CHARLES UNIVERSITY

FACULTY OF SCIENCE

Study program: Biology

Branch of study: Biology (B-BI)

ŠIMON KARBAN

The evolution of SARS-CoV-2 during the pandemic of COVID-19

Evoluce viru SARS-CoV-2 během pandemie COVID-19

BACHELOR'S THESIS

Supervisor: **RNDr. Ruth Tachezy, PhD.**

Consultant: **RNDr. Jaroslav Nunvář, PhD.**

Prague, 2023

Declaration / Prohlášení

I declare that I wrote this thesis by myself and listed all used resources and bibliography. Neither this thesis nor its important part was previously used to gain any academic degree.

Prohlašuji, že tuto práci jsem vypracoval samostatně a uvedl jsem všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla použita k získání žádného akademického titulu.

V Praze, 2023

Signature / Podpis

Acknowledgement

My heartfelt gratitude goes to my friends and family for the mental support I needed while writing. I would also like to thank the team in my laboratory for their patience and understanding of my complicated time management during the work on this thesis.

Also, I am grateful to my supervisor, RNDr. Ruth Tachezy, PhD., and consultant, RNDr. Jaroslav Nunvář, PhD., for their ongoing mentorship and endless ideas and suggestions that pushed me to my limits and led to this thesis's current form.

Abstract:

Coronaviruses are animal and human viruses which, in the case of humans, cause respiratory diseases. The genome of coronaviruses is non-segmented and encodes several structural and several non-structural proteins. As their genome consists of single-stranded RNA in a positive sense, they encode RNA-dependent RNA polymerase. The origin of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is uncertain and may never be known. However, this thesis covers up-to-date knowledge and the arguments for the main theories on the emergence of this virus. The pandemic of disease COVID-19 caused by the SARS-CoV-2 had an enormous impact on the health and lives of people worldwide. The length and severity of the pandemic were caused by the characteristic of the virus, transmissibility and asymptomatic type of infection with severe symptoms in elderly and chronically ill individuals and the fast evolution of the virus after its appearance in humans. This thesis will describe important characteristics of the most important variants of the virus and changes which gave them a selection advantage. In the end, trends in the evolution of SARS-CoV-2 will be discussed.

Keywords: SARS-CoV-2, *Coronaviridae*, variants, phylogeny, origin, evolution

Abstrakt

Koronaviry jsou lidské a zvířecí viry, které u lidí způsobují respirační onemocnění. Jejich genom je nesegmentovaný a kóduje několik strukturních a nestrukturních proteinů. Vzhledem k tomu, že je genom koronavirů tvořen jednovlákennou RNA v pozitivním smyslu, kódují si vlastní RNA dependentní RNA polymerázu. Existuje několik hypotéz o původu viru SARS-CoV-2, žádná však zatím není definitivní. Následná pandemie nemoci COVID-19 způsobena virem SARS-CoV-2, měla zásadní efekt na zdraví a život lidí po celém světě. Doba trvání a závažnost této pandemie byly ovlivněny několika vlastnostmi viru SARS-CoV-2 – přenosnost, možnost asymptomatického průběhu, vážné symptomatické onemocnění u starších a chronicky nemocných pacientů a rychlá evoluce viru po jeho vstupu do lidské populace. Tato práce shrnuje dosavadní a aktuální poznání včetně argumentů pro předpokládané teorie vzniku tohoto viru, hlavní charakteristiky a varianty viru SARS-CoV-2 včetně selekčních výhod, které představily. Závěrem jsou diskutovány trendy v evoluci viru SARS-CoV-2.

Klíčová slova: SARS-CoV-2, koronaviry, varianty, fylogeneze, původ, evoluce

TABLE OF CONTENTS

LIST OF ABBREVIATIONS USED IN THE THESIS

+ssRNA = positive sense single-stranded ribonucleic acid

3CL^{pro} = chymotrypsin-like cysteine proteinase

A = adenine

(m)Ab(s) = (monoclonal) antibody(-ies)

(h)ACE2 = (human) angiotensin-converting enzyme 2

ADRP = ADP-ribose-1'-phosphatase

(h)APN = (human) aminopeptidase N

C = cytosine

CH = calponin homology

CoV(s) = coronavirus(es)

COVID-19 = Coronavirus disease 2019

CTD = C-terminal domain

CtD = cytoplasmic tail domain

DENV = Dengue virus

DMV = double-membrane vesicles

DPP4 = dipeptidyl peptidase 4

E = envelope protein

FBI = Federal Bureau of Investigation

FCS = furin cleavage site

G = guanine

HCoV = human coronavirus

HE = hemagglutinin-esterase

HeV = Hendra virus

HGT = horizontal gene transfer

HIV = human immunodeficiency virus

HSM = Huanan Seafood market

HSPGs = heparan sulphate proteoglycans

IFN = interferon

IgG = Immunoglobulin G

IIR = innate immune response

M = transmembrane protein

MDA5 = melanoma differentiation-associated protein 5

MERS-CoV = Middle-east respiratory syndrome coronavirus

M^{pro} = main proteinase

mRNA = messenger RNA

mtDNA = mitochondrial DNA

N = nucleocapsid protein

N7-MTase = N7-methyltransferase

NendoU = Nidoviral RNA uridylate-specific endoribonuclease

NiV = Nipah virus

nsp(s) = non-structural protein(s)

NTD = N-terminal domain

O-MT = 2'-O-methyltransferase

ORF(s) = open reading frame(s)

PDCoV = porcine Deltacoronavirus

PEDV = porcine epidemic diarrhoea virus

PLN = Pango lineage dynamic nomenclature

PLP = papain-like proteinase

pp1a/pp1ab = polyprotein 1a / polyprotein 1ab

RBD = receptor-binding domain

RCA = recent common ancestor

RDR(s) = recurrent deletion region(s)

RdRp = RNA-dependent RNA polymerase

RTC = replication transcription complex

S = spike protein

SARS-CoV-1 / 2 = Severe Acute Respiratory Syndrome Coronavirus 1 / 2

SFTS = Severe Fever with Thrombocytopenia Syndrome

sgRNA(s) = subgenomic ribonucleic acid(s)

TMD = transmembrane domain

TMPRSS2 = transmembrane serine proteinase 2

U = uracil

UTR = untranslated region

VOC = variant of concern

VOHC = variant of high consequence

VOI = variant of interest

VUM = variant being monitored

WGS = whole genome sequencing

WHO = world health organisation

WIV = Wuhan Institute of Virology

WT = wild-type

αmp = multispanning membrane protein

PREFACE

In the past few years, people worldwide have been highly affected by the pandemic of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). This novel virus and its evolution are the focus of this thesis.

This literature review will summarise up-to-date knowledge of SARS-CoV-2's emergence, evolutionary changes, and trends. In the introductory part, SARS-CoV-2's significant taxonomical characteristics will be described, and its close relatives will be introduced. The second part will concern the theories of the origin of this new virus. The third part will list and describe the characteristics of the epidemiologically most important variants of this virus known so far. Lastly, trends in the evolution of SARS-CoV-2 during the pandemic will be analysed.

1. BASIC CHARACTERISTICS OF CORONAVIRUSES

Coronaviruses (CoVs) infect only vertebrates. Research on coronaviruses has been hugely boosted by severe diseases they have caused to pigs, cows, chickens, dogs, cats, and humans in the last hundred years. In animals, CoVs cause gastroenteritis, peritonitis, hepatitis, respiratory diseases and diseases of the central nervous system. All human-infecting coronaviruses cause respiratory diseases and are of zoonotic origin (see Figure 1). Due to their outbreaks in the early 2000s, SARS-CoV-1 and MERS-CoV were spotted even by the public. They infect the lower respiratory tract and cause severe conditions. The SARS-CoV-1 virus caused a small pandemic that quickly dampened and has not re-emerged since 2004. Middle East respiratory syndrome coronavirus (MERS-CoV) did not cause a pandemic; it is causing mostly outbreaks in Saudi Arabia and surrounding countries where the intermediate host of this virus – dromedary, is kept. The virus is spreading mostly in households and as a nosocomial infection. Other human coronaviruses – Human coronavirus HKU1 (HCoV-HKU1), HCoV-OC43, HCoV-NL63, and HCoV-229 – mostly cause mild upper respiratory tract diseases. Up to 15-30% of cases of the common cold are associated with the four less severe coronaviruses. (Nieto-Torres *et al.*, 2014; Chen, Liu and Guo, 2020; Hasöksüz, Kiliç and Saraç, 2020; Salajegheh Tazerji *et al.*, 2020; Holmes *et al.*, 2021)

Enveloped virions of *Coronaviridae* have a spherical shape with a size of 50-160 nm. Their nucleocapsid (N) protein shows helical symmetry. All coronaviruses have characteristic petal-shaped spike (S) proteins on their surface, which play a crucial role in the infection. The genome of CoVs consists of non-segmented positive-sense single-stranded ribonucleic acid (+ssRNA). They are classified as group IV regarding the Baltimore classification scheme. (Baltimore, 1971; Holmes, 1999; Hasöksüz, Kiliç and Saraç, 2020; Malik, 2020)

SARS-CoV-2 belongs to the order *Nidovirales*, family *Coronaviridae*. Four genera of the subfamily *Orthocoronavirinae* are *Alpha-*, *Beta-*, *Gamma-*, and *Deltacoronavirus*. SARS-CoV-1, SARS-CoV-2, MERS-

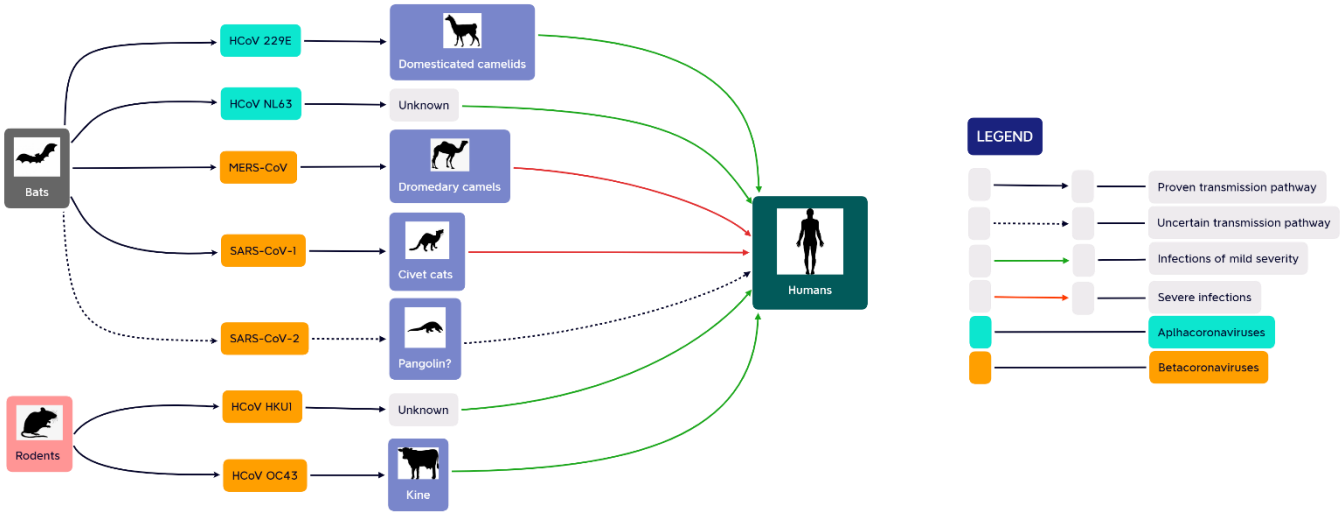

Figure 1: Pathways of transmission of human-infecting CoVs (adapted from Salajegheh Tazerji et al., 2020; modified).

CoV, HCoV-HKU1, and HCoV-OC43 belong to the genus *Betacoronavirus*. HCoV-NL63, together with HCoV-229E, is classified as *Alphacoronavirus*. These seven coronaviruses are the only ones known to infect humans. (Chen, Liu and Guo, 2020; Cuffari, 2020a; Hasöksüz, Kiliç and Saraç, 2020; Wan *et al.*, 2020; Y. Yang *et al.*, 2020; Liu, Liang and Fung, 2021)

1.1. Genome and Proteome

The average length of the coronaviral genome is 27-32 kb, the biggest among all RNA viruses. When RNA viruses replicate, the number of errors is high, resulting in many related genotypes called "quasispecies." This high error rate, caused by a low replication fidelity of their RNA polymerase, accelerates the adaptation of RNA viruses to different selective pressures. Compared to other RNA viruses, the replication competence in CoVs is enhanced by the exonuclease coded within the N-terminal domain (NTD) of non-structural protein (nsp) 14, which, apart from decreasing the number of changes created during each replication cycle, creates a barrier for using nucleoside analogues to treat the coronaviral infections through its proofreading activity. (Denison *et al.*, 2011; Chen, Liu and Guo, 2020; Hasöksüz, Kiliç and Saraç, 2020; Robson *et al.*, 2020)

The genomic RNA is used directly as a template for translation to polyprotein 1a/1ab (pp1a/pp1ab), which is then processed into 16 nsps, some of which form the complex necessary for replication and transcription called RTC. The coronaviral genome can be divided into six open reading frames (ORFs). Using discontinuous transcription, a set of subgenomic RNAs (sgRNAs) is synthesised by RTC thanks to the possession of common 5' and 3' leader and terminal sequences (see Figure 2). Most of the 16 nsps have their unique role in the replication process of CoVs. (Chen, Liu and Guo, 2020; Hasöksüz, Kiliç and Saraç, 2020; V'kovski *et al.*, 2021)

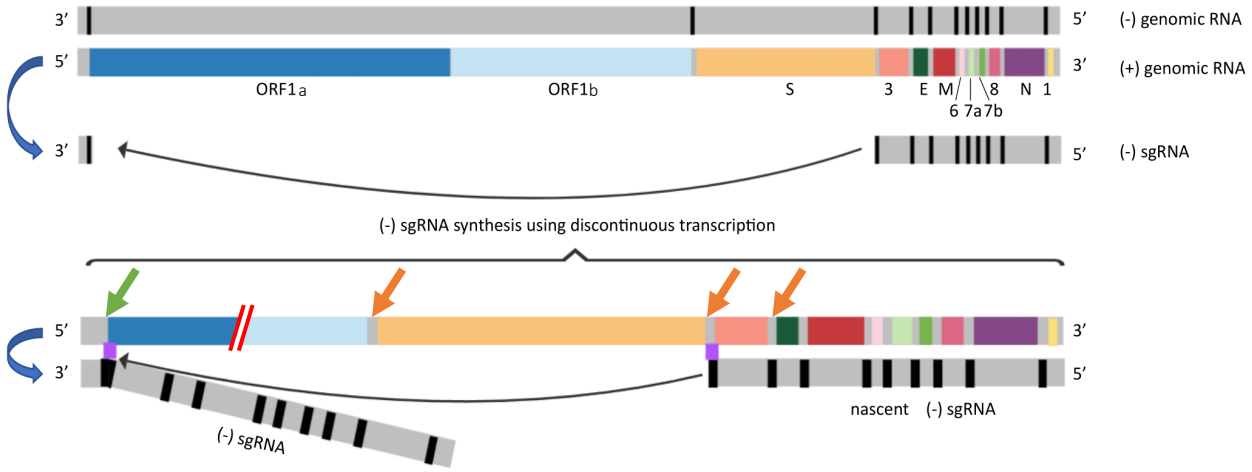

Figure 2: Scheme of RNA synthesis in CoVs (adapted from V'kovski et al., 2021; modified). The genomic RNA strand with positive sense is a template to produce both negative-sense genomic RNA and negative-sense sgRNAs. (-) sgRNAs are then used as a template for the (+) sg mRNA that encodes structural and non-structural proteins. The process of discontinuous transcription uses a pair of template switches. Body transcription regulatory sequence (TRS-B, orange arrows) moves the transcription process to the leader TRS (TRS-L, green arrow). Any present TRS-B can be used to start the discontinuous transcription leading to leader-body fusion and resulting in the synthesis of a characteristic set of (+) sg mRNAs. TRSs share conserved motif – ACGAAC recognised by RdRp.

1.1.1. Structural proteins

The capsid is assembled of four proteins. The encapsidation of RNA is mediated by the N protein, which consists of two domains. Both domains can bind viral RNA but utilise different mechanisms. N also binds nsp3, whose Papain-like proteinase 1 (PLP) domain is responsible for the cleavage of 1a/1ab polypeptides, promoting cytokine expression and blocking the host's innate immune response (IIR). It also promotes the genome's binding to RTC and packaging of the genome into virions. The second structural protein, envelope (E), is responsible for virus assembly and release and plays a significant role in viral pathogenesis. Transmembrane (M) protein accounts for the shape of the virion. It curves the membrane and binds protein N. These actions are mediated through the cooperation of the three transmembrane domains of the M protein. The primary function of the S protein of CoVs is the attachment of the virus to the host cell's receptor, the first step in the viral entry to the host cell. Spike proteins form homotrimers in open or closed conformation (see Figure 3). In some Betacoronaviruses fifth structural protein can be found – hemagglutinin-esterase (HE) protein which enhances cell entry mediated by the S protein. (Delmas and Laude, 1990; Holmes, 1999; Chang *et al.*, 2006; Nieto-Torres *et al.*, 2014; Cui *et al.*, 2015; Lei, Kusov and Hilgenfeld, 2018; Chen, Liu and Guo, 2020)

The structure of the S protein is conserved across the family *Coronaviridae*. It has two domains – S1 and S2. The S1 domain comprises the signal peptide, NTD, and receptor-binding domain (RBD). RBD is one of the most variable viral parts. Unlike the S1 domain, the central role of S2 is in the fusion of viral and cell membranes. (Fehr and Perlman, 2015; Cui, Li and Shi, 2019; Huang *et al.*, 2020; Ou *et al.*, 2020; Xia, 2021)

1.1.2. Non-structural proteins

Non-structural viral proteins play a significant role in the replication of CoVs. Known functions of all the nsps found in CoVs are listed in the table below. Nevertheless, some nsps might be involved in multiple processes, and some of the specific roles of some nsps still need to be described.

Non-structural protein	Functions
nsp1	Cellular mRNA degradation, inhibition of interferon (IFN) signalling
nsp2	Translation stimulation, enhancing microRNA-mediated translational repression (thus blocking host antiviral immunity)
nsp3	Papain-like-proteinase; cleavage of the polypeptides, blocking of the host innate immune response, promotion of cytokine expression.
nsp4	Formation of double-membrane vesicles (DMV)
nsp5	Chymotrypsin-like cysteine proteinase ($3CL^{pro}$), also known as main proteinase (M^{pro}); cleavage of the polypeptides, inhibition of IFN signalling.
nsp6	Restriction of autophagosome expansion, formation of DMV
nsp7	Cofactor with nsp8 and 12
nsp8	Primase; cofactor with nsp7 and 12
nsp9	RNA binding
nsp10	Scaffold protein for nsp14 and 16
nsp11	Partakes in the interaction between the virus and the host membrane; no independent function is known
nsp12	RNA-dependent RNA polymerase (RdRp; primer dependent)
nsp13	RNA helicase, 5' triphosphatase
nsp14	Exoribonuclease, guanine-N7-methyltransferase (N7-MTase)
nsp15	Nidoviral RNA uridylate-specific endoribonuclease (NendoU); evasion of dsRNA sensors
nsp16	2'-O-methyltransferase (O-MT); avoiding recognition by a melanoma differentiation-associated protein 5 (MDA5), negative regulation of innate immunity

Table 1: Functions of the coronaviral non-structural proteins, adapted from Chen et al., 2020, eked out regarding Korneeva et al., 2023; Naeli et al., 2023; Zhang and Yang, 2022

1.2. Mechanism of viral entry in the cell

CoVs enter the cell using two different mechanisms. The first possibility is binding to a specific receptor which leads to the fusion of viral and cell surface membranes and, thus, the delivery of the viral genetic information in the cell. The second way is through endocytosis of the whole viral particle; low pH in the endosome results in the fusion of the viral envelope with the endosomal membrane and the release of the RNA in the cytosol. (Nassar *et al.*, 2021)

Coronaviruses interact with several host receptors. SARS-CoV-1, SARS-CoV-2 and NL63 bind to the human angiotensin-converting enzyme 2 (hACE2) receptor, while cell entry of MERS-CoV is mediated by the dipeptidyl peptidase 4 (DPP4) receptor. Virus 229E uses the interaction with the aminopeptidase N (APN) receptor. CoVs can also use other receptors to facilitate cell entry, although those mentioned above are used primarily. (Kolb, Hegyi and Siddell, 1997; Yang *et al.*, 2014; Tang, Liu and Chen, 2022)

Fusion of the membranes, a key step in the cell entry for all enveloped viruses, has high energetic barriers that must be overcome to realize the process. Free energy used for enabling the fusion process comes from the fusion protein refolding. The fusion protein for coronaviruses is their spike protein. It exists in metastable prefusional (see Figure 3A) and a stable postfusional (see Figure 3B) conformation. (Rand and Parsegian, 1984; Weissenhorn *et al.*, 1999; Harrison *et al.*, 2006)

The fusion of membranes is often performed by fusion peptides that penetrate the target cell membrane and install an anchor there. Transmembrane (TMD) and cytoplasmic tail (CtD) domains, parts of the S2, then form the second anchor for the virion. The S2 domain then changes its conformation, thus bringing the two membranes closer, which enables fusion. This change in conformation depends on the signal indicating the

proximity of a viral particle to a target cell or between the infected and target cell. The change in the state of the S2 domain is catalysed by the transmembrane serine proteinase 2 (TMPRSS2) or a similar proteinase expressed on the surface of a target cell. (Fehr and Perlman, 2015; Cui, Li and Shi, 2019; Huang *et al.*, 2020; Ou *et al.*, 2020; Xia, 2021)

1.2.1. Open/Close conformation

The C-terminal trimerisation stabilises the structure of the prefusion S protein conformation (see Figure 3A) and can be further stabilised by proline substitutions F817P & A924P; these, however, were only present in the artificial mutants. Four domains of the S1 wreathe the threefold axis, which results in the hiding of the S2. The furin cleavage site (FCS) is exposed in a surface loop. Most of the S2 polypeptide coats the central helix of the calponin homology (CH) domain. (Cai *et al.*, 2020)

The postfusion conformation differs from the prefusion, mostly in the S2 part of the trimer. The S2 segment of the protein forms a long rigid structure consisting of CH and HR1 (rho binding) domains. N-linked glycans on the surface are a key characteristic of the postfusion conformation of the S2 fragment (see Figure 3B). (Cai *et al.*, 2020)

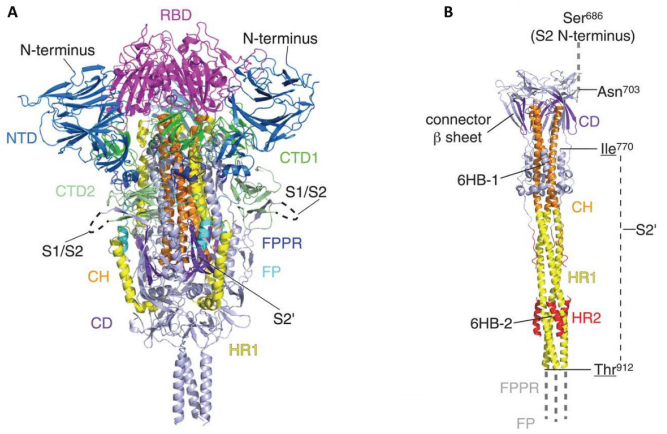

Figure 3: Conformational states of the SARS-CoV-2's S protein (adapted from Cai et al., 2020; modified).

