, G. K., Thilliez, G., Kay, G. L., Githinji, G., van Zyl, G., Awandare, G. A., Schubert, G., Maphalala, G. P., Ranaivoson, H. C., Lemriss, H., Anise, H., Abe, H., Karray, H. H., Nansumba, H., Elgahzaly, H. A.,

- Gumbo, H., Smeti, I., Ayed, I. B., Odia, I., Ben Boubaker, I. B., Gaaloul, I., Gazy, I., Mudau, I., Ssewanyana, I., Konstantinus, I., Lekana-Douk, J. B., Makangara, J. C., Tamfum, J. M., Heraud, J. M., Shaffer, J. G., Giandhari, J., Li, J., Yasuda, J., Mends, J. Q., Kiconco, J., Morobe, J. M., Gyapong, J. O., Okolie, J. C., Kayiwa, J. T., Edwards, J. A., Gyamfi, J., Farah, J., Nakasegu, J., Ngoi, J. M., Namulondo, J., Andeko, J. C., Lutwama, J. J., O'Grady, J., Siddle, K., Adeyemi, K. T., Tumedi, K. A., Said, K. M., Hae-Young, K., Duedu, K. O., Belyamani, L., Fki-Berrajah, L., Singh, L., Martins, L. O., Tyers, L., Ramuth, M., Mastouri, M., Aouni, M., El Hefnawi, M., Matsheka, M. I., Kebabonye, M., Diop, M., Turki, M., Paye, M., Nyaga, M. M., Mareka, M., Damaris, M. M., Mburu, M. W., Mpina, M., Nwando, M., Owusu, M., Wiley, M. R., Youtchou, M. T., Ayekaba, M. O., Abouelhoda, M., Seadawy, M. G., Khalifa, M. K., Sekhele, M., Ouadghiri, M., Diagne, M. M., Mwenda, M., Allam, M., Phan, M. V. T., Abid, N., Touil, N., Rujeni, N., Kharrat, N., Ismael, N., Dia, N., Mabunda, N., Hsiao, N. Y., Silochi, N. B., Nsenga, N., Gumede, N., Mulder, N., Ndodo, N., Razanajatovo, N. H., Iguosadolo, N., Judith, O., Kingsley, O. C., Sylvanus, O., Peter, O., Femi, O., Idowu, O., Testimony, O., Chukwuma, O. E., Ogah, O. E., Onwuamah, C. K., Cyril, O., Faye, O., Tomori, O., Ondo, P., Combe, P., Semanda, P., Oluniyi, P. E., Arnaldo, P., Quashie, P. K., Dussart, P., Bester, P. A., Mbala, P. K., Ayivor-Djanie, R., Njouom, R., Phillips, R. O., Gorman, R., Kingsley, R. A., Carr, R. A. A., El Kabbaj, S., Gargouri, S., Masmoudi, S., Sankhe, S., Lawal, S. B., Kassim, S., Trabelsi, S., Metha, S., Kammoun, S., Lemriss, S., Agwa, S. H. A., Calvignac-Spencer, S., Schaffner, S. F., Doumbia, S., Mandanda, S. M., Aryeetey, S., Ahmed, S. S., Elhamoumi, S., Andriamandimby, S., Tope, S., Lekana-Douki, S., Prosolek, S., Ouangraoua, S., Mundeke, S. A., Rudder, S., Panji, S., Pillay, S., Engelbrecht, S., Nabadda, S., Behillil, S., Budiaki, S. L., van der Werf, S., Mashe, T., Aanniz, T., Mohale, T., Le-Viet, T., Schindler, T., Anyaneji, U. J., Chinedu, U., Ramphal, U., Jessica, U., George, U., Fonseca, V., Enouf, V., Gorova, V., Roshdy, W. H., Ampofo, W. K., Preiser, W., Choga, W. T., Bediako, Y., Naidoo, Y., Butera, Y., de Laurent, Z. R., Sall, A. A., Rebai, A., von Gottberg, A., Kouriba, B., Williamson, C., Bridges, D. J., Chikwe, I., Bhiman, J. N., Mine, M., Cotten, M., Moyo, S., Gaseitsiwe, S., Saasa, N., Sabeti, P. C., Kaleebu, P., Tebeje, Y. K., Tessema, S. K., Happi, C., Nkengasong, J., & de Oliveira, T. (2021). A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa. *Science*, *374*(6566), 423-431. doi:10.1126/science.abj4336
- Woolhouse, M. E., & Gowtage-Sequeria, S. (2005). Host range and emerging and reemerging pathogens. *Emerg Infect Dis*, *11*(12), 1842-1847. doi:10.3201/eid1112.050997
- Worobey, M., Levy, J. I., Serrano, L. M., Crits-Christoph, A., Pekar, J. E., Goldstein, S. A., Rasmussen, A. L., Kraemer, M. U. G., Newman, C., Koopmans, M. P. G., Suchard, M. A., Wertheim, J. O., Lemey, P., Robertson, D. L., Garry, R. F., Holmes, E. C., Rambaut, A., & Andersen, K. G. (2022). The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic. *Science*, *abp8715*. doi:10.1126/science.abp8715
- Wright, F. (1990). The 'effective number of codons' used in a gene. *Gene*, *87*(1), 23-29.