***A** – prefusional S trimer*

***B** – postfusional S2 trimer*

1.2.2. TMPRSS2

TMPRSS2 is a protein that plays a crucial role as a mediator of cell entry for several viruses. Mutations in neither humans nor mice (including knockout mice) are known to be associated with any pathology. This can either mean that its function is redundant or that its function might be non-vital and highly specialised, leading to the effects of the mutation or knockout being observable only in the conditions such as disease or other significant stressors. The human TMPRSS2 gene consists of fourteen (Jacquinet *et al.*, 2000) or fifteen (Thunders and Delahunt, 2020) exons. Its total weight is about 70 kDa, but it undergoes autoregulated proteolytic cleavage, leading to the polypeptide of 32 kDa. (Jacquinet *et al.*, 2000; Kim *et al.*, 2006; Bertram *et al.*, 2010; Thunders and Delahunt, 2020; Peng *et al.*, 2021)

1.2.3. Cell entry of SARS-CoV-2

Early studies on SARS-CoV-2 have shown that the virus uses hACE2 to enter the host cell. The viral S protein's RBD binds to the hACE2, and then the proteolysis of the S2 fusion machine is executed by TMPRSS2 or a similar proteinase present on the host cell membrane. These actions lead to the activation of the spike protein, and activation promotes the entry of the virus into the cell. (Fehr and Perlman, 2015; Huang *et al.*, 2020; Letko, Marzi and Munster, 2020; Z. Liu *et al.*, 2020; Nassar *et al.*, 2021; Peng *et al.*, 2021)

Contrary to the case of SARS-CoV-1 and MERS-CoV, about 20% of the membrane fusion is performed even when the TMPRSS2 is inhibited. This means that SARS-CoV-2 can enter even the cells that do not have TMPRSS2 on their surface. It was shown by Yamamoto et al. (2022) that the cases of TMPRSS2-independent fusion decrease when the tissue culture is treated with metalloproteinase inhibitors and thus might be used as therapeutics to treat COVID-19 patients. However, no clinical tests have been performed so far. As such cell entry pathway was not observed in the case of any other human infecting CoVs, it is suggested that the metalloproteinase pathway of membrane fusion is specific only for SARS-CoV-2. The utilisation of the metalloproteinase (also known as endosomal) pathway is further dependent on the presence of the SARS-CoV-2-specific S1/S2 boundary and the S2 domain, as they play a crucial role in the cleavage of the furin site. Furthermore, the TMPRSS2-independent fusion of membranes can only be realized in the environment of the endosome, not on the cell surface. (Shen *et al.*, 2017; Meng *et al.*, 2022; Yamamoto *et al.*, 2022)

1.2.3.1. Furin[1] cleavage site (FCS)

Across the *Betacoronavirus* genus, the S protein sequence is relatively conserved with 75% similarity. One of the distinctions of SARS-CoV-2 (together with MERS-CoV) from other *Betacoronaviruses* is the FCS in their S protein. Introducing a cleavage site results in less specific cell entry in the case of several coronaviruses. Thanks to the presence of such a site, the spike cleavage is less specific as it can be cleaved by several different proteinases. In the case of SARS-CoV-2, it is an insertion of 4 amino acids (P-R-R-A) upstream of the S1 cleavage site (position 681-684 or 682-685 in the peptide; see Figure 4). The arginine in the FCS (and other viral parts) is encoded by the CGG codon, which is not used primarily in humans due to its hypermutability. Hence, the possibility of the virus gaining this codon in some intermediate host(s) was suggested. However, the intermediate host would have to differ in codon usage. Thus, the intermediate host would likely belong to other than the *Mammalia* class, as the codon bias is similar among all mammal species. (Jaimes *et al.*, 2020; Johnson *et al.*, 2020, 2021; Li X. *et al.*, 2020; Schulze, Hanchard and Wangler, 2020; Chan and Zhan, 2022)

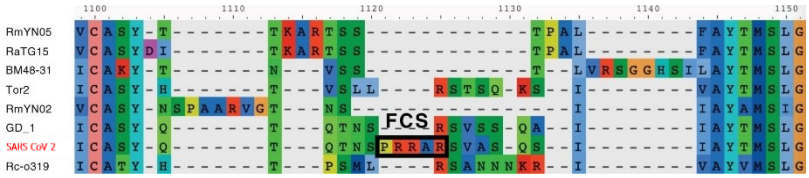

Figure 4: Multiple sequence alignment of SARS-CoV-2's and closely related CoVs' S protein (adapted from Chan and Zhan, 2022; modified). FCS is highlighted in a frame.

Johnson et al. (2020, 2021) and Peacock et al. (2021) have studied the importance of the FCS in SARS-CoV-2's pathogenesis. According to their results, the S protein of the viruses with a deletion in the FCS is less processed (folded less neatly), attenuating replication in respiratory cells and reducing pathogenesis. Competition assays revealed that the wild-type (WT) virus is predominant over ΔPRRA mutant in cell lineages derived from human colorectal and lung carcinoma. (Chan and Zhan, 2022; Johnson et al., 2021, 2020[2]; Li et al., 2020; Peacock et al., 2021)

[1] Furin is an endoproteinase activating several substrates, such as growth factors, pathogenic agents, receptors, etc., through hydrolysis at specific internal peptide bonds. (Cuffari, 2020b; UniProtKB, 2023a)

[2] Preprint

The mentioned studies by Johnson et al. and Peacock et al. presented different results of the pathogenicity of SARS-CoV-2's WT and ΔPRRA mutant dependent on the cell lineage used for the experiment. The used cell lines (Vero E6 and Calu-3) differ in the expression of host serine proteinases, including TMPRSS2. During competition assay, the WT and mutant remained at equal levels, suggesting that TMPRSS2 reduces the advantage in fitness and replication of the mutant over the WT when serine proteinases are expressed. Not only TMPRSS2 and similar host proteinases do facilitate the entry of SARS-CoV-2 inside the host cell, but another group of proteins with similar effects are also cathepsins (cysteine proteinases). When inhibitors of serine proteinases (such as TMPRSS2) and cysteine proteinases were used, infections of both SARS-CoV-1 and SARS-CoV-2 in the host cells were less effective. (Cai *et al.*, 2020; Johnson *et al.*, 2021; Peacock *et al.*, 2021)

1.3. Fusogenicity

Fusogenicity, the ability to promote membrane fusion, is one characteristic that differs among the SARS-CoV-2 variants through their evolution. The cooperative functions of multiple parts of S protein, such as RBD and FCS, or host proteinases, such as TMPRSS2 or cathepsins, mainly influence this characteristic. In the case of SARS-CoV-2 formation of syncytia of host cells happens already six hours post-infection *in vitro*. The advantage in the syncytium formation during the infection is earlier achieving the peak viral titre. It is also suggested that syncytia formation might be a strategy for immune system evasion. (Cifuentes-Muñoz and Ellis Dutch, 2019; Buchrieser *et al.*, 2020; Zhu *et al.*, 2020; Jessie and Dobrovolny, 2021)

2. Taxonomy of Coronaviruses

In the *Coronaviridae* family, the subfamilies and genera taxonomy is based on the sequence of several domains in the replicase and pp1a/pp1ab that are conserved across the whole family. These domains are ADP-ribose-1'-phosphatase (ADRP), nsp5 ($3CL^{pro}$), nsp12 (RdRp), nsp13 (helicase), nsp14 (exonuclease), nsp15 (NendoU), and nsp16 (O-MT). Specimens with more than 90% identity in these conserved domains are considered specimens of the same species. Specimens that share less than 46% identity in sequence with any established family member are considered new members of a new genus. (Keep *et al.*, 2018; Woo, 2023)

2.1. Alphacoronavirus

The main difference between Alphacoronaviruses from other genera of *Coronaviridae* is in the type of nsp1, which differs significantly from the genus *Betacoronavirus* and appears not to have any counterpart in the genus *Gammacoronavirus*. All species of this genus also share an accessory gene which can be found in ORF3. This accessory gene encodes the multi-spanning membrane protein (αmp; polytopic trans-membrane protein). Some "subspecies[3]" can carry as many as six accessory proteins (for example, canine coronaviruses belonging to the *Alphacoronavirus*1 specie). (Woo, 2023)

There are eight species in the genus *Alphacoronavirus*, many of which have several subspecies. The species are *Alphacoronavirus1, Scotophilus* bat CoV 512*, Rhinolophus* bat CoV HKU2, Porcine epidemic

[3] Subspecies are not officially recognised in the taxonomy of viruses.

diarrhoea virus (PEDV), *Minopterus* bat CoV HKU8*, Minopterus* bat CoV1 and two coronaviruses infecting humans – HCoV 229E, HCoV NL63. Among "candidate" viruses are, for example, Ferret CoV, Chinese ferret badger CoV DM95/03, Yellow-bellied weasel CoV GX/D726/2005 and many more, most of which infect bats and are both pathogenic and commensals. (Woo, 2023)

2.1.1. HCoV 229E

HCoV 229E, the first isolated human-infecting CoV, got its name from the code of the specimen it was isolated from – the standard tissue culture 229E. This culture was started from a nasal swab taken in 1965 by a patient suffering from a common cold. The most likely origin of this virus is in African hipposiderid bats. Before infecting humans, it seems to have infected domesticated camelids, considered intermediate hosts of this virus. The specific sequence in the APN receptor, required for the 229E binding, was identified in humans and the cell lineages of felines and cats, but there are likely more possible animal hosts. (Kolb, Hegyi and Siddell, 1997; Liu, Liang and Fung, 2021; Tang, Liu and Chen, 2022)

The symptoms in adult healthy patients are nearly indistinguishable from the common cold. Nevertheless, it can cause severe disease in elderly people and children. Furthermore, infections of HCoV 229E can be life-threatening if the patient is immunocompromised. (Liu, Liang and Fung, 2021)

2.1.2. HCoV NL63

HCoV NL63 was identified in 2004 during the post-SARS-CoV-1 era and was isolated from the phlegm of a seven-month-old child suffering from bronchitis, fever, and conjunctivitis. NL is an acronym for Netherlands – the state where the virus was isolated for the first time. Later studies have shown that this virus is spread worldwide and is probably circulating in the human population for nearly a millennium. Huynh et al. (2012) found a novel coronavirus that seems to share a recent common ancestor (RCA) with NL63 and infects the tricoloured bat (*Perimyotis subflavus*). Following their observations, it was suggested that HCoVs could infect multiple mammalian species and thus may lead to reverse zoonoses. Such reversion poses a risk of viral recombination (further covered in Chapter 3), which may result in a variant with higher virulence. (Pyrc *et al.*, 2006; Huynh *et al.*, 2012; Liu, Liang and Fung, 2021)

HCoV NL63 and HCoV OC43 (see Chapter 2.2.) cause most human coronaviral infections requiring hospitalisation. NL63 is also responsible for the development of croup in children. Children under five years of age are infected most often. Every year it is estimated that up to 10% of people are infected. Contrary to 229E, which peaks in winter, epidemics of NL63 peak during the spring and summer (in Hong Kong); this indicates that seasonality is not restricted to winter periods. (Liu, Liang and Fung, 2021)

2.2. BETACORONAVIRUS

Same as in the case of *Alphacoronavirus,* the unique, yet conserved across *Betacoronavirus*, nsp1 sequence is used for the distinction. Four main lineages in the genus *Betcoronavirus* (A-D) are distinguishable regarding each lineage's unique set of accessory genes (as mentioned above). (Woo, 2023)

Seven species in total are classified into the *Betacoronavirus* specie. *Tylonycetris* bat CoV HKU4, SARS-related CoVs, *Rousettus* bat CoV HKU9, *Pipistrellus* bat CoV HKU5, Murine CoV, HCoV HKU1 and *Betacoronavirus1* that has several subspecies including viruses infecting cows, pigs, horses and human-infecting HCoV OC43. (Woo, 2023)

2.2.1. HCoV HKU1

HKU1 was isolated at Hong Kong University (hence the name HKU1) from an adult patient suffering from pneumonia. HKU1 has been identified only in humans so far. According to genetic studies, HKU1 is phylogenetically close to mouse hepatitis and rat sialodacryoadenitis viruses. It was suggested that the original host of an HKU1 predecessor was some rodent species. (Woo *et al.*, 2005; Corman *et al.*, 2018; Liu, Liang and Fung, 2021)

The HCoV HKU1 is generally frequent in adults and causes rhinorrhoea, cough, fever, nasal congestion, sore throat, chills, etc. HKU1 is also associated with exacerbations of asthma. It usually co-circulates with the respiratory syncytial virus and has an epidemic peak shortly before the influenza season. (Liu, Liang and Fung, 2021)

2.2.2. HCoV OC43

OC43 was recovered from tracheal tissue culture in 1967. The specimen that it was isolated from had assigned code Organ culture 43. There are seven known genotypes labelled A-G. Although at first considered closely relative to 229E, it was later found that these two viruses differ serologically. OC43 belongs to the species *Betacoronavirus1*. This species infects a broad spectrum of hosts. The phylogenetically closest are the bovine coronaviruses; cows and pigs are considered intermediate hosts of this virus. More than two-thirds of patients infected with OC43 usually suffer co-infection with other respiratory diseases, such as rhinovirus, parainfluenza, or enterovirus. (Vijgen *et al.*, 2005; Corman *et al.*, 2018; Liu, Liang and Fung, 2021)

This virus has also been shown to infect neurons *in vivo* and cause encephalitis. The epidemic peak of OC43 is during winter in temperate climate areas. (Liu, Liang and Fung, 2021)

2.2.3. MERS-CoV

MERS-CoV was first identified in September 2012. The most probable major reservoir is insectivorous bats from the genus *Pipistrellus* or *Nycteris,* with dromedary camel being the intermediate hosts. (Annan *et al.*, 2013; Hijawi *et al.*, 2013; Raj *et al.*, 2014; ECDC, 2017, 2020; Tang, Liu and Chen, 2022)

All cases diagnosed since the first emergence of MERS-CoV have a strong link to Middle Eastern countries. Most outbreak cases diagnosed are restricted to the Arabian Peninsula, with Saudi Arabia having the highest number of cases. The camel trade is the cause of the high number of Saudi Arabian cases. There are also cases of MERS-CoV in Yemen, Egypt, Syria, Western Africa etc. However, many confirmed cases are asymptomatic (often diagnosed among blood donors). (Annan *et al.*, 2013; Hijawi *et al.*, 2013; Raj *et al.*, 2014; ECDC, 2017, 2020; Degnah *et al.*, 2020)

Patients infected with MERS-CoV have shown shortness of breath, fever, and cough. In a minority of cases, nausea and diarrhoea can also be symptoms of infection. Many patients with more severe infections even developed pneumonia or failure of kidneys. The mortality among those infected with MERS-CoV reaches up to 35%. (CDC, 2019)

2.2.4. SARS-CoV-1

SARS-CoV-1 (previously SARS-CoV) emerged in 2002 in South China. A vast diversity of SARS-like CoVs can be found in racoon dogs, civets, and bats. Horseshoe bats – genus *Rhinolophus* are the original reservoirs of the SARS-CoV-1. It is suggested that SARS-CoV-1 originates in recombination of bat viruses. The recombined virus was later transmitted into civets, and humans were infected following the circulation of the virus in civets (probably in civet farms). Bidirectional transmissions – from civets to humans and back – are documented. (Guan *et al.*, 2003; Li *et al.*, 2005; Janies *et al.*, 2008; Bolles, Donaldson and Baric, 2011; Wang *et al.*, 2018)

The symptoms of SARS-CoV-1 do not differ significantly from those of SARS-CoV-2. It may even be asymptomatic. Patients usually have high fevers, coughs, and shortness of breath. These symptoms, common for most respiratory diseases, are often accompanied by diarrhoea. In the later stages of the disease, pneumonia may progress to fatal respiratory failure. The overall mortality of patients suffering from SARS-CoV-1 infections has reached beyond 7%. Interhuman transmission of the virus happens via the inhalation of droplets. The SARS-CoV-1 pandemic in 2003 and 2004 was terminated quickly due to the severity of symptoms causing the infected people to be easily identified and remain isolated. (ECDC, 2010)

2.3. GAMMACORONAVIRUS

Unlike the viruses from the previously mentioned genera, viruses from the *Gammacoronavirus* genus lack the gene for nsp1 and do not generally share any characteristics in their gene composition, genome organisation, replication or virion morphology with other species belonging to the genera**.** Two species are known to belong in this genus – Avian coronavirus and Beluga whale coronavirus. (Woo, 2023)

2.4. DELTACORONAVIRUS

Genus *Deltacoronavirus* is the phylogenetically youngest genus belonging to the family *Coronaviridae,* as it was established in 2011. Porcine Deltacoronavirus (PDCoV), Munia coronavirus HKU13 and Thrush coronavirus HKU12 are species belonging to the *Deltacoronavirus* genus. (Woo, 2023)

3. ON THE ORIGIN OF SARS-COV-2

About 60% of the known human-infecting pathogens (viruses, bacteria and fungi; circa 1,400 species as of 2012) are zoonotic. Out of emerging and re-emerging human zoonotic diseases, more than 37% are caused by RNA viruses. Several bat species are the original hosts of previously mentioned viral infections and, together with rodents, are considered the most important reservoir of zoonotic infections. Bats have been the

original hosts of multiple zoonotic viruses introduced to humans, such as SARS-CoV-1 or Nipah virus[4] (NiV). Hantaviruses, whose reservoirs are rodents, caused a major outbreak of severe and sometimes fatal respiratory disease in the United States in the 1990s. Apart from Hantaviruses, Arenaviruses, which from the beginning of the 1990s often infected humans in South America, were detected in rodents. These outbreaks are likely to be connected to the El Niño events, which occurred in 1991-92 and led to the overpopulation of rodents on the American continent. It is also reported that some viruses historically emerged from birds, mosquitoes, primates and several other mammals, hardly ever. However, they were the original reservoir of the infection. It is common, though not necessary, for a virus to infect several species of animals before finally infecting humans, who are often dead-end hosts. In many cases, transmission via an intermediate host is reported, for example, in SARS-CoV-1, Hendra virus[5] (HeV), NiV, and MERS-CoV. (Brunt *et al.*, 1993; Engelthaler *et al.*, 1999; Chua *et al.*, 2000; Halpin *et al.*, 2000; Taylor, Latham and woolhouse, 2001; Mickleburgh, Hutson and Racey, 2002; Guan *et al.*, 2003; Weiss and McMichael, 2004; Li *et al.*, 2005; Heeney, 2006; WHO, 2012; Raj *et al.*, 2014; Lednicky *et al.*, 2021; Muga *et al.*, 2021; Xiao *et al.*, 2021; Vlasova *et al.*, 2022)

The virus must adapt to the new hosts' internal environment to cross the barrier. Viruses can accomplish this in numerous ways. Reassortment is a method of adaptation of viruses with the segmented genome (e.g., influenza). Such an event occurs when two viral strains (e.g., avian and human) infect animal species susceptible to infection of strains originating in different animal species (e.g., pig). In this intermediate host, genomic segments of these viruses are mixed and create a new (reassorted or changed) viral strain with new characteristics. Due to newly gained characteristics, such reassorted viruses may pose a significant risk of becoming epidemic or pandemic. (Nichol, Arikawa and Kawaoka, 2000; Taylor, Latham and Woolhouse, 2001; Blancou *et al.*, 2005; Cunningham, 2005; Vorou, Papavassiliou and Tsiodras, 2007)

Another way how viruses evolve and can also gain the ability to infect new hosts is by antigenic drift. Antigenic drift is described as small changes in the viral genome sequence – point mutations. It occurs in partially immune host populations and occasionally leads to new viral strains. Most such strains do not outcompete the parental strain after undergoing antigenic drift, as they do not show any growth or selection advantage. (Archetti and Horsfall, 1950; Gerhard and Webster, 1978; Heeney, 2006)

Antigenic shift leads to a more dramatic virus change and occurs via horizontal gene transfer (HGT). It results in a significant change in phenotype. HGT is a process of moving the genetic information between two or more individuals, thus fuelling the organism's evolution. It can be observed in various microorganisms, including viruses and bacteria. In bacteria, the antigenic shift plays a significant role in building antibiotic resistance. It infrequently happens in viruses, yet its consequences might be severe. This process alters the antigens on the pathogen's surface to such an extent that the immunity gained by the host against the previous strain of the virus is no longer protective. In the case of SARS-CoV-2, it was associated with the emergence of the Omicron variant being vastly different from its predecessors. Recombination is a special case of

[4] Member of family *Paramyxoviridae*, associated with encephalitis, mild to severe illnesses and also a cause of death (*Nipah Virus (NiV) | CDC*, 2022)

[5] Member of the family *Paramyxoviridae* cause respiratory and neurological diseases in horses and humans (*Hendra Virus Disease | CDC*, 2022)

antigenic shift happening when viruses with segmented genomes of at least two different parent strains infect the same host resulting in progeny having some genes from each parental virus. Earlier, it was thought that for the recombination to occur, the virus must be of the same type (e.g., two encephalitis viruses, the recombination of which created the Western equine encephalitis virus). Nonetheless, proven examples of recombination between viruses belonging to different families exist. The new recombinant viruses have often altered virulence and are more likely to evade the host's immune response. Nonetheless, recombinant viruses are also often not viable. (McLean and Donohue, 1959; Narayan, Griffin and Chase, 1977; Fleischmann, 1996; Lambert, Martin and Lanciotti, 2003; Harrison *et al.*, 2006; Heeney, 2006; Lefeuvre *et al.*, 2009; Burmeister, 2015; CDC, 2022)

Nevertheless, another possibility of viral introduction to humans exists. Since the pandemic started, it has been speculated whether the SARS-CoV-2 might have emerged from a laboratory incident or even if the virus was not constructed in the laboratory. Arguments in favour of both these theories ("zoonotic emergence" and "laboratory leak") exist.

3.1. ZOONOTIC EMERGENCE THEORY

This hypothesis suggests that SARS-CoV-2 was transmitted from an animal reservoir into humans. As previously described, this can happen with or without viral adaptations in an intermediate (animal) host. The high frequency of recombination and genomic plasticity of CoVs underlie their ability to cross species during transmissions. (Cyranoski, 2020; Holmes *et al.*, 2021)

In favour of this theory is a solid link to the Huanan seafood market (HSM) in Wuhan, China, as more than a quarter of all the cases of COVID-19 reported in December 2019 were directly linked to this market selling live animals. As presented by Worobey *et al*. in their study from 2022, the HSM is located within the highest density of residences of people diagnosed with COVID-19 at an earlier date. On the contrary, the pattern of cases from early January to Mid-February is scattered significantly more across the city map, and the epicentre moves closer to the places of greater population density. According to data from Worobey *et al*., no other region in Wuhan shows a density of cases that could be comparable to the areas closely surrounding the HSM. The authors of the cited study also mention that the data might be affected by population density and age structure in the centre of Wuhan city. (WHO, 2021; Worobey *et al.*, 2022)

Data from Worobey *et al*. show that the greatest concentration of positive swab samples taken at the HSM surfaces was in the south of the area's Western side. With following-up time, it has been recorded that the positive human cases have moved with time to the northern part of the Western market. This suggests the infection spread between humans from the epicentre to the outer rims. (WHO, 2021; Gao *et al.*, 2022; Worobey *et al.*, 2022)

The study by Xiao *et al*. from 2021 reported that nearly fifty thousand individuals of 38 species of farmed and wildlife animals were sold in Wuhan wildlife markets, including those that might have been the vector of the zoonotic emergence of SARS-CoV-2. All these animals were traded between May 2017 and November

2019[6]. However, the animals were not screened for the presence of SARS-CoV-2 virus. All analyses were done only on the swabs of the surfaces and cages that sold animals were in contact with and wastewater samples. As the hygiene conditions and welfare of the sold wild animals were poor[7], the possibility of transmission of different pathogens was vast. Many animal species sold in the observed markets are ordinary hosts of multiple zoonotic diseases or parasites. Besides the officially listed species of animals sold at the HSM, DNA of other animal species were identified, some of which were later proven susceptible to SARS-CoV-2 infection, suggesting illegal trade with wildlife as a culprit of the first SARS-CoV-2 outbreak. A more recent study by Crits-Christoph et al. from 2023[8] also identified mammals' mitochondrial DNA (mtDNA) sequences predicted to be prone to SARS-CoV-2 infection. This study adds new suspected intermediate host species: raccoon dogs, hedgehogs, bamboo rats, marmots, weasels and porcupines. Moreover, species whose DNA was analysed in this study were sold alive in significant numbers in the likely epicentre of the HSM. Besides the transmission from living animal hosts, frozen seafood was examined as a possible source of the infection with negative results. (P. Liu *et al.*, 2020; P. Zhou *et al.*, 2020; Zhao, Cui and Tian, 2020; WHO, 2021; Xiao *et al.*, 2021; Chand, 2022; Worobey *et al.*, 2022; Crits-Christoph *et al.*, 2023)

The study by Latinne et al. from 2020 analysed 1,246 CoV sequences[9]. They have concluded that the closest SARS-CoV-2 relative is a group of viruses commonly infecting horseshoe bats (*Rhinolophus sp.*) inhabiting the area of Yunnan province in China. This finding further supports earlier reports of the close similarity between SARS-CoV-2 and the bat coronavirus RaTG13 (BatCoVRaTG13 or RaTG13), found in anal swabs of *Rhinolophus affinis* living in the Yunnan province. The RaTG13 shares more than 96% of its genetic sequence with SARS-CoV-2. Nevertheless, it has been suggested by Prof Zhengli[10] (as cited by Cyranoski, 2020) that even such high genomic similarity is not enough for the RaTG13 to be considered the direct ancestor of SARS-CoV-2. Another bat species (*Rhinolophus malayanus*) lives in Myanmar and Laos, neighbouring China's Yunnan province. The BANAL-52[11] virus found in these bats has even greater sequence similarity than that of RaTG13 (over 96%). Therefore, there is a possibility that SARS-CoV-2's ancestor originating in this species also exists. However, a substantial geographical gap exists between the natural habitat of *Rhinolophus affinis* and *malayanus* bats and the first human cases of the infection by the novel

[6] The study first concerned Severe Fever with Thrombocytopenia Syndrome (SFTS), showing no Human-to-human transmissibility, which emerged in 2009.

[7] Living animals were often kept in unclean cages that were in stacks. Most vendors also offered butcher services of the animals right on the spot, in conditions raising questions on animal welfare and food hygiene. In addition, about 30% of sold animals suffered from wounds caused by guns or traps, indicating hunting illegal wildlife. (Xiao *et al.*, 2021; Worobey *et al.*, 2022)

[8] Preprint

[9] 630 partial RNA-dependent RNA polymerase (RdRp) sequences collected from bat rectal swabs in China, 608 bat's CoV and eight pangolin's CoV sequences from China loaded from GenBank. (Latinne *et al.*, 2020)

[10] Prof Shi Zhengli is the principal researcher in WIV. Her main area of expertise is viruses originating in bats (which earned her the nickname "bat-woman"). Her research group works on coronaviruses, adenoviruses, orthoreoviruses, circoviruses, paramyxoviruses, filoviruses, viruses of hepatitis, etc. and has discovered a range of novel viruses as well as viral antibodies (mainly in bats). (Chinese Academy of Sciences and Wuhan Institute of Virology, 2023)

[11] SARS-CoV-2 related coronavirus. Shares up to 96.8% of the genome sequence. (Mallapaty, 2021)

coronavirus SARS-CoV-2. Therefore, other possibilities for introducing the viruses to Wuhan were explored. (Forni *et al.*, 2017; Cyranoski, 2020; Latinne *et al.*, 2020; Holmes *et al.*, 2021; WHO, 2021; Xiao *et al.*, 2021)

Wang *et al.,* in their study from 2018, state that SARS-CoV-1-related viruses continuously infect humans living nearby the caves in Yunnan province, which *Rhinolophus* bats inhabit. None of the positively tested individuals has been close to any SARS-CoV-1 outbreak. Most of them, however, own livestock animals, which might be intermediate hosts. One handled a bat corpse, and every tenth individual confirmed seeing bats fly close to their households. Regarding mentioned results, it was thought that SARS-CoV-2 might have gotten into the population via bat catchers. However, no antibodies (Abs) against SARS-CoV-2 were found in people living near the bat caves or anyone involved in catching bats. (Wang *et al.*, 2018; Lam *et al.*, 2020; Holmes *et al.*, 2021; WHO, 2021)

Considering the transmission via an intermediate host, several species are considered: pangolins, mink, dogs, ferrets, turtles, snakes, yaks, cats, and pigs. Several species have already been ruled out, e.g., mice.[12] As two distinct types of SARS-CoV-2-like CoVs are found in Malayan pangolins, these animals were also considered a possible host of SARS-CoV-2. These SARS-CoV-2-like CoVs got assigned names GD/1/2019 and GX/P2V/2017. While they display great sequence similarity in RBD with SARS-CoV-2, the similarity of the whole genome is smaller than with RaTG13 (90.4% and 85.48%, respectively). Contrary to bat SARS-CoVs, pangolin CoVs bind, due to their RBD similarity, quite efficiently but less strongly than SARS-CoV-2 to human ACE2 receptors. (Lam *et al.*, 2020; Zhao, Cui and Tian, 2020; Niu *et al.*, 2021; WHO, 2021)

SARS-CoV-2 with RBD of pangolin CoVs were successfully transmitted to humans, monkeys, rabbits, horses, civets, racoon dogs, chickens, hedgehogs, etc. (18 species total) *vivo*. In *in vitro* experiments, the GD/1/2019 virus binds to human HeLa-hACE2 cell receptors more effectively than SARS-CoV-2. Pangolins were also positive for Abs which reacted with SARS-CoV-2's S protein. Furthermore, pangolins and bats are the only known animal species being provably infected by SARS-CoV-2-like CoVs years before the emergence of SARS-CoV-2 and the respective pandemic. However, the study of Xiao et al. (2021) points out that no pangolins were traded at the HSM, and neither were any of the relevant bat species. In addition, the exact sequence of SARS-CoV-2 was not isolated from any of the considered host animal species. The viruses closely related to SARS-CoV-2 found in bats and pangolins were too distant to be the SARS-CoV-2 direct predecessor (origin in recombination event is hardly refutable). Most likely, the SARS-CoV-2 did not originate from any of these viruses, with mentioned species as intermediate hosts. Still, pangolin CoVs, pose a risk of new zoonotic infection and thus should be studied further. (Lam *et al.*, 2020; Niu *et al.*, 2021; WHO, 2021; Xiao *et al.*, 2021; Gao *et al.*, 2022)

Guan et al. (2003; during the SARS-CoV-1 pandemic) isolated the causing virus from racoon dogs, showing the susceptibility of racoon dogs to the SARS-CoV-1. Noting the results of the mentioned study and the number of racoon dogs held captive in fur farms in China, Freuling et al. (2020) assumed that it is probable

[12] HeLa cells transfected with the mouse gene for ACE2 culture were not susceptible to SARS-CoV-2 *in vitro*, whereas HeLa cell lines presenting the ACE2 of humans, pigs, civets, or bats were.

that racoon dogs might be susceptible to SARS-CoV-2 as well, which was later proven. Considering that the infection was spreading among observed individuals and that racoon dogs were sold in the pandemic's epicentre in the HSM's Western side, racoon dogs might have been the intermediate host of SARS-CoV-2. However, all the evidence supporting this hypothesis is indirect (Guan *et al.*, 2003; Freuling *et al.*, 2020; Mallapaty, 2023)

The main argument against the zoonotic theory is the geographical gap between the geographical location of Yunnan horseshoe bats, the place of the first emergence of this infection, and several decades of evolutionary space between the known bat viruses and SARS-CoV-2. Furthermore, only a few identified bat viruses with large similarities to human SARS-CoV-2 can bind to human ACE2 receptors. Genetic and epidemiological studies suggest some species commonly farmed in Southeast Asia were a source of human infection rather than direct transmission from bats. But, no evidence of the repeated introduction of early SARS-CoV-2 strains into humans, apart from the original lineages A and B (see chapter 4.2.1.), was detected, suggesting a limited number of animal reservoirs. It is also possible that after gaining the ability to infect humans, the virus lost its ability to infect the original host (WHO, 2021; Worobey *et al.*, 2022)

3.2. Laboratory leak theory

This theory claims that SARS-CoV-2 might have found its way into the human population via the accidental infection of laboratory personnel or the escape of infected animal models. Since SARS-CoV-2 is frequently causing asymptomatic or mild disease, the infected person with asymptomatic or prodromal infection might spread the infection in a populated area of the city, such as wet markets. There are precedents for the escape of viral infections due to laboratory incidents, as shown by Parry (2004). For example, the outbreak of SARS-CoV-1 in early 2004 in Beijing, after the pandemic was mitigated, was most likely caused by breaching the biosafety procedures in the best Chinese virology laboratories. Moreover, SARS-CoV-2 has been proven to have infected a research assistant in a Taiwanese lab who worked with infected animals. Also, the simulation study on the influenza virus done by Merler et al. (2013) suggested that with a probability between 5 and 15%, it is likely that the escape event of a potential pandemic virus is not detected. (Parry, 2004; Merler *et al.*, 2013; Cyranoski, 2017b; Andersen *et al.*, 2020; Holmes *et al.*, 2021; Silver, 2022)

The laboratory escape hypothesis of SARS-CoV-2 origin is supported by the biosafety level 4[13] (BSL-4) laboratory in Wuhan, located about 30 km from the wet market where the first human cases occurred.

[13] Biosafety level (BSL) is used to identify protective measures in the laboratories set to protect the environment, staff and the public. There are four levels in total. Infectious agents or toxins that consistently cause disease in healthy adults are studied in BSL-1 labs. There is no requirement for special equipment. In BSL-2 labs, moderate-risk infectious agents are studied. Such agents pose a risk if they are accidentally swallowed, exposed to the skin or inhaled. There Must be washing sinks, eye-washing stations, and automatic closing and locking doors. Equipment for the decontamination of laboratory waste is also a necessity. Agents or toxins that might be transmitted through the air and cause potentially lethal infection if inhaled must be studied in BSL-3 labs. Biosafety cabinets with controlled air flow or sealed enclosures must be used when handling such agents. Interlocked doors, sealed windows, wall surfaces, and filtered ventilation systems must be found in these biosafety levels. BSL-4, the highest biosafety level, labs are where aerosol-transmitted infectious agents or toxins are hazardous. These agents cause life-threatening infections for which no therapy or vaccine is available. Access to such laboratories is carefully controlled, and there is a requirement for special training. (Public Helath Emergency, no date)

Additionally, this laboratory was moved to its current place shortly before the first human cases were identified. According to a WHO study, such a situation might pose a risk if the personnel do not abide by the strict regulation for handling dangerous materials. (WHO, 2021)

The laboratory most considered to be the culprit of the COVID-19 pandemic in case of a laboratory escape scenario is the Wuhan Institute of Virology (WIV). The construction of the mentioned laboratory was finished in 2014. However, certain discrepancies were pointed out by Western scientists about it. Most of them mentioned the previous escapes of SARS-CoV-1 from Beijing labs and the strict Chinese societal hierarchy decreasing the effectiveness of self-control mechanisms (such as paradigm diversity, speaking up, and openly available information, all uncommon in China). Furthermore, it is now proven that the National Institute of Health (NIH) of the USA provided the financial background for the gain-of-function[14] (GOF) research at the WIV. (Cyranoski, 2017a; Dance, 2021; Holmes *et al.*, 2021; Lerner *et al.*, 2021; WHO, 2021)

On the other hand, the work involving cell cultures and humanised model animals was not routinely performed at this laboratory. The research was focused on viral genomic sequencing, which poses a negligible threat of personnel infection. Although several tabloids (e.g., Daily Mail, New York Post) claim that humanised mice susceptible to SARS-CoV-2 were present and used for military research at WIV, no scientific evidence supports this. No pandemic so far was ever caused by the laboratory escape of any novel virus. In addition, none of the viruses listed in the documentation for the GOF experiments is related to SARS-CoV-2 enough to evolve into it. However, there are speculations about whether the GOF experiments may have played a role in the SARS-CoV-2 emergence as the virus research was performed on naturally infected bats. However, no evidence exists that WIV worked on SARS-CoV-2 or a prospective ancestor. Also, despite extensive contact tracing, it has been shown that the early cases had no link to the staff of WIV, and members of the suspected team of the WIV were all seronegative for SARS-CoV-2 specific antibodies when tested in March 2020. (Cyranoski, 2017a; Dance, 2021; Holmes *et al.*, 2021; Lerner *et al.*, 2021; WHO, 2021)

According to WHO, the virus carries no evidence suggesting its construction in the laboratory. However, as stated by Prof David Baltimore,[15] it is impossible to reveal the origin of some sequence just by analysing it. In the cited interview, he also confirms that the FCS sequence found in SARS-CoV-2 is absent across the whole taxonomy class, which he claims is not likely to happen naturally through the evolutionary processes. He also argues using the results of the experiment done by his team. They modified the spike protein of SARS-CoV-1 insertion of S1/S2 FCS, and whether this sequence was inserted by humans or evolution was indistinguishable. Also, the similarity of the FCS sequence of SARS-CoV-2 with the sequence of MERS-CoV was used as an argument supporting the hypothesis of laboratory origin caused by GOF coinfection experiments. The theory considering deliberate bioengineering of SARS-CoV-2 or its release is no longer considered (by WHO), even though some members of the scientific community still argue in favour of this theory. Intelligence reports (which in most cases consider laboratory leak theory quite probable), such as the

[14] It is not exactly defined what the gain-of-function research means so far. In the case of viruses, it usually means altering the genome for the virus, favouring the better infectivity and transmissibility of the pathogen.