- Wu, Y., & Zhao, S. (2020). Furin cleavage sites naturally occur in coronaviruses. *Stem Cell Res*, *50*, 102115. doi:10.1016/j.scr.2020.102115
- Xiao, K., Zhai, J., Feng, Y., Zhou, N., Zhang, X., Zou, J. J., Li, N., Guo, Y., Li, X., Shen, X., Zhang, Z., Shu, F., Huang, W., Li, Y., Zhang, Z., Chen, R. A., Wu, Y. J., Peng, S. M., Huang, M., Xie, W. J., Cai, Q. H., Hou, F. H., Chen, W., Xiao, L., & Shen, Y. (2020). Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature*, *583*(7815), 286-289. doi:10.1038/s41586-020-2313-x
- Xiao, X., Newman, C., Buesching, C. D., Macdonald, D. W., & Zhou, Z. M. (2021). Animal sales from Wuhan wet markets immediately prior to the COVID-19 pandemic. *Sci Rep*, *11*(1), 11898. doi:10.1038/s41598-021-91470-2
- Xiong, Q., Cao, L., Ma, C., Liu, C., Si, J., Liu, P., Gu, M., Wang, C., Shi, L., Tong, F., Huang, M., Li, J., Zhao, C., Shen, C., Chen, Y., Zhao, H., Lan, K., Wang, X., & Yan, H. (2022). Close relatives of MERS-CoV in bats use ACE2 as their functional receptors. *bioRxiv*, 2022.2001.2024.477490. doi:10.1101/2022.01.24.477490
- Xu, B., & Yang, Z. (2013). PAMLX: a graphical user interface for PAML. *Mol Biol Evol*, *30*(12), 2723-2724. doi:10.1093/molbev/mst179

- Yadouleton, A., Sander, A. L., Adewumi, P., de Oliveira Filho, E. F., Tchiboza, C., Hounkanrin, G., Rene, K. K., Ange, D., Kohoun, R. K., Nari, R. C., Salifou, S., Saizonou, R., Kakai, C. G., Bedie, S. V., Al Onifade, F., Nagel, M., Aissi, M. A. J., Akogbeto, P., Drosten, C., Wulf, B., Moreira-Soto, A., Djingarey, M. H., Hounkpatin, B., & Drexler, J. F. (2021). Emergence of SARS-CoV-2 Delta Variant, Benin, May-July 2021. *Emerg Infect Dis*, 28(1). doi:10.3201/eid2801.211909
- Yang, Y., Liang, Y., Qu, L., Chen, Z., Yi, M., Li, K., & Lemon, S. M. (2007). Disruption of innate immunity due to mitochondrial targeting of a picornaviral protease precursor. *Proc Natl Acad Sci U S A*, 104(17), 7253-7258. doi:10.1073/pnas.0611506104
- Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*, 13(5), 555-556. doi:10.1093/bioinformatics/13.5.555
- Yen, H. L., Sit, T. H. C., Brackman, C. J., Chuk, S. S. Y., Gu, H., Tam, K. W. S., Law, P. Y. T., Leung, G. M., Peiris, M., Poon, L. L. M., & team, H.-S. s. (2022). Transmission of SARS-CoV-2 delta variant (AY.127) from pet hamsters to humans, leading to onward human-to-human transmission: a case study. *Lancet*, 399(10329), 1070-1078. doi:10.1016/S0140-6736(22)00326-9
- Yin, C. (2020). Genotyping coronavirus SARS-CoV-2: methods and implications. *Genomics*, 112(5), 3588-3596. doi:10.1016/j.ygeno.2020.04.016
- Yu, J. M., Li, L. L., Xie, G. C., Zhang, C. Y., Ao, Y. Y., & Duan, Z. J. (2018). Experimental infection of *Marmota monax* with a novel hepatitis A virus. *Arch Virol*. doi:10.1007/s00705-018-3715-z
- Yu, J. M., Li, L. L., Zhang, C. Y., Lu, S., Ao, Y. Y., Gao, H. C., Xie, Z. P., Xie, G. C., Sun, X. M., Pang, L. L., Xu, J. G., Lipkin, W. I., & Duan, Z. J. (2016). A novel hepatovirus identified in wild woodchuck *Marmota himalayana*. *Sci Rep*, 6, 22361. doi:10.1038/srep22361