[15] Professor of virology at Caltech, Nobel Prize laureate for his work in viral genetics

report of the FBI from early March 2023, do not show any scientific proof for such statements and hence are not considered among scientists. (Follis, York and Nunberg, 2006; Dajose, 2021; Kaina, 2021; Mecklin, 2021; WHO, 2021; Lenharo and Wolf, 2023; Rofer, 2023)

The lack of record of SARS-CoV-2, and its closely related viruses or genomes, discovered before December 2019, argue against the laboratory leak hypothesis. Additionally, the hypothesis considering viral escape via laboratory animals (i.e., mice) was ruled out as mice used as models at the WIV were not susceptible to infection by the early strains of SARS-CoV-2. (H. Zhou *et al.*, 2020; Wan *et al.*, 2020; WHO, 2021)

4. DIVERSITY OF SARS-COV-2

SARS-CoV-2, after its introduction to humans, evolved into different lineages and variants. A variant is defined as a virus that contains several mutations in its genome, while the viral lineage is a group of viruses which share a common ancestor (see Figure 5). A lineage or a group of lineages with analogous mutations can be classified as a variant of interest (VOI), a variant of concern (VOC), a variant of high consequence (VOHC), or a variant being monitored (VUM). This nomenclature is based on characteristics which can lead to changes in the behaviour of the virus and which need to be reflected in public health policies (CDC, 2020)

The Wuhan variant had undergone many evolutionary changes throughout the pandemic. The evolution of SARS-CoV-2 can be visualised in the phylogenetic tree (see Figure 6). Some changes arose consecutively, while others appeared simultaneously in different geographical areas during the pandemic. There are numerous SARS-CoV-2 variants, so only the most epidemiologically important ones will be presented in this thesis. The complete list of SARS-CoV-2 lineages and variants can be found at https://cov-lineages.org/lineage_list.html. As far as the up-to-date knowledge reaches, the most distant subvariant from the original Wuhan virus is variant BA.5.2.1.5.3 (or BF.5.3), a subvariant of an Omicron lineage. This variant has accumulated 114 nucleotide changes in its sequence compared to the original Wuhan strain (see Figure 6). ("https://nextstrain.org/nextclade/sars-cov-2" system created by Hadfield *et al.*, 2018; *Cov-Lineages*, no date)

Initially, there were two lineages identified, lineage A and B, that started the whole pandemic. Lineage B was first discovered on 24th December 2019, while lineage A was identified six days later[16]. The B haplotype is a predecessor of most variants emerging in the later stages of the pandemic, but some important variants have their origins also in the A lineage. The A lineage was detected in China, Japan, South Korea, Australia, the USA, and several European countries. (Rambaut *et al.*, 2020; O'Toole *et al.*, 2021; O'Toole, Scher and Rambaut, 2023)

For distinguishing the variants, different nomenclatures were used. WHO names the variants using the Greek alphabet. The UK developed their nomenclature system, which this thesis will not cover. Pango lineage

[16] The time that passed by between the identification of the two lineages might have been caused by data processing. (Rambaut *et al.*, 2020; O'Toole *et al.*, 2021)

dynamic nomenclature (PLN) was designed to identify epidemiologically relevant lineages of SARS-CoV-2, and it is the most used nomenclature today. (Rambaut *et al.*, 2020; 'Pango Network – Helping track the transmission and spread of SARS-CoV-2', 2022)

4.1. PANGO LINEAGES CLASSIFICATION SYSTEM

A group of SARS-CoV-2 genome sequences must meet a strict set of characteristics to be considered a new lineage. The official list of rules for assigning and naming Pango lineages can be viewed at https://www.pango.network/the-pango-nomenclature-system/statement-of-nomenclature-rules/. As lineage and variant are terms for a monophyletic and genetically defined group of viruses, both terms will be used synonymously in the respective part of the thesis.

The Pango system for naming SARS-CoV-2 genetic lineages is unique thanks to its ability to reflect the changes in the pandemic and the accumulation of new genomic data. Each new Pango lineage must share ancestry representing a branch of the evolutionary tree. Secondly, the appearance of lineages must be supported by epidemiological characteristics such as the emergence of the virus in a new geographical location. Each Pango lineage is assigned its unique alphanumerical code containing basic information about its position in the phylogenetic tree (see Figure 5). ('What are Pango lineages? – Pango Network', 2022)

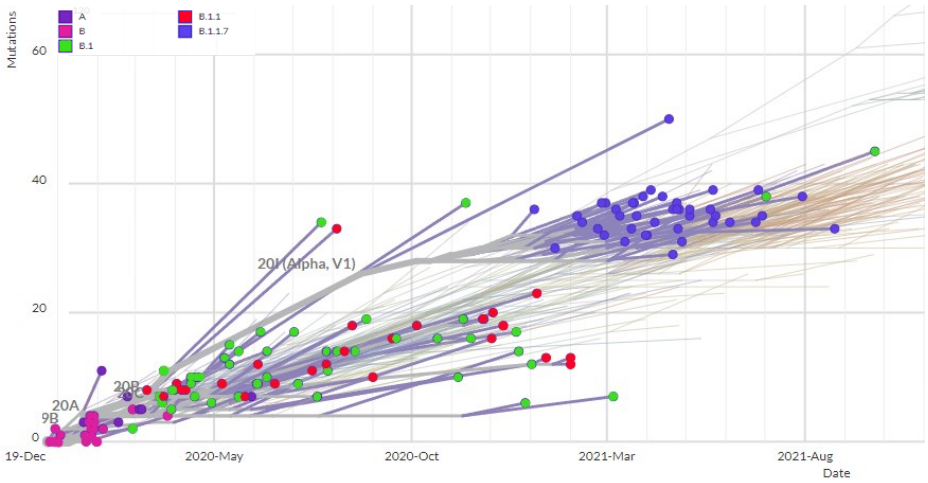

Figure 5: Phylogenetic tree of the Alpha variant (B.1.1.7; nextstrain.org; modified). In the picture the important predecessor variants are highlighted. As is visible from the plot Alpha variant acquired significant number of mutations that are not present in the hitherto variants, and all its progeny kept the vast majority of newly acquired mutations. To see the location of important mutations, see Figure 7. Alpha strain meet the virological definition of a lineage while each dot in the plot represents its own variant.

The name in the Pango system consists of alphabetical characters representing parental lineage and the numerical suffix formed from one or more numbers separated by dots. Each dot represents "being a descendant of". Thus, in the case of the B.1.1.28.1 variant, also known as P.1 or gamma, the first number 1 stands for the first described lineage of B linage, and a similar situation comes with the second number 1. Therefore, the partial name "B.1.1.28" means the 28th descendant lineage of B.1.1. ('Pango lineage names – Pango Network', 2022)

As the pandemic continued, the names of the newly formed lineages became so long and complicated that aliases were introduced. When there are more than three numbers in a row in the suffix, the letter changes to the first following in the alphabet that has not been assigned to any variant. For example, C.1 is an alias for B.1.1.1.1, meaning that the C and B.1.1.1 lineage are identical. After the letter Z is used for the alphabetical prefix, the newly discovered lineages would get the prefixes AA, AB, …, and ZZ, from which the names

would go back in the alphabet to form the AAA alias.[17] However, a rule prevents the usage of letters I, O and X as I and O, which might confuse numbers 1 and 0, and the letter X is reserved for recombinants of SARS-CoV-2's variants. ('Pango lineage names – Pango Network', 2022)

There are two exceptional cases among the lineages of SARS-CoV-2. One of these categories is reserved for lineages A and B. In their case, we do not know the root of their phylogeny with certainty, so their ancestry remains ambiguous. ('Statement of Nomenclature Rules – Pango Network', 2022)

The second special case consists of recombinant variants. That means, by definition, that those are lineages that have more than one parental lineage. The naming of recombinants remains similar to using aliases in the newly described lineages. The first known recombinant got the name XA, followed by XB. The terms of recombinant variants need more information on their parental lineages, which must be found in the variant description list. ('Statement of Nomenclature Rules – Pango Network', 2022)

These special case lineages often lack the numerical suffix, as the prefix of each lineage stands for a single unequivocal ancestor. However, if a non-recombinant descendant appears, the numerical suffix is again obligatory. Thus, the first non-recombinant descendent of recombinant lineage XB is XB.1. Rules for creating aliases also apply in case of need. This means that XA.1.1.1.1 would be AJ.1 if the following available top-level prefix would be AJ. ('Statement of Nomenclature Rules – Pango Network', 2022)

4.2. Variants of the most significant epidemiological concern

4.2.1. Original Wuhan virus

The first cases of SARS-CoV-2 are dated to mid-November 2019, although the first official report was issued at the beginning of December 2019. When the Wuhan variant is mentioned, two lineages (according to PLN) are meant – A and B. Genomic studies concerning the genomes known until late January have divided the SARS-CoV-2 into several haplogroups[18] – A, B, B1, B1a, B2 & B4. During the first weeks of the pandemic, the A clade was predominant worldwide. However, it was out-competed by the B clade (further covered in Chapter 4.2.2.1.), giving rise to most later variants. One of the distinctions when this variant is compared to its successors is the age of patients with severe disease symptoms. The Wuhan variant targeted mainly elderly people, while some later variants infected a similar proportion of 20- and 60-year-olds. Although the average age of patients requiring hospitalisation stayed quite high during the occurrence and dominance of later variants, it was significantly lower than the mean of 60 years in the case of the Wuhan strain. (Gómez-Carballa *et al.*, 2020; X. Yang *et al.*, 2020)

[17] The last used alias up to 14th April 2023 is FJ in FJ.1 (B.1.1.529.2.75.3.4.1.1.1.1.19.1) variant discovered in the UK on 18th December 2022; overall through the pandemic, 3056 PANGO lineages were described. (O'Toole, Scher and Rambaut, 2023)

[18] Haplogroups A and B are homological terms for the variants A and B in the Pango nomenclature. Haplogroups B1, B1a etc., are temporary denominations and will not be covered further in the thesis. (Gómez-Carballa *et al.*, 2020; Y. Yang *et al.*, 2020)

4.2.1.1. B.1

B.1 variant is known to have two important mutations that enabled it to out-compete previously predominant variants. These mutations are D614G in the S protein and P314L (see Figure 6) in ORF1b in the segment encoding nsp12 (RdRp). When cells were infected with SARS-CoV-2 with D614G mutation *in vitro,* the titter levels were significantly higher in the cells infected by the mutant virus. This mutation also increased the replication efficacy in the upper respiratory pathway *in vitro* (primary human airway tissue cultures) and *in vivo* (Syrian golden hamster model). (Ogawa *et al.*, 2020; Plante *et al.*, 2021; Archana, Long and Chandran, 2022; Carroll *et al.*, 2022; Gangavarapu *et al.*, 2023; Goldswain *et al.*, 2023)

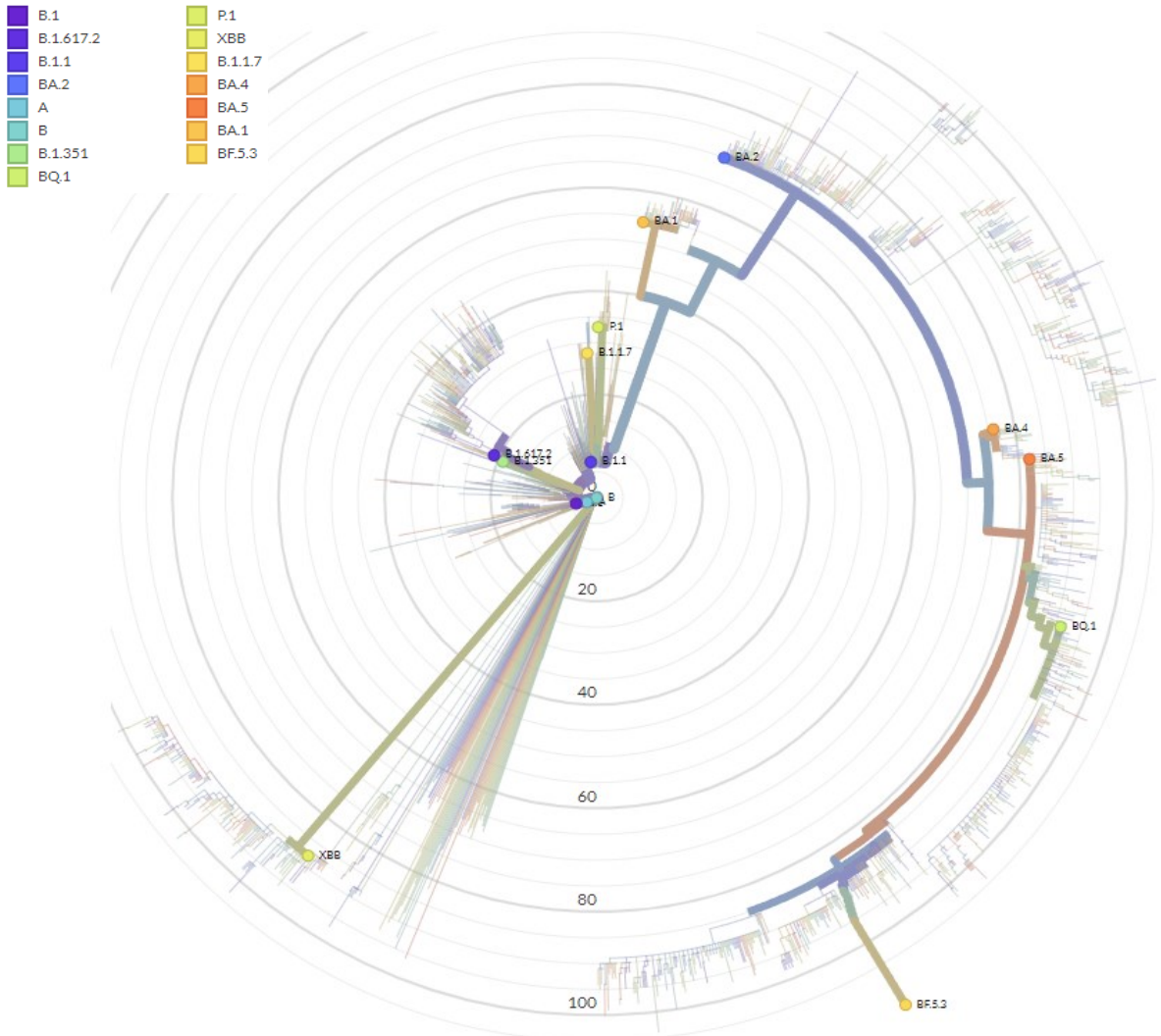

Figure 6: Divergence plot of mentioned SARS-CoV-2 variants (nexstrain.org). All variants are labelled by their PLN name. Each circle stands for 5 acquired mutations.

4.2.1.2. B.1.1

As shown in Figure 7, there are no significant changes in sequence between B.1 and B.1.1 variants in either the S protein or the non-structural proteins. However, important changes occurred in the sequence of the N protein, specifically in positions R203K and G204R. These alterations streamline the ribonucleocapsid assembly, giving the B.1.1 replication advantage compared to the previous B.1 variant. Also, the increase of the infectivity in the lung cells was documented. (Wu *et al.*, 2021; Gangavarapu *et al.*, 2023)

4.2.2. ALPHA

An important VOC first identified in the UK in September 2020 was assigned the name B.1.1.7 in PLN, and it constitutes the clade Alpha in the nomenclature of WHO and its subvariants. This variant shows several differences in the sequence of the S gene as compared to the Wuhan strain (see Figure 7A). These differences caused its selection superiority compared to previous variants, leading to its predominance in late 2020 and early 2021 (in the UK). (Davies *et al.*, 2021; Walker *et al.*, 2021; *Cov-Lineages*, 2023)

The N501Y (in S) substitution of this variant increased the affinity of SARS-CoV-2 to the ACE2 receptor, and H69del/V70del was most likely responsible for the immune evasion of the virus as it changed the structure of the epitope (part of the antigen to which antibodies attach themselves). The third of the most critical changes was the P681H mutation, potentially facilitating cell entry as it is part of the FCS. The mentioned changes in the genome also probably affected the severity of the disease caused by SARS-CoV-2 since the mortality rate was higher in the population infected with this variant. (Challen *et al.*, 2021; Walker *et al.*, 2021; Grint *et al.*, 2022)

4.2.3. BETA

In mid-December 2020, a new variant of concern assigned PLN B.1.351 and named Beta by WHO appeared in South Africa. This lineage has spread globally rapidly. The N501Y mutation likely causes

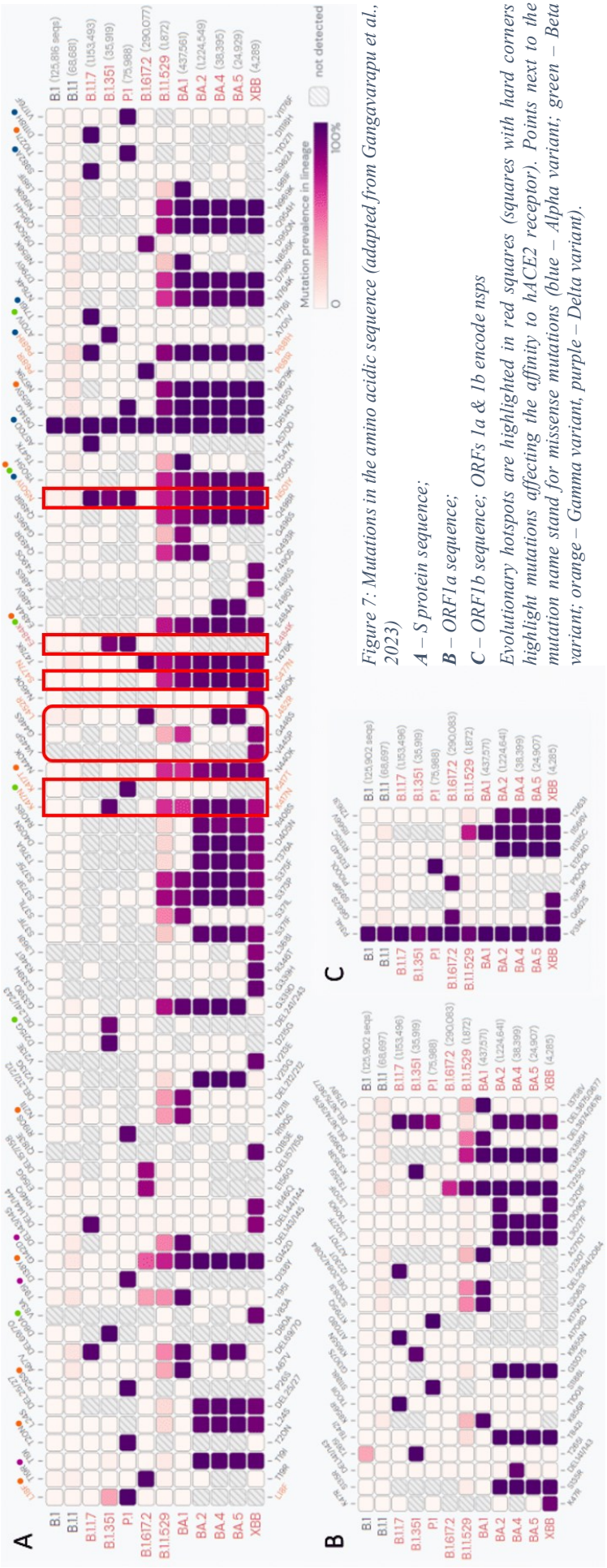

Figure 7: Mutations in the amino acidic sequence (adapted from Gangavarapu et al., 2023)
A *– S protein sequence;*
B *– ORF1a sequence;*
C *– ORF1b sequence; ORFs 1a & 1b encode nsps*
Evolutionary hotspots are highlighted in red squares (squares with hard corners highlight mutations affecting the affinity to hACE2 receptor). Points next to the mutation name stand for missense mutations (blue – Alpha variant; green – Beta variant; orange – Gamma variant, purple – Delta variant).

the increased transmission, a characteristic shared with the Alpha variant. (Mwenda *et al.*, 2021; Tegally *et al.*, 2021)

N501Y, K417N, and E484K substitutions affect RBD residues of key importance (all mutations can be seen in Figure 7). Besides substitutions, the deletion of three amino acids in positions 242-244 in the N-terminal domain (NTD) was discovered[19]. The combination of changes in positions 417, 484 and 501 is crucial in lowering the affinity of neutralising Abs, as they are localised in their binding regions. The E484K substitution also increases the binding affinity to the ACE2 receptor and thus increases the selection advantage of B.1.351. (Starr *et al.*, 2020; Radvak *et al.*, 2021; Tegally *et al.*, 2021; Wang *et al.*, 2021)

4.2.4. GAMMA

In November 2020, a new VOC of SARS-CoV-2 emerged in Brazil. This variant was denoted as P.1 by Pango nomenclature, and WHO assigned the Greek letter Gamma for its description. This variant has spread rapidly in Brazil. (Faria *et al.*, 2021; Prete *et al.*, 2022)

P.1 variant of SARS-CoV-2 carries several mutations (see Figure 7), some of which have been previously considered important in other variants. Compared to its direct ancestor – variant B.1.1.28, seventeen changes of amino acids increasing the affinity to the hACE2 receptor were detected. Ten of these amino acid substitutions are in the S protein, three deletions, four synonymous mutations and four insertions. The three crucial mutations in the P.1 variant are K417T, E484K, and N501Y. The same three mutations were also present in the B.1.351 VOC, and N501Y was also detected in the B.1.1.7. The independent appearance of variants with similar changes in such geographically distant places is an example of convergent evolution. (Faria *et al.*, 2021)

4.2.5. DELTA

Among the most breakthrough VOCs was the Delta variant, B.1.617.2, in PLN nomenclature. It was first identified at the end of 2020 in West India. It spread rapidly across the Indian subcontinent and outcompeted previously predominating variants – B.1.1.7 and B.1.617.1 (Kappa). Pre-existing neutralising antibodies were six times less effective against Delta than against the original Wuhan strain. Vaccination-elicited antibodies were *in vitro* eight times less effective. Breaching the vaccine-elicited immunity was reported for all three vaccines used in the EU (vaccines from Pfizer, Moderna and AstraZeneca) at that time. (Ferreira *et al.*, 2021; Mlcochova *et al.*, 2021; Zhang *et al.*, 2022)

The S protein of the delta variant contains several important mutations. Two crucial mutations are found in the RBD – L452R, T478K. The latter, together with D614G substitution in the S1, enhanced the binding of the ACE2 receptor. L452R increased the overall stability of the spike protein and decreased the effectivity of binding of some monoclonal antibodies (mAbs) used for treatment. The more stable S protein leads to a more effective membrane connection and thereby promotion of transmission through more effective cell entry. (Mlcochova *et al.*, 2021; Saito *et al.*, 2022; Zhang *et al.*, 2022)

[19] This area is complicated to align correctly, so the deletion might be in positions 241-243.

Compared with other variants, B.1.617.2 has much higher fusogenicity thanks to the P681R mutation, which enhances S protein cleavage. This variant creates over three times larger syncytia than the B.1.1 variant. Most of the other important mutations introduced by the Delta variant are located within the NTD sequence, and it was suggested that these mutations enhance the viral escape from the immune system (T19R, G142D, Δ156-7, R158G, and D950N). (Mlcochova *et al.*, 2021; Saito *et al.*, 2022; Zhang *et al.*, 2022)

4.2.6. OMICRON

Omicron is the last VOC to appear so far. The PLN name for the Omicron variant is B.1.1.529, and its successors are BA.1, BA.2, etc. It was identified in mid-November 2021 in the South African region. The Omicron variant spread very fast worldwide and gained dominance. (Cui *et al.*, 2022; *Cov-Lineages*, 2023)

The number of mutations in the B.1.1.529 variant is unprecedented (see Figure 7). There are 37 mutations in the sequence of the spike protein, fifteen of which are located within the RBD. Six amino acid residues are deleted, and one is inserted. These substitutions affect the ability of pre-existing antibodies to neutralise it. (Cameroni et al., 2022; Cui et al., 2022; Liu et al., 2022; Meng et al., 2022; Peacock et al., 2022[20])

On the other hand, the S protein sequence changes made Omicron variants significantly less reliant on TMPRSS2-dependent cell entry. Due to the differences in the architecture of cells in the upper and lower respiratory pathways, the endosomal pathway leads to a preference for upper respiratory tract cell infection as the cells in the upper respiratory tract do not have the TMPRSS2 on their surface. Infection of these cells is linked to less severe disease. (Cameroni et al., 2022; Cui et al., 2022; Liu et al., 2022; Meng et al., 2022; Peacock et al., 2022)

4.2.6.1. Omicron descendants

BA.1 (B.1.1.529.1) was detected in Botswana and South Africa at the start of November 2021 and showed 30 amino acid substitutions compared to the Wuhan-Hu-1 reference sequence. (Shrestha *et al.*, 2022)

BA.2 was discovered nearly simultaneously with BA.1. The difference between BA.2 and its sister lineage, BA.1, is an example of divergent evolution since they split their evolutionary paths from their last common ancestor. The divergence between these lineages reaches 21 differences in amino-acidic sequence, while 8 of these changes are mutations occurring in BA.2 but not in BA.1 and vice versa. Surprisingly, the differences between these two variants are almost as big as between Alpha, Gamma, and Delta. During the first months of 2022, in South Africa, the BA.2 outgrew the BA.1 variant (Shrestha *et al.*, 2022)

Omicron variants BA.4 and BA.5 are quite similar. Their most recent common ancestor is estimated to originate in mid-November 2021. The spike protein sequence of both variants is closely related to the BA.2 variant. In South Africa, these new variants outcompeted hitherto dominant BA.2. (Tegally *et al.*, 2021; Malato *et al.*, 2022; Shrestha *et al.*, 2022)

[20] Preprint

Similarly to other Omicron lineages, BA.4 and BA.5 are capable of immunity evasion. Compared to BA.2, BA.4 and BA.5 exhibit deletions in positions 69 and 70 (these can also be found in the Alpha variant and BA.1), substitutions L452R (also present in Delta variant) and F486V, and restoration of wild-type Glutamine on position 493. Further mutations can be found in the nsp1 and nsp8. It was proven that BA.2, BA.4, and BA.5 could evade the immunity elicited by the prior Omicron variants. (Tegally *et al.*, 2021; Malato *et al.*, 2022; Nadig *et al.*, 2022[21]; Shrestha *et al.*, 2022)

4.2.7. XBB

XBB is a recombinant variant of BJ.1 and BM.1.1.1 Omicron variants (see Figure 8). It was first identified in the USA and Singapore in December 2022. The XBB variant predominated in the first half of 2023in South and Southeast Asia and Peru. In the United States, the prevalence was above 10%, similar to Europe. XBB shows a level of infectivity significantly higher than other variants identified before. Only the infectivity of the BQ.1 (descendant of the BA.5 variant) omicron variant is comparable. The strategy for the immune evasion used by XBB might be linked to the smaller surface of the S protein compared to BA.2. But it is also suggested that the changes in the S protein sequence alone might be causing the immune evasion. (Arora et al., 2023; Tamura et al., 2022[22])

XBB descendant – XBB.1.5 (a. k. a "Kraken") subvariant – grew in numbers in early 2023. At the end of January, it caused nearly 50% of all confirmed cases in the USA. The main difference between the XBB.1.5 subvariant and its parental variant is the possession of the F486P mutation. This mutation seems to improve the binding of the virus to the hACE2 receptor enabling the predominance of Kraken subvariant. All the XBB subvariants that circulate today have convergently acquired the F486P substitution. (Callaway, 2023; Parums, 2023)

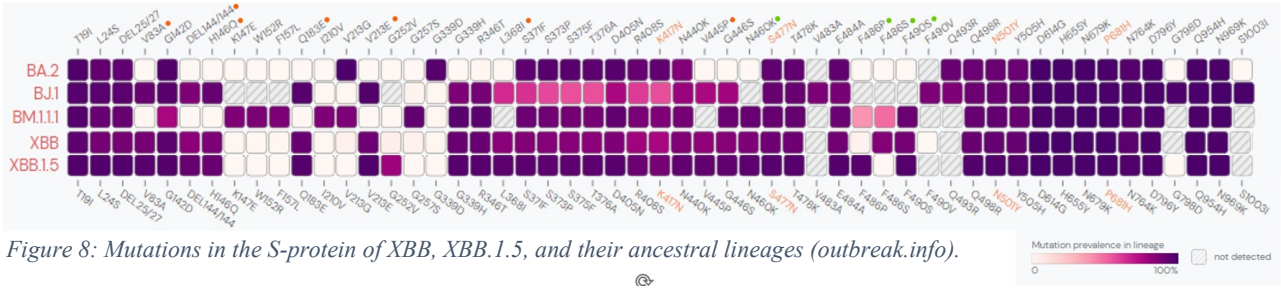

Figure 8: Mutations in the S-protein of XBB, XBB.1.5, and their ancestral lineages (outbreak.info).

BJ.1 variant is a descendant of BA.2 and later recombined with BM.1.1.1, resulting in XBB receiving specific mutations from each parental lineage. Mutations highlighted by the orange dot XBB acquired from the BJ.1 and mutations acquired from the BM.1.1.1 are highlighted by the green dot.

5. TRENDS OBSERVED IN THE SARS-COV-2 EVOLUTION

As can be assumed from the basics of the evolutionary biology curriculum, the most common changes observed in evolution are the changes of characteristics with significant effects on the ability to reproduce and the likelihood of survival – fitness (or selection) advantages. Looking back in history, only a few lethal viral infections circulate in the population to this day (such as HIV). In most cases, the evolution of the virus to

[21] Preprint
[22] Preprint

cause less severe disease is observable as it is advantageous for its propagation. Spanish flu, for example, struck Europe right at the end of World War I and, until 1920, took millions of lives. After 1920, however, it became a much less severe and seasonal infection. This can be explained by a transmission-virulence trade-off phenomenon (see Equation 1). In the case of RNA viruses, a quick decline in virulence is observed shortly after the infection outbreak. (Anderson and May, 1979; May and Anderson, 1979; Jedwab, Johnson and Koyama, 2019; Telenti *et al.*, 2021; Xu *et al.*, 2023)

$$R = \frac{\beta}{\mu + \alpha + \gamma}$$

Equation 1: Equation of transmission-virulence trade-off (adapted from Xu et al., 2023)

***R** = fitness; **β** = transmission rate; **μ** = host mortality; **α** = host disease-induced mortality; **γ** = recovery rate*

Viruses lose their transmission vector if the host dies or is isolated because of severe symptoms. Hence, these viruses will not reproduce. The sublineages of the Omicron variant are significantly less virulent than previous variants or the original Wuhan strain of the SARS-Cov-2. However, this characteristic was only observable after the emergence of the Omicron variant. (Anderson and May, 1979; May and Anderson, 1979; Jedwab, Johnson and Koyama, 2019; Xu *et al.*, 2023)

Many trends in the evolution of the SARS-CoV-2 accelerated significantly when the Omicron VOC and its filial variants started to occur. Among else, the quickly accelerating convergent evolution of the RBD. The newly acquired mutations significantly affected the tropism of the novel human coronavirus, and the newly occurring lineages were able to evade the immunity of their hosts. Omicron variants BA.2 and BA.5 were predominant co-circulating variants for a considerable time. When antibodies elicited by the infection of these variants were studied by Cao et al. in 2023, it was shown that breakthrough infections rarely induce the creation of new Abs specific to the respective variant. In the case of post-vaccination infections, the variants BA.2 and BA.5 are recognised by the immune cells bearing pathogen-recognising receptors elicited by the vaccine based on the WT virus (yet, the same effect might be caused by previous infection). This phenomenon is also known as immune imprinting – vaccination or infection with another viral strain induces a sort of boost of antibodies formed to combat some previously circulating strain. Immune imprinting often helps the immunity to fight off known diseases faster, as substantial changes within the pathogen may otherwise prolong the recovery from the disease. (St. John and Rathore, 2019; Wheatley *et al.*, 2021; Cao *et al.*, 2023)

Antibody-dependent enhancement, a phenomenon when the antibodies developed in response to infection of different variants do not neutralise the variant currently present in the host efficiently and potentiate the disease severity, was considered in SARS-CoV-2. Such effect is documented in secondary infection by Dengue virus (DENV) of a different serotype than the primary infection. Fortunately, antibody-dependent enhancement in the case of SARS-CoV-2 was documented only during *in vitro* experiments and has not been shown *in vivo* so far. (St. John and Rathore, 2019; Gan *et al.*, 2022)

Within the receptor-binding domain, some evolutionary hotspots can be found. These places in the amino acid sequence are prone to point mutations more than others. The hotspots are highlighted in Figure 7. Some hotspots occur only in the Omicron variants, partly due to the accelerated mutation rate (see Figure 7). One of the mutations not present in the S-protein, yet found in all VOCs throughout the pandemic, is the P314L

alteration in the ORF1b gene encoding RdRp. This alteration was convergently conserved among the variants. (Barton *et al.*, 2021; Hossain *et al.*, 2021; Chakraborty *et al.*, 2022; Cao *et al.*, 2023)

FCS of the SARS-CoV-2, similarly to FCS of other Betacoronaviruses, is adaptable, and the coding sequence has been optimised throughout the evolution in VOCs. Mutation P681H (P681R present in Delta variant enhances the S protein cleavage and fusogenicity), located within the sequence of the FCS present in Alpha and Omicron variants, seems to be linked to prolonged infections. The N679K mutation found in Omicron is located outside the core FCS domain; however, it is part of the amino acid sequence of the furin cleavage (673-687) site and is crucial in constructing the furin-binding pocket. (Tian, Huajun and Wu, 2012; Whittaker, 2021; Lubinski, Jaimes and Whittaker, 2022; Lubinski and Whittaker, 2023; UniProtKB, 2023b)

The effect and nature of the SARS-CoV-2 mutations differ, with more than half being missense mutations and around 35% synonymous ones. The mutation rate during the first year of the pandemic was 8×10^{-4} nucleotides per genome per year. Before the emergence of the Delta strain, the variants Alpha, Beta, and Gamma could not over-compete one another. Their similar fitness and co-circulation might have caused this inability to gain dominance over other strains in mostly naïve populations. (Rahimi, Mirzazadeh and Tavakolpour, 2021; Chakraborty *et al.*, 2022; Chen *et al.*, 2023)

At the time of the co-circulation of Alpha, Beta and Gamma strains were, European countries mostly infected by the Alpha strain then. However, some cases of variants were also detected in the population. Although the Gamma variant was not effectively recognised by the immunity elicited by prior infection with the Alpha variant, it could not over-compete it due to its lower relative transmissibility. (Nasif *et al.*, 2022; Stefanelli *et al.*, 2022)

Another trend observable in the evolution of the SARS-CoV-2 during the pandemic of COVID-19 is the overall reduction of the genome size and the increase in the abundance of uracil (U) in the sequence of the RNA genome. This topic was deeply studied by Wang et al. in their study from 2022; however, it excludes Omicron variants as they emerged shortly after the paper was sent to the publisher. All VOCs included in the study (Alpha-Delta) have shown a high mutation rate; on average, 13 to 15 amino acids[23] change in the lineage sequence after seven to twelve months of circulation. The number of cytosines (C) decreased in all the variants observed. In contrast, the number of U has risen, increasing the abundance of amino acids encoded by U-rich codons. On the other hand, as C-rich codons decrease, so do the respective amino acids in abundance. Mutations of the C, guanine (G), and adenine (A) take place regardless of the position in a codon, even if it would cause the formation of a premature STOP codon. Results for U are ambiguous. As the A-U pairing creates only two hydrogen bonds, compared to three hydrogen bonds of the C-G pair, the secondary structure of the genome with U→C mutations is less stable. These sequential changes reduced the overall stability by up to five per cent at the 5'-untranslated region (UTR). The less stable nucleic acid is easier to unfold, facilitating the replication process. The biggest number of U nucleotides that Wang et al. found in their study

[23] This number of mutations occurs within one variant (diversification from parental lineage to all its subvariants). The emergence of a new VOC usually introduces many new mutations.

was 9,591 in the sequence of 29,769 nucleotides in length. This amount is 31 higher than in the reference sequence for SARS-CoV-2. These changes resulted in a decrease in the stability of the 5' UTR of the genome – such changes in this region are crucial for the interaction with the ribosome, resulting in higher translation efficiency. (Hinnebusch, Ivanov and Sonenberg, 2016; Masone, Alvarez and Polo, 2022; Wang *et al.*, 2022)

During the pandemic, the proportion of deletions has increased. Venkatakrishnan et al. have performed an analysis of more than 2.1 million genome sequences of SARS-CoV-2. Most mutations were substitutions (more than 95%), while deletions posed only 4.3% of all mutations. In total, 92 mutations were associated with surges (sudden increases in the number of cases), 18 being deletions (19.5%), significantly more than could be expected regarding the low proportion of mutations being deletions. The 18 deletions mentioned earlier, such as ΔH69-V70 and ΔY144 (see Figure 7), are located in the NTD of the S protein. The evolution of the highly transmissible variants is likely associated with deletions in the NTD and substitutions in segments outside of NTD but critical for the function of the S protein (such as L452R). There seems to be a link between deletions and immunity throughout the population. Since the massive vaccination started (January 2021), deleted regions have expanded greatly. As a significant proportion of the world population has been infected before the vaccination, the combination of vaccine and an infection-elicited immune response is suggested as the likely selection pressure. (Venkatakrishnan *et al.*, 2023)

Seven regions within the NTD have shown an increasing deletion rate after January 2021. Recurrent deletion regions (RDRs) are present at positions 14-18, 62-77, 136-147, 149-159, 210-211, 242-244, and 256-260 in the amino acid sequence. Most RDRs identified are close to the antigenic epitope – the segment of amino acid sequence that NTD-targeting Abs can identify. It seems likely that the expansion of several deleted regions in this segment of the S protein sequence is the attempt to discard the residues the host can detect in the sequence, hence evolving so-called antigenic minimalism. Identifying newly acquired deleted regions and continuing their monitoring is important due to their link to sudden epidemic surges. (Venkatakrishnan *et al.*, 2023)

Missense mutations, although less common than synonymous substitutions, altered the structure of the SARS-CoV-2 significantly. Variants Alpha and Beta introduced six missense mutations in their S protein sequence, ten such mutations were found in the Gamma variant, and Delta had four. All missense mutations are highlighted in Figure 7. Physiochemical analysis by Mahmood *et al.*, published in 2022, found that the spike's molecular weight has risen by nearly 300 g/mol (comparing the original Wuhan strain and Delta variant) due to the missense mutations. Furthermore, the charge of the spike has changed its polarity from negative to positive in the neutral pH. This change in charge resulted in more efficient immune evasion of the virus through heparan sulphate proteoglycans (HSPGs). Most of the missense mutations had stabilising effects on the structure of the S-protein. Missense mutations within the RBD (such as L452R and T478K) have increased the effectivity of the binding between the virus and the hACE2 receptor. (Rambaut *et al.*, 2020; Walls *et al.*, 2020; Klinakis, Cournia and Rampias, 2021; Mahmood *et al.*, 2022; Zhang, Zhang and Wang, 2022)

Main proteinase (M^{pro}, nsp5, $3CL^{pro}$) is a key enzyme for the replication process of SARS-CoV-2. It cleaves the pp1a/pp1ab together with papain-like proteinase (nsp3). Nsp5 is responsible for the cleavage at 11 cleavage sites of pp1a/pp1ab producing 13 end products. The M^{pro} is irreplaceable for the maturation of SARS-CoV-2's replicase and inhibits the hosts' immune response pathway. As shown in the study by Diessner *et al.* from 2023, mutations in M^{pro} have several effects. The increase in the surface area and the increase in hydrophobicity of M^{pro} are observable throughout time. In the dimeric state, M^{pro} in VOCs tends to be more flexible in the active site with time. For example, mutations in the M^{pro} are K90R (in Beta) and P132H (in Omicron). Increased flexibility of the M^{pro} leads to further stabilisation of the enzyme-substrate complex. (Ziebuhr, 2005; Graham *et al.*, 2008; Ullrich *et al.*, 2022; Diessner *et al.*, 2023; Rocha *et al.*, no date)

Recombination is highly difficult to detect as it requires whole genome sequencing (WGS) of large genome sets to be discovered. One of the ways of possible identification of recombination is the reversion of deletions, as their reversion is highly unlikely through any other mechanism. This may result in sort of "rescuing" genomes with deletions that would be disadvantageous. Recombinations of bat and pangolin CoVs were even considered a way of introducing the whole RBD. However, a later study from Hassanin, Rambaud and Klein from 2022 has shown no direct link between the RBD and pangolin CoVs. The first example of documented recombination has come with the Alpha VOC. Recombination events often change a variant so vastly that the prime immune system cannot recognise the virus. Therefore, the virus propagates more easily in the host and has a higher transmission rate. Nevertheless, in general, most recombinant viruses are incapable of effective infection and propagation. As the first successful recombinant variant was XBB, recombination is not considered that important evolutionary trend in the case of SARS-CoV-2. (X. Li *et al.*, 2020; Jackson *et al.*, 2021; Focosi and Maggi, 2022; Hassanin, Rambaud and Klein, 2022; Lytras *et al.*, 2022; Turakhia *et al.*, 2022)

Questions regarding the risk of SARS-CoV-2 acquiring new mutations, possibly leading to a new selection-advantageous variant in immunocompromised patients, were raised during the pandemic. Studies performed on HIV patients have not indicated that coinfection with HIV would be responsible for persisting replication which would potentially pose a risk of selection of a new, more severe variant. Several studies focused on the persistence of SARS-CoV-2 in immunocompromised patients were published. A study by Avanzato *et al.* from 2020 was performed with a 71-year-old patient suffering from lymphocytic leukaemia and acquired hypogammaglobulinemia. Infectious viral particles were formed as late as 70 days post-initial infection, and subgenomic RNAs were detected for over another month. Multiple swabs were collected from the patient and were deeply analysed. Although the viral replication kinetics remained unchanged compared to observations of the then-circulating virus in the rest of the population, within-host evolution was observed. Similarly to the evolution of the virus throughout the pandemic, most of the changes occurred within the sequence of the S protein. Changes in the S protein, however, were achieved mostly by deletions. These deletions provided the S protein of viruses within this host with higher flexibility. Although the within host-diversity was observed, the predominant SARS-CoV-2 population remained identical for the whole period. Choi *et al.* (2020) studied a 45-year-old male suffering from severe antiphospholipid syndrome and diffuse

alveolar haemorrhage. In this patient, despite the antiviral treatment (Remdesivir), the viral titer continuously grew, and the patient was positive for the viral RNA until his death 154 days after the first positive test. Overall, nine different virus sequences were isolated from the patient throughout the infection. The majority of differences in the genome sequences were non-synonymous mutations. Most substitutions were observed in the S protein, but changes in ORF1a, RdRp, E and N were also observed. Immunocompromised patients are at higher risk of persisting infection, and SARS-CoV-2 has been shown to acquire significant diversity within such hosts. Although cases of selection-advantageous variants originating in immunocompromised patients were not reported, their infection with SARS-CoV-2 should remain observed as it is capable of significant changes during their long-term infection. (Avanzato *et al.*, 2020; Choi *et al.*, 2020; Corey *et al.*, 2021; Karim *et al.*, 2021; DeWolf *et al.*, 2022)

Several animal species are prone to SARS-CoV-2 infection. However, only those that can be infected persistently by the SARS-CoV-2 and the virus is also capable of circulation and evolution in them will be further discussed, as they pose a risk of viral reintroduction to humans. One of them is the white-tailed deer (*Odocoileus virginianus*). Chandler *et al.* from 2021 detected SARS-CoV-2-specific antibodies in 40% of samples, and it has been shown that the virus has a high affinity to its ACE2 receptor. Furthermore, these animals live in herds and transmission between these animals has been proven and as well as the evolution of the virus while circulating in white-tailed deer herds. Importantly, VOCs rarely diagnosed today are still circulating among white-tail deer, constituting the risk of introducing changed viruses to humans. (Chandler *et al.*, 2021; Palmer *et al.*, 2021; Telenti *et al.*, 2021; Caserta *et al.*, 2023)

Another animal species that can be infected by SARS-CoV-2 is *Neovison vison*. Minks are the only known species from which SARS-CoV-2 was reportedly transmitted back to humans. Such events were detected at several mink farms, and nucleotide changes, some of which caused amino acid substitutions, were observed when minks were sampled repeatedly. Hammer *et al.*, in their study from 2021, have analysed the SARS-CoV-2 genomes present in swabs of minks and their infected caretakers. Changes in mink samples were later discovered in the swabs collected from the farms' personnel. Such findings have raised awareness of the spillover events, which could result in the introduction of significantly changed SARS-CoV-2 back into the human population. As many mink farms in China mainly supply the fur market. Therefore, they might have been considered another potential intermediate host of SARS-CoV-2. (Devaux *et al.*, 2021; Fenollar *et al.*, 2021; Hammer *et al.*, 2021; Sharun *et al.*, 2021; Zhou and Shi, 2021)

6. Conclusion

SARS-CoV-2, the virus responsible for the pandemic that humanity has been dealing with from the end of 2019 until now, is a +ssRNA virus from the family *Coronaviridae*. It has several unique characteristics, distinguishing it from other family members, such as the FCS, with a unique amino acid sequence making the virus more efficient in entering the host cells.