- Zell, R., Delwart, E., Gorbalenya, A. E., Hovi, T., King, A. M. Q., Knowles, N. J., Lindberg, A. M., Pallansch, M. A., Palmenberg, A. C., Reuter, G., Simmonds, P., Skern, T., Stanway, G., Yamashita, T., & Ictv Report, C. (2017). ICTV Virus Taxonomy Profile: Picornaviridae. *J Gen Virol*, 98(10), 2421-2422. doi:10.1099/jgv.0.000911
- Zhang, L., Jackson, C. B., Mou, H., Ojha, A., Peng, H., Quinlan, B. D., Rangarajan, E. S., Pan, A., Vanderheiden, A., Suthar, M. S., Li, W., Izard, T., Rader, C., Farzan, M., & Choe, H. (2020). SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nat Commun*, 11(1), 6013. doi:10.1038/s41467-020-19808-4
- Zhang, Y., Liu, Y., Liu, W., Zhou, J., Chen, H., Wang, Y., Ma, L., Ding, Y., & Zhang, J. (2011). Analysis of synonymous codon usage in hepatitis A virus. *Virol J*, 8, 174. doi:10.1186/1743-422X-8-174
- Zhang, Y., Zhang, L., Wu, J., Yu, Y., Liu, S., Li, T., Li, Q., Ding, R., Wang, H., Nie, J., Cui, Z., Wang, Y., Huang, W., & Wang, Y. (2022). A second functional furin site in the SARS-CoV-2 spike protein. *Emerg Microbes Infect*, 11(1), 182-194. doi:10.1080/22221751.2021.2014284
- Zhao, Y., Zheng, H., Xu, A., Yan, D., Jiang, Z., Qi, Q., & Sun, J. (2016). Analysis of codon usage bias of envelope glycoprotein genes in nuclear polyhedrosis virus (NPV) and its relation to evolution. *BMC Genomics*, 17, 677. doi:10.1186/s12864-016-3021-7
- Zhou, H., Chen, X., Hu, T., Li, J., Song, H., Liu, Y., Wang, P., Liu, D., Yang, J., Holmes, E. C., Hughes, A. C., Bi, Y., & Shi, W. (2020a). A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein. *Curr Biol*, 30(11), 2196-2203 e2193. doi:10.1016/j.cub.2020.05.023
- Zhou, H., Ji, J., Chen, X., Bi, Y., Li, J., Wang, Q., Hu, T., Song, H., Zhao, R., Chen, Y., Cui, M., Zhang, Y., Hughes, A. C., Holmes, E. C., & Shi, W. (2021). Identification of novel bat coronaviruses sheds light on the evolutionary origins of SARS-CoV-2 and related viruses. *Cell*, 184(17), 4380-4391 e4314. doi:10.1016/j.cell.2021.06.008
- Zhou, J., Li, C., Liu, X., Chiu, M. C., Zhao, X., Wang, D., Wei, Y., Lee, A., Zhang, A. J., Chu, H., Cai, J. P., Yip, C. C., Chan, I. H., Wong, K. K., Tsang, O. T., Chan, K. H., Chan, J. F., To, K. K., Chen, H., & Yuen, K.

- Y. (2020b). Infection of bat and human intestinal organoids by SARS-CoV-2. *Nat Med*, 26(7), 1077-1083. doi:10.1038/s41591-020-0912-6
- Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., Si, H. R., Zhu, Y., Li, B., Huang, C. L., Chen, H. D., Chen, J., Luo, Y., Guo, H., Jiang, R. D., Liu, M. Q., Chen, Y., Shen, X. R., Wang, X., Zheng, X. S., Zhao, K., Chen, Q. J., Deng, F., Liu, L. L., Yan, B., Zhan, F. X., Wang, Y. Y., Xiao, G. F., & Shi, Z. L. (2020c). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798), 270-273. doi:10.1038/s41586-020-2012-7
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, 31(13), 3406-3415. doi:10.1093/nar/gkg595