The origin of this novel virus remains unclear and may never be known for certain. Although multiple investigations were carried out regarding its genesis and introduction in the human population, we still need

unambiguous proof to confirm either zoonotic emergence or laboratory leak theory. Yet, among the scientific public, the theory of zoonotic introduction from bats through pangolin in humans is considered more likely.

SARS-CoV-2 has acquired many mutations, diversifying it into many lineages and variants. Some of these viral strains had characteristics which resulted in increased transmission or disease severity and were listed as variants of concern. Alpha, Beta, Gamma, Delta, and Omicron variants, including filial lineages like XBB and XBB.1.5, are the most important. In the evolution of the SARS-CoV-2, important trends were observed, such as the genome size reduction, lowering the stability in the 5'UTR, high mutation rates in the NTD and RBD of the spike protein and many other changes altering its ability to enter the cells and evade the host's immune response. The virulence of the Omicron variant is significantly lower than that of the preceding variants. However, its ability to spread in the naïve population is higher than its predecessors. It was proven that SARS-CoV-2 could infect many animal species, circulate in their population and, in the case of minks, can also evolve. Infected animals may pose a risk of reintroduction of altered variants in the human population that could potentially evade our immunity. Immunocompromised patients show high within-host diversity of viral genomes, and e significantly altered viruses could form during persistent infection of such patients. Furthermore, if strains circulating among humans and animals would recombine with one another, as happens with influenza viruses, the risk of creating potential pandemic viruses rises.

SARS-CoV-2 and the pandemic provided a unique opportunity for scientists to observe the evolution of the new virus in humans in real time. The number of lives the pandemic took is enormous, and the impact on life worldwide was immense. It is desirable to gain from this experience the ability to prepare better and react when a new pandemic virus eventually appears in the upcoming years and decades.

BIBLIOGRAPHY

Andersen, K.G. *et al.* (2020) 'The proximal origin of SARS-CoV-2', *Nature Medicine*, 26(4), pp. 450–452. Available at: https://doi.org/10.1038/s41591-020-0820-9.

Anderson, R.M. and May, R.M. (1979) 'Population biology of infectious diseases: Part I', *Nature*, 280(5721), pp. 361–367. Available at: https://doi.org/10.1038/280361a0.

Annan, A. *et al.* (2013) 'Human Betacoronavirus 2c EMC/2012–related Viruses in Bats, Ghana and Europe', *Emerging Infectious Diseases*, 19(3), pp. 456–459. Available at: https://doi.org/10.3201/eid1903.121503.

Archana, A., Long, C. and Chandran, K. (2022) 'Analysis of SARS-CoV-2 amino acid mutations in New York City Metropolitan wastewater (2020-2022) reveals multiple traits with human health implications across the genome and environment-specific distinctions'. medRxiv, p. 2022.07.15.22277689. Available at: https://doi.org/10.1101/2022.07.15.22277689.

Archetti, I. and Horsfall, F.L., Jr. (1950) 'Persistent antigenic variation of influenza A viruses after incomplete neutralisation in ovo with heterologous immune serum', *Journal of Experimental Medicine*, 92(5), pp. 441–462. Available at: https://doi.org/10.1084/jem.92.5.441.

Arora, P. *et al.* (2023) 'Neutralisation sensitivity of the SARS-CoV-2 XBB.1 lineage', *The Lancet Infectious Diseases*, 23(2), pp. 147–148. Available at: https://doi.org/10.1016/S1473-3099(22)00831-3.

Avanzato, V.A. *et al.* (2020) 'Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an Asymptomatic Immunocompromised Individual with Cancer', *Cell*, 183(7), pp. 1901-1912.e9. Available at: https://doi.org/10.1016/j.cell.2020.10.049.

Baltimore, D. (1971) 'Expression of animal virus genomes.', *Bacteriological Reviews*, 35(3), pp. 235–241.

Barton, M.I. *et al.* (2021) 'Effects of common mutations in the SARS-CoV-2 Spike RBD and its ligand, the human ACE2 receptor on binding affinity and kinetics', *eLife*. Edited by R.A. Fouchier, J.W. Van der Meer, and R.A. Fouchier, 10, p. e70658. Available at: https://doi.org/10.7554/eLife.70658.

Bertram, S. *et al.* (2010) 'TMPRSS2 and TMPRSS4 facilitate trypsin-independent spread of influenza virus in Caco-2 cells', *Journal of Virology*, 84(19), pp. 10016–10025. Available at: https://doi.org/10.1128/JVI.00239-10.

Blancou, J. *et al.* (2005) 'Emerging or re-emerging bacterial zoonoses: factors of emergence, surveillance and control', *Veterinary Research*, 36(3), p. 507. Available at: https://doi.org/10.1051/vetres:2005008.

Bolles, M., Donaldson, E. and Baric, R. (2011) 'SARS-CoV and emergent coronaviruses: viral determinants of interspecies transmission', *Current Opinion in Virology*, 1(6), pp. 624–634. Available at: https://doi.org/10.1016/j.coviro.2011.10.012.

Brunt, J. *et al.* (1993) *The Hantavirus epidemic in the Southwest:Rodent population dynamics and the implications for transmission ofHantavirus-associated adult respiratory distress syndrome (HARDS) in the Four Corners Region: Report to the Federal Centers for Disease Control and Prevention, Atlanta, Georgia*. Available at: https://doi.org/10.13140/RG.2.1.2623.1841.

Buchrieser, J. *et al.* (2020) 'Syncytia formation by SARS-CoV-2-infected cells', *The EMBO Journal*, 39(23), p. e106267. Available at: https://doi.org/10.15252/embj.2020106267.

Burmeister, A.R. (2015) 'Horizontal Gene Transfer', *Evolution, Medicine, and Public Health*, 2015(1), pp. 193–194. Available at: https://doi.org/10.1093/emph/eov018.

Cai, Y. *et al.* (2020) 'Distinct conformational states of SARS-CoV-2 spike protein', *Science*, 369(6511), pp. 1586–1592. Available at: https://doi.org/10.1126/science.abd4251.

Callaway, E. (2023) 'Coronavirus variant XBB.1.5 rises in the United States — is it a global threat?', *Nature*, 613(7943), pp. 222–223. Available at: https://doi.org/10.1038/d41586-023-00014-3.

Cameroni, E. *et al.* (2022) 'Broadly neutralizing antibodies overcome SARS-CoV-2 Omicron antigenic shift', *Nature*, 602(7898), pp. 664–670. Available at: https://doi.org/10.1038/s41586-021-04386-2.

Cao, Y. *et al.* (2023) 'Imprinted SARS-CoV-2 humoral immunity induces convergent Omicron RBD evolution', *Nature*, 614(7948), pp. 521–529. Available at: https://doi.org/10.1038/s41586-022-05644-7.

Carroll, T. *et al.* (2022) 'The B.1.427/1.429 (epsilon) SARS-CoV-2 variants are more virulent than ancestral B.1 (614G) in Syrian hamsters', *PLOS Pathogens*, 18(2), p. e1009914. Available at: https://doi.org/10.1371/journal.ppat.1009914.

Caserta, L.C. *et al.* (2023) 'White-tailed deer (Odocoileus virginianus) may serve as a wildlife reservoir for nearly extinct SARS-CoV-2 variants of concern', *Proceedings of the National Academy of Sciences*, 120(6), p. e2215067120. Available at: https://doi.org/10.1073/pnas.2215067120.

CDC (2019) *MERS Symptoms & Complications*, *Centers for Disease Control and Prevention*. Available at: https://www.cdc.gov/coronavirus/mers/about/symptoms.html (Accessed: 13 April 2023).

CDC (2020) *Coronavirus Disease 2019 (COVID-19)*, *Centers for Disease Control and Prevention*. Available at: https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html (Accessed: 14 March 2023).

CDC (2022) *How Flu Viruses Can Change*, *Centers for Disease Control and Prevention*. Available at: https://www.cdc.gov/flu/about/viruses/change.htm (Accessed: 14 April 2023).

Chakraborty, C. *et al.* (2022) 'Continent-wide evolutionary trends of emerging SARS-CoV-2 variants: dynamic profiles from Alpha to Omicron', *GeroScience*, 44(5), pp. 2371–2392. Available at: https://doi.org/10.1007/s11357-022-00619-y.

Challen, R. *et al.* (2021) 'Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: matched cohort study', *The BMJ*, 372, p. n579. Available at: https://doi.org/10.1136/bmj.n579.

Chan, Y.A. and Zhan, S.H. (2022) 'The Emergence of the Spike Furin Cleavage Site in SARS-CoV-2', *Molecular Biology and Evolution*, 39(1), p. msab327. Available at: https://doi.org/10.1093/molbev/msab327.

Chand, A. (2022) 'Live animal sales and SARS-CoV-2', *Nature Food*, 3(9), pp. 675–675. Available at: https://doi.org/10.1038/s43016-022-00606-8.

Chandler, J.C. *et al.* (2021) 'SARS-CoV-2 exposure in wild white-tailed deer (Odocoileus virginianus)', *Proceedings of the National Academy of Sciences*, 118(47), p. e2114828118. Available at: https://doi.org/10.1073/pnas.2114828118.

Chang, C. *et al.* (2006) 'Modular organization of SARS coronavirus nucleocapsid protein', *Journal of Biomedical Science*, 13(1), pp. 59–72. Available at: https://doi.org/10.1007/s11373-005-9035-9.

Chen, J. *et al.* (2023) 'Competition of SARS-CoV-2 variants on the pandemic transmission dynamics', *Chaos, Solitons & Fractals*, 169, p. 113193. Available at: https://doi.org/10.1016/j.chaos.2023.113193.

Chen, Y., Liu, Q. and Guo, D. (2020) 'Emerging coronaviruses: Genome structure, replication, and pathogenesis', *Journal of Medical Virology*, 92(4), pp. 418–423. Available at: https://doi.org/10.1002/jmv.25681.

Chinese Academy of Sciences and Wuhan Institute of Virology (2023) 'Wuhan Institute of virology; Our Team: Zhengli SHI'. Available at: http://www.whiov.cas.cn/sourcedb_whiov_cas/yw/rck/200907/t20090718_2100074.html (Accessed: 31 March 2023).

Choi, B. *et al.* (2020) 'Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host', *The New England Journal of Medicine*, 383(23), pp. 2291–2293. Available at: https://doi.org/10.1056/NEJMc2031364.

Chua, K.B. *et al.* (2000) 'Nipah Virus: A Recently Emergent Deadly Paramyxovirus', *Science*, 288(5470), pp. 1432–1435. Available at: https://doi.org/10.1126/science.288.5470.1432.

Cifuentes-Muñoz, N. and Ellis Dutch, R. (2019) 'To assemble or not to assemble: The changing rules of pneumovirus transmission', *Virus Research*, 265, pp. 68–73. Available at: https://doi.org/10.1016/j.virusres.2019.03.002.

Corey, L. *et al.* (2021) 'SARS-CoV-2 Variants in Patients with Immunosuppression', *New England Journal of Medicine*, 385(6), pp. 562–566. Available at: https://doi.org/10.1056/NEJMsb2104756.

Corman, V.M. *et al.* (2018) 'Chapter Eight - Hosts and Sources of Endemic Human Coronaviruses', in M. Kielian, T.C. Mettenleiter, and M.J. Roossinck (eds) *Advances in Virus Research*. Academic Press, pp. 163–188. Available at: https://doi.org/10.1016/bs.aivir.2018.01.001.

Cov-Lineages (2023). Available at: https://cov-lineages.org/lineage_list.html (Accessed: 27 February 2023).

Crits-Christoph, A. *et al.* (2023) *Genetic evidence of susceptible wildlife in SARS-CoV-2 positive samples at the Huanan Wholesale Seafood Market, Wuhan: Analysis and interpretation of data released by the Chinese Center for Disease Control*. Zenodo. Available at: https://doi.org/10.5281/zenodo.7754299.

Cuffari, B. (2020a) *The Size of SARS-CoV-2 and its Implications*, *News-Medical.net*. Available at: https://www.news-medical.net/health/The-Size-of-SARS-CoV-2-Compared-to-Other-Things.aspx (Accessed: 27 February 2023).

Cuffari, B. (2020b) *What are Furin Proteases?*, *News-Medical.net*. Available at: https://www.news-medical.net/health/What-are-Furin-Proteases.aspx (Accessed: 14 April 2023).

Cui, J., Li, F. and Shi, Z.-L. (2019) 'Origin and evolution of pathogenic coronaviruses', *Nature Reviews Microbiology*, 17(3), pp. 181–192. Available at: https://doi.org/10.1038/s41579-018-0118-9.

Cui, L. *et al.* (2015) 'The Nucleocapsid Protein of Coronaviruses Acts as a Viral Suppressor of RNA Silencing in Mammalian Cells', *Journal of Virology*, 89(17), pp. 9029–9043. Available at: https://doi.org/10.1128/JVI.01331-15.

Cui, Z. *et al.* (2022) 'Structural and functional characterizations of infectivity and immune evasion of SARS-CoV-2 Omicron', *Cell*, 185(5), pp. 860-871.e13. Available at: https://doi.org/10.1016/j.cell.2022.01.019.

Cunningham, A.A. (2005) 'A walk on the wild side—emerging wildlife diseases', *BMJ*, 331(7527), pp. 1214–1215. Available at: https://doi.org/10.1136/bmj.331.7527.1214.

Cyranoski, D. (2017a) *China to Permit Lab Poised to Study World's Most Dangerous Pathogens*, *Scientific American*. Available at: https://www.scientificamerican.com/article/china-to-permit-lab-poised-to-study-worlds-most-dangerous-pathogens/ (Accessed: 26 April 2023).

Cyranoski, D. (2017b) 'Inside the Chinese lab poised to study world's most dangerous pathogens', *Nature*, 542(7642), pp. 399–400. Available at: https://doi.org/10.1038/nature.2017.21487.

Cyranoski, D. (2020) 'The biggest mystery: what it will take to trace the coronavirus source', *Nature* [Preprint]. Available at: https://doi.org/10.1038/d41586-020-01541-z.

Dajose, L. (2021) *The Debate over Origins of SARS-CoV-2*, *California Institute of Technology*. Available at: https://www.caltech.edu/about/news/the-debate-over-origins-of-sars-cov-2 (Accessed: 1 April 2023).

Dance, A. (2021) 'The shifting sands of "gain-of-function" research', *Nature*, 598(7882), pp. 554–557. Available at: https://doi.org/10.1038/d41586-021-02903-x.

Davies, N.G. *et al.* (2021) 'Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England', *Science (New York, N.y.)*, 372(6538), p. eabg3055. Available at: https://doi.org/10.1126/science.abg3055.

Degnah, A.A. *et al.* (2020) 'Seroprevalence of MERS-CoV in healthy adults in western Saudi Arabia, 2011–2016', *Journal of Infection and Public Health*, 13(5), pp. 697–703. Available at: https://doi.org/10.1016/j.jiph.2020.01.001.

Delmas, B. and Laude, H. (1990) 'Assembly of coronavirus spike protein into trimers and its role in epitope expression', *Journal of Virology*, 64(11), pp. 5367–5375. Available at: https://doi.org/10.1128/JVI.64.11.5367-5375.1990.

Denison, M.R. *et al.* (2011) 'Coronaviruses', *RNA Biology*, 8(2), pp. 270–279. Available at: https://doi.org/10.4161/rna.8.2.15013.

Devaux, C.A. *et al.* (2021) 'Spread of Mink SARS-CoV-2 Variants in Humans: A Model of Sarbecovirus Interspecies Evolution', *Frontiers in Microbiology*, 12. Available at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.675528 (Accessed: 6 July 2023).

DeWolf, S. *et al.* (2022) 'SARS-CoV-2 in immunocompromised individuals', *Immunity*, 55(10), pp. 1779–1798. Available at: https://doi.org/10.1016/j.immuni.2022.09.006.

Diessner, E.M. *et al.* (2023) 'Mutation Effects on Structure and Dynamics: Adaptive Evolution of the SARS-CoV-2 Main Protease', *Biochemistry*, 62(3), pp. 747–758. Available at: https://doi.org/10.1021/acs.biochem.2c00479.

ECDC (2010) *Severe acute respiratory syndrome (SARS)*. Available at: https://www.ecdc.europa.eu/en/severe-acute-respiratory-syndrome (Accessed: 13 April 2023).

ECDC (2017) *Threats and outbreaks of Middle East respiratory syndrome coronavirus (MERS-CoV)*. Available at: https://www.ecdc.europa.eu/en/middle-east-respiratory-syndrome-coronavirus/threats-and-outbreaks (Accessed: 13 April 2023).

ECDC (2020) *MERS-CoV worldwide overview*. Available at: https://www.ecdc.europa.eu/en/middle-east-respiratory-syndrome-coronavirus-mers-cov-situation-update (Accessed: 13 April 2023).

Engelthaler, D.M. *et al.* (1999) 'Climatic and environmental patterns associated with hantavirus pulmonary syndrome, Four Corners region, United States.', *Emerging Infectious Diseases*, 5(1), pp. 87–94.

Faria, N.R. *et al.* (2021) 'Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil', *Science*, 372(6544), pp. 815–821. Available at: https://doi.org/10.1126/science.abh2644.

Fehr, A.R. and Perlman, S. (2015) 'Coronaviruses: An Overview of Their Replication and Pathogenesis', in H.J. Maier, E. Bickerton, and P. Britton (eds) *Coronaviruses: Methods and Protocols*. New York, NY: Springer (Methods in Molecular Biology), pp. 1–23. Available at: https://doi.org/10.1007/978-1-4939-2438-7_1.

Fenollar, F. *et al.* (2021) 'Mink, SARS-CoV-2, and the Human-Animal Interface', *Frontiers in Microbiology*, 12. Available at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.663815 (Accessed: 6 July 2023).

Ferreira, I.A.T.M. *et al.* (2021) 'SARS-CoV-2 B.1.617 Mutations L452R and E484Q Are Not Synergistic for Antibody Evasion', *The Journal of Infectious Diseases*, 224(6), pp. 989–994. Available at: https://doi.org/10.1093/infdis/jiab368.

Fleischmann, W.R. (1996) 'Viral Genetics', in S. Baron (ed.) *Medical Microbiology*. 4th edn. Galveston (TX): University of Texas Medical Branch at Galveston. Available at: http://www.ncbi.nlm.nih.gov/books/NBK8439/ (Accessed: 1 May 2023).

Focosi, D. and Maggi, F. (2022) 'Recombination in Coronaviruses, with a Focus on SARS-CoV-2', *Viruses*, 14(6), p. 1239. Available at: https://doi.org/10.3390/v14061239.

Follis, K.E., York, J. and Nunberg, J.H. (2006) 'Furin cleavage of the SARS coronavirus spike glycoprotein enhances cell-cell fusion but does not affect virion entry', *Virology*, 350(2), pp. 358–369. Available at: https://doi.org/10.1016/j.virol.2006.02.003.

Forni, D. *et al.* (2017) 'Molecular Evolution of Human Coronavirus Genomes', *Trends in Microbiology*, 25(1), pp. 35–48. Available at: https://doi.org/10.1016/j.tim.2016.09.001.

Freuling, C.M. *et al.* (2020) 'Susceptibility of Raccoon Dogs for Experimental SARS-CoV-2 Infection', *Emerging Infectious Diseases*, 26(12), pp. 2982–2985. Available at: https://doi.org/10.3201/eid2612.203733.

Gan, L. *et al.* (2022) 'Does potential antibody-dependent enhancement occur during SARS-CoV-2 infection after natural infection or vaccination? A meta-analysis', *BMC Infectious Diseases*, 22(1), p. 742. Available at: https://doi.org/10.1186/s12879-022-07735-2.

Gangavarapu, K. *et al.* (2023) 'Outbreak.info genomic reports: scalable and dynamic surveillance of SARS-CoV-2 variants and mutations', *Nature Methods*, 20(4), pp. 512–522. Available at: https://doi.org/10.1038/s41592-023-01769-3.

Gao, G. *et al.* (2022) *Surveillance of SARS-CoV-2 in the environment and animal samples of the Huanan Seafood Market*. preprint. In Review. Available at: https://doi.org/10.21203/rs.3.rs-1370392/v1.

Gerhard, W. and Webster, R. (1978) 'Antigenic drift in influenza A viruses. I. Selection and characterization of antigenic variants of A/PR/8/34 [HON1] influenza virus with monoclonal antibodies', *Journal of Experimental Medicine*, 148(2), pp. 383–392. Available at: https://doi.org/10.1084/jem.148.2.383.

Goldswain, H. *et al.* (2023) 'The P323L substitution in the SARS-CoV-2 polymerase (NSP12) confers a selective advantage during infection', *Genome Biology*, 24(1), p. 47. Available at: https://doi.org/10.1186/s13059-023-02881-5.

Gómez-Carballa, A. *et al.* (2020) 'Mapping genome variation of SARS-CoV-2 worldwide highlights the impact of COVID-19 super-spreaders', *Genome Research*, 30(10), pp. 1434–1448. Available at: https://doi.org/10.1101/gr.266221.120.

Graham, R.L. *et al.* (2008) 'SARS coronavirus replicase proteins in pathogenesis', *Virus Research*, 133(1), pp. 88–100. Available at: https://doi.org/10.1016/j.virusres.2007.02.017.

Grint, D.J. *et al.* (2022) 'Severity of Severe Acute Respiratory System Coronavirus 2 (SARS-CoV-2) Alpha Variant (B.1.1.7) in England', *Clinical Infectious Diseases*, 75(1), pp. e1120–e1127. Available at: https://doi.org/10.1093/cid/ciab754.

Guan, Y. *et al.* (2003) 'Isolation and Characterization of Viruses Related to the SARS Coronavirus from Animals in Southern China', *Science*, 302(5643), pp. 276–278. Available at: https://doi.org/10.1126/science.1087139.

Hadfield, J. *et al.* (2018) 'Nextstrain: real-time tracking of pathogen evolution', *Bioinformatics*. Edited by J. Kelso, 34(23), pp. 4121–4123. Available at: https://doi.org/10.1093/bioinformatics/bty407.

Halpin, K. *et al.* (2000) 'Isolation of Hendra virus from pteropid bats: a natural reservoir of Hendra virus', *Journal of General Virology*, 81(8), pp. 1927–1932. Available at: https://doi.org/10.1099/0022-1317-81-8-1927.

Hammer, A.S. *et al.* (2021) 'SARS-CoV-2 Transmission between Mink (Neovison vison) and Humans, Denmark', *Emerging Infectious Diseases*, 27(2), pp. 547–551. Available at: https://doi.org/10.3201/eid2702.203794.

Harrison, L.H. *et al.* (2006) 'Antigenic Shift and Increased Incidence of Meningococcal Disease', *The Journal of Infectious Diseases*, 193(9), pp. 1266–1274. Available at: https://doi.org/10.1086/501371.

Hasöksüz, M., Kiliç, S. and Saraç, F. (2020) 'Coronaviruses and SARS-COV-2', *Turkish Journal of Medical Sciences*, 50(9), pp. 549–556. Available at: https://doi.org/10.3906/sag-2004-127.

Hassanin, A., Rambaud, O. and Klein, D. (2022) 'Genomic Bootstrap Barcodes and Their Application to Study the Evolution of Sarbecoviruses', *Viruses*, 14(2), p. 440. Available at: https://doi.org/10.3390/v14020440.

Heeney, J.L. (2006) 'Zoonotic viral diseases and the frontier of early diagnosis, control and prevention', *Journal of Internal Medicine*, 260(5), pp. 399–408. Available at: https://doi.org/10.1111/j.1365-2796.2006.01711.x.

Hendra Virus Disease | CDC (2022). Available at: https://www.cdc.gov/vhf/hendra/index.html (Accessed: 28 March 2023).

Hijawi, B. *et al.* (2013) 'Novel coronavirus infections in Jordan, April 2012: epidemiological findings from a retrospective investigation', *Eastern Mediterranean Health Journal = La Revue De Sante De La Mediterranee Orientale = Al-Majallah Al-Sihhiyah Li-Sharq Al-Mutawassit*, 19 Suppl 1, pp. S12-18.

Hinnebusch, A.G., Ivanov, I.P. and Sonenberg, N. (2016) 'Translational control by 5'-untranslated regions of eukaryotic mRNAs', *Science (New York, N.Y.)*, 352(6292), pp. 1413–1416. Available at: https://doi.org/10.1126/science.aad9868.

Holmes, E.C. *et al.* (2021) 'The origins of SARS-CoV-2: A critical review', *Cell*, 184(19), pp. 4848–4856. Available at: https://doi.org/10.1016/j.cell.2021.08.017.

Holmes, K.V. (1999) 'Coronaviruses (Coronaviridae)', *Encyclopedia of Virology*, pp. 291–298. Available at: https://doi.org/10.1006/rwvi.1999.0055.

Hossain, M.S. *et al.* (2021) 'Genome-wide identification and prediction of SARS-CoV-2 mutations show an abundance of variants: Integrated study of bioinformatics and deep neural learning', *Informatics in Medicine Unlocked*, 27, p. 100798. Available at: https://doi.org/10.1016/j.imu.2021.100798.

Huang, Y. *et al.* (2020) 'Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19', *Acta Pharmacologica Sinica*, 41(9), pp. 1141–1149. Available at: https://doi.org/10.1038/s41401-020-0485-4.

Huynh, J. *et al.* (2012) 'Evidence Supporting a Zoonotic Origin of Human Coronavirus Strain NL63', *Journal of Virology*, 86(23), pp. 12816–12825. Available at: https://doi.org/10.1128/JVI.00906-12.

Jackson, B. *et al.* (2021) 'Generation and transmission of interlineage recombinants in the SARS-CoV-2 pandemic', *Cell*, 184(20), pp. 5179-5188.e8. Available at: https://doi.org/10.1016/j.cell.2021.08.014.

Jacquinet, E. *et al.* (2000) 'Cloning, genomic organization, chromosomal assignment and expression of a novel mosaic serine proteinase: epitheliasin', *FEBS Letters*, 468(1), pp. 93–100. Available at: https://doi.org/10.1016/S0014-5793(00)01196-0.

Jaimes, J.A. *et al.* (2020) 'Phylogenetic Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop', *Journal of Molecular Biology*, 432(10), pp. 3309–3325. Available at: https://doi.org/10.1016/j.jmb.2020.04.009.

Janies, D. *et al.* (2008) 'Evolution of genomes, host shifts and the geographic spread of SARS-CoV and related coronaviruses', *Cladistics*, 24(2), pp. 111–130. Available at: https://doi.org/10.1111/j.1096-0031.2008.00199.x.

Jedwab, R., Johnson, N.D. and Koyama, M. (2019) 'Pandemics, Places, and Populations: Evidence from the Black Death'. Rochester, NY. Available at: https://doi.org/10.2139/ssrn.3343855.

Jessie, B. and Dobrovolny, H.M. (2021) 'The role of syncytia during viral infections', *Journal of Theoretical Biology*, 525, p. 110749. Available at: https://doi.org/10.1016/j.jtbi.2021.110749.

Johnson, B.A. *et al.* (2020) 'Furin Cleavage Site Is Key to SARS-CoV-2 Pathogenesis', *bioRxiv*, p. 2020.08.26.268854. Available at: https://doi.org/10.1101/2020.08.26.268854.

Johnson, B.A. *et al.* (2021) 'Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis', *Nature*, 591(7849), pp. 293–299. Available at: https://doi.org/10.1038/s41586-021-03237-4.

Kaina, B. (2021) 'On the Origin of SARS-CoV-2: Did Cell Culture Experiments Lead to Increased Virulence of the Progenitor Virus for Humans?', *In Vivo (Athens, Greece)*, 35(3), pp. 1313–1326. Available at: https://doi.org/10.21873/invivo.12384.

Karim, F. *et al.* (2021) 'HIV status alters disease severity and immune cell responses in Beta variant SARS-CoV-2 infection wave', *eLife*. Edited by M.P. Davenport and L. Ndhlovu, 10, p. e67397. Available at: https://doi.org/10.7554/eLife.67397.

Keep, S. *et al.* (2018) 'The ADRP domain from a virulent strain of infectious bronchitis virus is not sufficient to confer a pathogenic phenotype to the attenuated Beaudette strain', *The Journal of General Virology*, 99(8), pp. 1097–1102. Available at: https://doi.org/10.1099/jgv.0.001098.

Kim, T.S. *et al.* (2006) 'Phenotypic Analysis of Mice Lacking the Tmprss2-Encoded Protease', *Molecular and Cellular Biology*, 26(3), pp. 965–975. Available at: https://doi.org/10.1128/MCB.26.3.965-975.2006.

Klinakis, A., Cournia, Z. and Rampias, T. (2021) 'N-terminal domain mutations of the spike protein are structurally implicated in epitope recognition in emerging SARS-CoV-2 strains', *Computational and Structural Biotechnology Journal*, 19, pp. 5556–5567. Available at: https://doi.org/10.1016/j.csbj.2021.10.004.

Kolb, A.F., Hegyi, A. and Siddell, S.G. (1997) 'Identification of residues critical for the human coronavirus 229E receptor function of human aminopeptidase N', *The Journal of General Virology*, 78 (Pt 11), pp. 2795–2802. Available at: https://doi.org/10.1099/0022-1317-78-11-2795.

Korneeva, N. *et al.* (2023) 'SARS-CoV-2 viral protein Nsp2 stimulates translation under normal and hypoxic conditions', *Virology Journal*, 20(1), p. 55. Available at: https://doi.org/10.1186/s12985-023-02021-2.

Lam, T.T.-Y. *et al.* (2020) 'Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins', *Nature*, 583(7815), pp. 282–285. Available at: https://doi.org/10.1038/s41586-020-2169-0.

Lambert, A.J., Martin, D.A. and Lanciotti, R.S. (2003) 'Detection of North American Eastern and Western Equine Encephalitis Viruses by Nucleic Acid Amplification Assays', *Journal of Clinical Microbiology*, 41(1), pp. 379–385. Available at: https://doi.org/10.1128/jcm.41.1.379-385.2003.

Latinne, A. *et al.* (2020) 'Origin and cross-species transmission of bat coronaviruses in China', *Nature Communications*, 11(1), p. 4235. Available at: https://doi.org/10.1038/s41467-020-17687-3.

Lednicky, J.A. *et al.* (2021) 'Independent infections of porcine deltacoronavirus among Haitian children', *Nature*, 600(7887), pp. 133–137. Available at: https://doi.org/10.1038/s41586-021-04111-z.

Lefeuvre, P. *et al.* (2009) 'Widely Conserved Recombination Patterns among Single-Stranded DNA Viruses', *Journal of Virology*, 83(6), pp. 2697–2707. Available at: https://doi.org/10.1128/jvi.02152-08.

Lei, J., Kusov, Y. and Hilgenfeld, R. (2018) 'Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein', *Antiviral Research*, 149, pp. 58–74. Available at: https://doi.org/10.1016/j.antiviral.2017.11.001.

Lenharo, M. and Wolf, L. (2023) 'US COVID-origins hearing renews debate over lab-leak hypothesis', *Nature*, 615(7952), pp. 380–381. Available at: https://doi.org/10.1038/d41586-023-00701-1.

Lerner, S. *et al.* (2021) *NIH Documents Provide New Evidence U.S. Funded Gain-of-Function Research in Wuhan*, *The Intercept*. Available at: https://theintercept.com/2021/09/09/covid-origins-gain-of-function-research/ (Accessed: 14 April 2023).

Letko, M., Marzi, A. and Munster, V. (2020) 'Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses', *Nature Microbiology*, 5(4), pp. 562–569. Available at: https://doi.org/10.1038/s41564-020-0688-y.

Li, W. *et al.* (2005) 'Bats Are Natural Reservoirs of SARS-Like Coronaviruses', *Science*, 310(5748), pp. 676–679. Available at: https://doi.org/10.1126/science.1118391.

Li X. *et al.* (2020) 'A furin cleavage site was discovered in the S protein of the 2019 novel coronavirus', *Chinese Journal of Bioinformatics*, pp. 103–108.

Li, X. *et al.* (2020) 'Emergence of SARS-CoV-2 through recombination and strong purifying selection', *Science Advances*, 6(27), p. eabb9153. Available at: https://doi.org/10.1126/sciadv.abb9153.

Liu, D.X., Liang, J.Q. and Fung, T.S. (2021) 'Human Coronavirus-229E, -OC43, -NL63, and -HKU1 (Coronaviridae)', *Encyclopedia of Virology*, pp. 428–440. Available at: https://doi.org/10.1016/B978-0-12-809633-8.21501-X.

Liu, Lihong *et al.* (2022) 'Striking antibody evasion manifested by the Omicron variant of SARS-CoV-2', *Nature*, 602(7898), pp. 676–681. Available at: https://doi.org/10.1038/s41586-021-04388-0.

Liu, P. *et al.* (2020) 'Cold-chain transportation in the frozen food industry may have caused a recurrence of COVID-19 cases in destination: Successful isolation of SARS-CoV-2 virus from the imported frozen cod package surface', *Biosafety and Health*, 2(4), pp. 199–201. Available at: https://doi.org/10.1016/j.bsheal.2020.11.003.

Liu, Z. *et al.* (2020) 'Composition and divergence of coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2', *Journal of Medical Virology*, 92(6), pp. 595–601. Available at: https://doi.org/10.1002/jmv.25726.

Lubinski, B., Jaimes, J.A. and Whittaker, G.R. (2022) 'Intrinsic furin-mediated cleavability of the spike S1/S2 site from SARS-CoV-2 variant B.1.1.529 (Omicron)', *bioRxiv*, p. 2022.04.20.488969. Available at: https://doi.org/10.1101/2022.04.20.488969.

Lubinski, B. and Whittaker, G.R. (2023) 'The SARS-CoV-2 furin cleavage site: natural selection or smoking gun?', *The Lancet Microbe*, 0(0). Available at: https://doi.org/10.1016/S2666-5247(23)00144-1.

Lytras, S. *et al.* (2022) 'Exploring the Natural Origins of SARS-CoV-2 in the Light of Recombination', *Genome Biology and Evolution*, 14(2), p. evac018. Available at: https://doi.org/10.1093/gbe/evac018.

Mahmood, T.B. *et al.* (2022) 'Missense mutations in spike protein of SARS-CoV-2 delta variant contribute to the alteration in viral structure and interaction with hACE2 receptor', *Immunity, Inflammation and Disease*, 10(9), p. e683. Available at: https://doi.org/10.1002/iid3.683.

Malato, J. *et al.* (2022) 'Risk of BA.5 Infection among Persons Exposed to Previous SARS-CoV-2 Variants', *New England Journal of Medicine*, 387(10), pp. 953–954. Available at: https://doi.org/10.1056/NEJMc2209479.

Malik, Y.A. (2020) 'Properties of Coronavirus and SARS-CoV-2', *The Malaysian Journal of Pathology*, 42(1), pp. 3–11.

Mallapaty, S. (2021) 'Closest known relatives of virus behind COVID-19 found in Laos', *Nature*, 597(7878), pp. 603–603. Available at: https://doi.org/10.1038/d41586-021-02596-2.

Mallapaty, S. (2023) 'COVID-origins study links raccoon dogs to Wuhan market: what scientists think', *Nature*, 615(7954), pp. 771–772. Available at: https://doi.org/10.1038/d41586-023-00827-2.

Masone, D., Alvarez, M.S. and Polo, L.M. (2022) 'The SARS-CoV-2 mutation landscape is shaped before replication starts'. bioRxiv, p. 2022.09.29.510149. Available at: https://doi.org/10.1101/2022.09.29.510149.

May, R.M. and Anderson, R.M. (1979) 'Population biology of infectious diseases: Part II', *Nature*, 280(5722), pp. 455–461. Available at: https://doi.org/10.1038/280455a0.

McLean, D.M. and Donohue, W.L. (1959) 'Powassan Virus: Isolation of Virus from a Fatal Case of Encephalitis', *Canadian Medical Association Journal*, 80(9), pp. 708–711.

Mecklin, J. (2021) 'The origin of COVID: Did people or nature open Pandora's box at Wuhan?', *Bulletin of the Atomic Scientists*, 5 May. Available at: https://thebulletin.org/2021/05/the-origin-of-covid-did-people-or-nature-open-pandoras-box-at-wuhan/ (Accessed: 7 June 2023).

Meng, B. *et al.* (2022) 'Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and fusogenicity', *Nature*, 603(7902), pp. 706–714. Available at: https://doi.org/10.1038/s41586-022-04474-x.

Merler, S. *et al.* (2013) 'Containing the accidental laboratory escape of potential pandemic influenza viruses', *BMC Medicine*, 11(1), p. 252. Available at: https://doi.org/10.1186/1741-7015-11-252.

Mickleburgh, S.P., Hutson, A.M. and Racey, P.A. (2002) 'A review of the global conservation status of bats', *Oryx*, 36(1), pp. 18–34. Available at: https://doi.org/10.1017/S0030605302000054.

Mlcochova, P. *et al.* (2021) 'SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion', *Nature*, 599(7883), pp. 114–119. Available at: https://doi.org/10.1038/s41586-021-03944-y.

Muga, G.O. *et al.* (2021) 'Indigenous knowledge of Rift Valley Fever among Somali nomadic pastoralists and its implications on public health delivery approaches in Ijara sub-County, North Eastern Kenya', *PLOS Neglected Tropical Diseases*, 15(2), p. e0009166. Available at: https://doi.org/10.1371/journal.pntd.0009166.

Mwenda, M. *et al.* (2021) 'Detection of B.1.351 SARS-CoV-2 Variant Strain — Zambia, December 2020', *Morbidity and Mortality Weekly Report*, 70(8), pp. 280–282. Available at: https://doi.org/10.15585/mmwr.mm7008e2.

Nadig, M. *et al.* (2022) 'Individuals with recent prior SARS-CoV-2 infection are at reduced risk of Omicron infection and associated hospitalization'. medRxiv, p. 2022.08.10.22278641. Available at: https://doi.org/10.1101/2022.08.10.22278641.

Naeli, P. *et al.* (2023) 'SARS-CoV-2 protein NSP2 enhances microRNA-mediated translational repression'. bioRxiv, p. 2023.01.01.522328. Available at: https://doi.org/10.1101/2023.01.01.522328.

Narayan, O., Griffin, D.E. and Chase, J. (1977) 'Antigenic Shift of Visna Virus in Persistently Infected Sheep', *Science*, 197(4301), pp. 376–378. Available at: https://doi.org/10.1126/science.195339.

Nasif, M.A.O. *et al.* (2022) 'Co-circulation of Multiple SARS-CoV-2 Variants of Concern in Dhaka, Bangladesh, during the Second Wave of the COVID-19 Pandemic', *Microbiology Resource Announcements*, 11(8), pp. e00549-22. Available at: https://doi.org/10.1128/mra.00549-22.

Nassar, A. *et al.* (2021) 'A Review of Human Coronaviruses' Receptors: The Host-Cell Targets for the Crown Bearing Viruses', *Molecules*, 26(21), p. 6455. Available at: https://doi.org/10.3390/molecules26216455.

Nichol, S.T., Arikawa, J. and Kawaoka, Y. (2000) 'Emerging viral diseases', *Proceedings of the National Academy of Sciences*, 97(23), pp. 12411–12412. Available at: https://doi.org/10.1073/pnas.210382297.

Nieto-Torres, J.L. *et al.* (2014) 'Severe acute respiratory syndrome coronavirus envelope protein ion channel activity promotes virus fitness and pathogenesis', *PLoS pathogens*, 10(5), p. e1004077. Available at: https://doi.org/10.1371/journal.ppat.1004077.

Nipah Virus (NiV) | CDC (2022). Available at: https://www.cdc.gov/vhf/nipah/index.html (Accessed: 28 March 2023).

Niu, S. *et al.* (2021) 'Molecular basis of cross-species ACE2 interactions with SARS-CoV-2-like viruses of pangolin origin', *The EMBO journal*, 40(16), p. e107786. Available at: https://doi.org/10.15252/embj.2021107786.

Ogawa, J. *et al.* (2020) 'The D614G mutation in the SARS-CoV2 Spike protein increases infectivity in an ACE2 receptor dependent manner'. bioRxiv, p. 2020.07.21.214932. Available at: https://doi.org/10.1101/2020.07.21.214932.

O'Toole, Á. *et al.* (2021) 'Tracking the international spread of SARS-CoV-2 lineages B.1.1.7 and B.1.351/501Y-V2 with grinch', *Wellcome Open Research*, 6, p. 121. Available at: https://doi.org/10.12688/wellcomeopenres.16661.2.

O'Toole, Á., Scher, E. and Rambaut, A. (2023) 'CoV Lineages'. Available at: https://cov-lineages.org/index.html (Accessed: 1 March 2023).

Ou, X. *et al.* (2020) 'Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV', *Nature Communications*, 11(1), p. 1620. Available at: https://doi.org/10.1038/s41467-020-15562-9.

Palmer, M.V. *et al.* (2021) 'Susceptibility of White-Tailed Deer (Odocoileus virginianus) to SARS-CoV-2', *Journal of Virology*, 95(11), p. 10.1128/jvi.00083-21. Available at: https://doi.org/10.1128/jvi.00083-21.

'Pango lineage names – Pango Network' (2022). Available at: https://www.pango.network/how-does-the-system-work/pango-lineage-names/ (Accessed: 7 March 2023).

'Pango Network – Helping track the transmission and spread of SARS-CoV-2' (2022). Available at: https://www.pango.network/ (Accessed: 1 March 2023).

Parry, J. (2004) 'Breaches of safety regulations are probable cause of recent SARS outbreak, WHO says', *BMJ : British Medical Journal*, 328(7450), p. 1222.

Parums, D.V. (2023) 'Editorial: The XBB.1.5 ("Kraken") Subvariant of Omicron SARS-CoV-2 and its Rapid Global Spread', *Medical Science Monitor : International Medical Journal of Experimental and Clinical Research*, 29, pp. e939580-1-e939580-3. Available at: https://doi.org/10.12659/MSM.939580.

Peacock, T.P. *et al.* (2021) 'The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets', *Nature Microbiology*, 6(7), pp. 899–909. Available at: https://doi.org/10.1038/s41564-021-00908-w.

Peacock, T.P. *et al.* (2022) 'The altered entry pathway and antigenic distance of the SARS-CoV-2 Omicron variant map to separate domains of spike protein'. bioRxiv, p. 2021.12.31.474653. Available at: https://doi.org/10.1101/2021.12.31.474653.

Peng, R. *et al.* (2021) 'Cell entry by SARS-CoV-2', *Trends in Biochemical Sciences*, 46(10), pp. 848–860. Available at: https://doi.org/10.1016/j.tibs.2021.06.001.

Plante, J.A. *et al.* (2021) 'Spike mutation D614G alters SARS-CoV-2 fitness', *Nature*, 592(7852), pp. 116–121. Available at: https://doi.org/10.1038/s41586-020-2895-3.

Prete, C.A. *et al.* (2022) 'Reinfection by the SARS-CoV-2 Gamma variant in blood donors in Manaus, Brazil', *BMC Infectious Diseases*, 22(1), p. 127. Available at: https://doi.org/10.1186/s12879-022-07094-y.

Public Helath Emergency (no date) *Science Safety Security, Biosafety Levels*, *Public Health Emergency*. Available at: https://www.phe.gov/s3/BioriskManagement/biosafety/Pages/Biosafety-Levels.aspx (Accessed: 22 March 2023).

Pyrc, K. *et al.* (2006) 'Mosaic structure of human coronavirus NL63, one thousand years of evolution', *Journal of Molecular Biology*, 364(5), pp. 964–973. Available at: https://doi.org/10.1016/j.jmb.2006.09.074.

Radvak, P. *et al.* (2021) 'SARS-CoV-2 B.1.1.7 (alpha) and B.1.351 (beta) variants induce pathogenic patterns in K18-hACE2 transgenic mice distinct from early strains', *Nature Communications*, 12(1), p. 6559. Available at: https://doi.org/10.1038/s41467-021-26803-w.

Rahimi, A., Mirzazadeh, A. and Tavakolpour, S. (2021) 'Genetics and genomics of SARS-CoV-2: A review of the literature with the special focus on genetic diversity and SARS-CoV-2 genome detection', *Genomics*, 113(1, Part 2), pp. 1221–1232. Available at: https://doi.org/10.1016/j.ygeno.2020.09.059.

Raj, V.S. *et al.* (2014) 'MERS: emergence of a novel human coronavirus', *Current Opinion in Virology*, 5, pp. 58–62. Available at: https://doi.org/10.1016/j.coviro.2014.01.010.

Rambaut, A. *et al.* (2020) 'A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology', *Nature Microbiology*, 5(11), pp. 1403–1407. Available at: https://doi.org/10.1038/s41564-020-0770-5.

Rand, R.P. and Parsegian, V.A. (1984) 'Physical force considerations in model and biological membranes', *Canadian Journal of Biochemistry and Cell Biology*, 62(8), pp. 752–759. Available at: https://doi.org/10.1139/o84-097.

Robson, F. *et al.* (2020) 'Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting', *Molecular Cell*, 79(5), pp. 710–727. Available at: https://doi.org/10.1016/j.molcel.2020.07.027.

Rocha, R.E.O. *et al.* (no date) 'A higher flexibility at the SARS-CoV-2 main protease active site compared to SARS-CoV and its potentialities for new inhibitor virtual screening targeting multi-conformers', *Journal of Biomolecular Structure & Dynamics*, pp. 1–21. Available at: https://doi.org/10.1080/07391102.2021.1924271.

Rofer, C. (2023) *Lab-Leak Intelligence Reports Aren’t Scientific Conclusions*, *Scientific American*. Available at: https://www.scientificamerican.com/article/lab-leak-intelligence-reports-arent-scientific-conclusions/ (Accessed: 7 June 2023).

Saito, A. *et al.* (2022) 'Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R mutation', *Nature*, 602(7896), pp. 300–306. Available at: https://doi.org/10.1038/s41586-021-04266-9.

Salajegheh Tazerji, S. *et al.* (2020) 'Transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) to animals: an updated review', *Journal of Translational Medicine*, 18(1), p. 358. Available at: https://doi.org/10.1186/s12967-020-02534-2.

Schulze, K.V., Hanchard, N.A. and Wangler, M.F. (2020) 'Biases in arginine codon usage correlate with genetic disease risk', *Genetics in Medicine*, 22(8), pp. 1407–1412. Available at: https://doi.org/10.1038/s41436-020-0813-6.

Sharun, K. *et al.* (2021) 'SARS-CoV-2 in animals: potential for unknown reservoir hosts and public health implications', *Veterinary Quarterly*, 41(1), pp. 181–201. Available at: https://doi.org/10.1080/01652176.2021.1921311.

Shen, L.W. *et al.* (2017) 'TMPRSS2: A potential target for treatment of influenza virus and coronavirus infections', *Biochimie*, 142, pp. 1–10. Available at: https://doi.org/10.1016/j.biochi.2017.07.016.

Shrestha, L.B. *et al.* (2022) 'Evolution of the SARS-CoV-2 omicron variants BA.1 to BA.5: Implications for immune escape and transmission', *Reviews in Medical Virology*, 32(5), p. e2381. Available at: https://doi.org/10.1002/rmv.2381.

Silver, A. (2022) *Taiwan's science academy fined for biosafety lapses after lab worker contracts COVID-19*, *Science insider news*. Available at: https://www.science.org/content/article/taiwan-s-science-academy-fined-biosafety-lapses-after-lab-worker-contracts-covid-19 (Accessed: 26 April 2023).

St. John, A.L. and Rathore, A.P.S. (2019) 'Adaptive immune responses to primary and secondary dengue virus infections', *Nature Reviews Immunology*, 19(4), pp. 218–230. Available at: https://doi.org/10.1038/s41577-019-0123-x.

Starr, T.N. *et al.* (2020) 'Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding', *Cell*, 182(5), pp. 1295-1310.e20. Available at: https://doi.org/10.1016/j.cell.2020.08.012.

'Statement of Nomenclature Rules – Pango Network' (2022). Available at: https://www.pango.network/the-pango-nomenclature-system/statement-of-nomenclature-rules/ (Accessed: 7 March 2023).

Stefanelli, P. *et al.* (2022) 'Co-circulation of SARS-CoV-2 Alpha and Gamma variants in Italy, February and March 2021', *Eurosurveillance*, 27(5), p. 2100429. Available at: https://doi.org/10.2807/1560-7917.ES.2022.27.5.2100429.

Tamura, T. *et al.* (2022) 'Virological characteristics of the SARS-CoV-2 XBB variant derived from recombination of two Omicron subvariants'. bioRxiv, p. 2022.12.27.521986. Available at: https://doi.org/10.1101/2022.12.27.521986.

Tang, G., Liu, Z. and Chen, D. (2022) 'Human coronaviruses: Origin, host and receptor', *Journal of Clinical Virology*, 155, p. 105246. Available at: https://doi.org/10.1016/j.jcv.2022.105246.

Taylor, L.H., Latham, S.M. and woolhouse, M.E.J. (2001) 'Risk factors for human disease emergence', *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* [Preprint]. Available at: https://doi.org/10.1098/rstb.2001.0888.

Tegally, H. *et al.* (2021) 'Detection of a SARS-CoV-2 variant of concern in South Africa', *Nature*, 592(7854), pp. 438–443. Available at: https://doi.org/10.1038/s41586-021-03402-9.

Telenti, A. *et al.* (2021) 'After the pandemic: perspectives on the future trajectory of COVID-19', *Nature*, 596(7873), pp. 495–504. Available at: https://doi.org/10.1038/s41586-021-03792-w.

Thunders, M. and Delahunt, B. (2020) 'Gene of the month: TMPRSS2 (transmembrane serine protease 2)', *Journal of Clinical Pathology*, 73(12), pp. 773–776. Available at: https://doi.org/10.1136/jclinpath-2020-206987.

Tian, S., Huajun, W. and Wu, J. (2012) 'Computational prediction of furin cleavage sites by a hybrid method and understanding mechanism underlying diseases', *Scientific Reports*, 2, p. 261. Available at: https://doi.org/10.1038/srep00261.

Turakhia, Y. *et al.* (2022) 'Pandemic-scale phylogenomics reveals the SARS-CoV-2 recombination landscape', *Nature*, 609(7929), pp. 994–997. Available at: https://doi.org/10.1038/s41586-022-05189-9.

Ullrich, S. *et al.* (2022) 'Main protease mutants of SARS-CoV-2 variants remain susceptible to nirmatrelvir', *Bioorganic & Medicinal Chemistry Letters*, 62, p. 128629. Available at: https://doi.org/10.1016/j.bmcl.2022.128629.

UniProtKB (2023a) *FURIN - Furin - Homo sapiens (Human) | UniProtKB | UniProt*. Available at: https://www.uniprot.org/uniprotkb/P09958/entry (Accessed: 14 April 2023).

UniProtKB (2023b) *S - Spike glycoprotein - Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) | Feature viewer | UniProtKB | UniProt*, *UniProtKB*. Available at: https://www.uniprot.org/uniprotkb/P0DTC2/feature-viewer (Accessed: 22 July 2023).

Venkatakrishnan, A.J. *et al.* (2023) 'Expanding repertoire of SARS-CoV-2 deletion mutations contributes to evolution of highly transmissible variants', *Scientific Reports*, 13(1), p. 257. Available at: https://doi.org/10.1038/s41598-022-26646-5.

Vijgen, L. *et al.* (2005) 'Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event', *Journal of Virology*, 79(3), pp. 1595–1604. Available at: https://doi.org/10.1128/JVI.79.3.1595-1604.2005.

V'kovski, P. *et al.* (2021) 'Coronavirus biology and replication: implications for SARS-CoV-2', *Nature Reviews. Microbiology*, 19(3), pp. 155–170. Available at: https://doi.org/10.1038/s41579-020-00468-6.

Vlasova, A.N. *et al.* (2022) 'Novel Canine Coronavirus Isolated from a Hospitalized Patient With Pneumonia in East Malaysia', *Clinical Infectious Diseases*, 74(3), pp. 446–454. Available at: https://doi.org/10.1093/cid/ciab456.

Vorou, R.M., Papavassiliou, V.G. and Tsiodras, S. (2007) 'Emerging zoonoses and vector-borne infections affecting humans in Europe', *Epidemiology & Infection*, 135(8), pp. 1231–1247. Available at: https://doi.org/10.1017/S0950268807008527.

Walker, A.S. *et al.* (2021) 'Tracking the Emergence of SARS-CoV-2 Alpha Variant in the United Kingdom', *New England Journal of Medicine*, 385(27), pp. 2582–2585. Available at: https://doi.org/10.1056/NEJMc2103227.

Walls, A.C. *et al.* (2020) 'Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein', *Cell*, 181(2), pp. 281-292.e6. Available at: https://doi.org/10.1016/j.cell.2020.02.058.

Wan, Y. *et al.* (2020) 'Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus', *Journal of Virology*, 94(7), pp. e00127-20. Available at: https://doi.org/10.1128/JVI.00127-20.

Wang, N. *et al.* (2018) 'Serological Evidence of Bat SARS-Related Coronavirus Infection in Humans, China', *Virologica Sinica*, 33(1), pp. 104–107. Available at: https://doi.org/10.1007/s12250-018-0012-7.

Wang, P. *et al.* (2021) 'Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7', *Nature*, 593(7857), pp. 130–135. Available at: https://doi.org/10.1038/s41586-021-03398-2.

Wang, Y. *et al.* (2022) 'Human SARS-CoV-2 has evolved to increase U content and reduce genome size', *International Journal of Biological Macromolecules*, 204, pp. 356–363. Available at: https://doi.org/10.1016/j.ijbiomac.2022.02.034.

Weiss, R.A. and McMichael, A.J. (2004) 'Social and environmental risk factors in the emergence of infectious diseases', *Nature Medicine*, 10(12 Suppl), pp. S70-76. Available at: https://doi.org/10.1038/nm1150.

Weissenhorn, W. *et al.* (1999) 'Structural basis for membrane fusion by enveloped viruses', *Molecular Membrane Biology*, 16(1), pp. 3–9. Available at: https://doi.org/10.1080/096876899294706.

'What are Pango lineages? – Pango Network' (2022). Available at: https://www.pango.network/how-does-the-system-work/what-are-pango-lineages/ (Accessed: 1 March 2023).

Wheatley, A.K. *et al.* (2021) 'Immune imprinting and SARS-CoV-2 vaccine design', *Trends in Immunology*, 42(11), pp. 956–959. Available at: https://doi.org/10.1016/j.it.2021.09.001.

Whittaker, G.R. (2021) 'SARS-CoV-2 spike and its adaptable furin cleavage site', *The Lancet. Microbe*, 2(10), pp. e488–e489. Available at: https://doi.org/10.1016/S2666-5247(21)00174-9.

WHO (2012) *Research priorities for zoonoses and marginalized infections: WHO TRS N°971*. Available at: https://www.who.int/publications-detail-redirect/WHO-TRS-971 (Accessed: 4 July 2023).

WHO (2021) *WHO-convened global study of origins of SARS-CoV-2: China Part*. Available at: https://www.who.int/publications-detail-redirect/who-convened-global-study-of-origins-of-sars-cov-2-china-part (Accessed: 27 February 2023).

Woo, P.C.Y. *et al.* (2005) 'Characterization and Complete Genome Sequence of a Novel Coronavirus, Coronavirus HKU1, from Patients with Pneumonia', *Journal of Virology*, 79(2), pp. 884–895. Available at: https://doi.org/10.1128/JVI.79.2.884-895.2005.

Woo, P.C.Y. (2023) 'ICTV Virus Taxonomy Profile: Coronaviridae 2023', *Journal of General Virology* [Preprint]. Edited by P. Simmonds and S.G. Siddell. Available at: https://ictv.global/report_9th/RNApos/Nidovirales/Coronaviridae (Accessed: 9 April 2023).

Worobey, M. *et al.* (2022) 'The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic', *Science*, 377(6609), pp. 951–959. Available at: https://doi.org/10.1126/science.abp8715.

Wu, H. *et al.* (2021) 'Nucleocapsid mutations R203K/G204R increase the infectivity, fitness, and virulence of SARS-CoV-2', *Cell Host & Microbe*, 29(12), pp. 1788-1801.e6. Available at: https://doi.org/10.1016/j.chom.2021.11.005.

Xia, X. (2021) 'Domains and Functions of Spike Protein in SARS-Cov-2 in the Context of Vaccine Design', *Viruses*, 13(1), p. 109. Available at: https://doi.org/10.3390/v13010109.

Xiao, X. *et al.* (2021) 'Animal sales from Wuhan wet markets immediately prior to the COVID-19 pandemic', *Scientific Reports*, 11(1), p. 11898. Available at: https://doi.org/10.1038/s41598-021-91470-2.

Xu, Z. *et al.* (2023) 'More or less deadly? A mathematical model that predicts SARS-CoV-2 evolutionary direction', *Computers in Biology and Medicine*, 153, p. 106510. Available at: https://doi.org/10.1016/j.compbiomed.2022.106510.

Yamamoto, M. *et al.* (2022) 'Metalloproteinase-Dependent and TMPRSS2-Independent Cell Surface Entry Pathway of SARS-CoV-2 Requires the Furin Cleavage Site and the S2 Domain of Spike Protein', *mBio*, 13(4), pp. e00519-22. Available at: https://doi.org/10.1128/mbio.00519-22.

Yang, X. *et al.* (2020) 'Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study', *The Lancet Respiratory Medicine*, 8(5), pp. 475–481. Available at: https://doi.org/10.1016/S2213-2600(20)30079-5.

Yang, Y. *et al.* (2014) 'Receptor usage and cell entry of bat coronavirus HKU4 provide insight into bat-to-human transmission of MERS coronavirus', *Proceedings of the National Academy of Sciences of the United States of America*, 111(34), pp. 12516–12521. Available at: https://doi.org/10.1073/pnas.1405889111.

Yang, Y. *et al.* (2020) 'SARS-CoV-2: characteristics and current advances in research', *Virology Journal*, 17(1), p. 117. Available at: https://doi.org/10.1186/s12985-020-01369-z.

Zhang, C. and Yang, M. (2022) 'Newly Emerged Antiviral Strategies for SARS-CoV-2: From Deciphering Viral Protein Structural Function to the Development of Vaccines, Antibodies, and Small Molecules', *International Journal of Molecular Sciences*, 23(11), p. 6083. Available at: https://doi.org/10.3390/ijms23116083.

Zhang, M. *et al.* (2022) 'A systematic review of Vaccine Breakthrough Infections by SARS-CoV-2 Delta Variant', *International Journal of Biological Sciences*, 18(2), pp. 889–900. Available at: https://doi.org/10.7150/ijbs.68973.

Zhang, Z., Zhang, J. and Wang, J. (2022) 'Surface charge changes in spike RBD mutations of SARS-CoV-2 and its variant strains alter the virus evasiveness via HSPGs: A review and mechanistic hypothesis', *Frontiers in Public Health*, 10. Available at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.952916 (Accessed: 22 July 2023).

Zhao, J., Cui, W. and Tian, B. (2020) 'The Potential Intermediate Hosts for SARS-CoV-2', *Frontiers in Microbiology*, 11. Available at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.580137 (Accessed: 2 April 2023).

Zhou, H. *et al.* (2020) 'A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein', *Current Biology*, 30(11), pp. 2196-2203.e3. Available at: https://doi.org/10.1016/j.cub.2020.05.023.

Zhou, P. *et al.* (2020) 'A pneumonia outbreak associated with a new coronavirus of probable bat origin', *Nature*, 579(7798), pp. 270–273. Available at: https://doi.org/10.1038/s41586-020-2012-7.

Zhou, P. and Shi, Z.-L. (2021) 'SARS-CoV-2 spillover events', *Science*, 371(6525), pp. 120–122. Available at: https://doi.org/10.1126/science.abf6097.

Zhu, Y. *et al.* (2020) *Design of Potent Membrane Fusion Inhibitors against SARS-CoV-2, an Emerging Coronavirus with High Fusogenic Activity*. Available at: https://doi.org/10.1128/JVI.00635-20.

Ziebuhr, J. (2005) 'The Coronavirus Replicase', in L. Enjuanes (ed.) *Coronavirus Replication and Reverse Genetics*. Berlin, Heidelberg: Springer (Current Topics in Microbiology and Immunology), pp. 57–94. Available at: https://doi.org/10.1007/3-540-26765-4_3.