MEDICAL UNIVERSITY OF VIENNA

Genomic epidemiology of SARS-CoV-2 in Austria

Doctoral thesis at the Medical University of Vienna
for obtaining the academic degree

Doctor of Philosophy

Submitted by

Jakob-Wendelin Genger, MSc

Supervisor:
Univ.-Prof. Dr. Andreas Bergthaler

CeMM Research Center for Molecular Medicine
of the Austrian Academy of Sciences
Lazarettgasse 14, AKH BT 25.3
1090 Vienna, Austria

Medical University of Vienna
Institute for Hygiene and Applied Immunology
Center for Pathophysiology, Infectiology and Immunology
Kinderspitalgasse 15
1090 Vienna, Austria

Vienna, 04/2023

Declaration

This thesis was composed in cumulative format and includes a first author publication published in **Science Translational Medicine** on **December 9th, 2020**. This study is the product of a collaborative effort which I managed and coordinated with Alexandra Popa and Andreas Bergthaler. I contributed to study design, performed experiments, and created, curated, managed, analyzed, and interpreted data as indicated below. Finally, I wrote the manuscript for the publication with Alexandra Popa, Christoph Bock and Andreas Bergthaler.

My individual contributions to the manuscript are:

- Contribution to conceptualization and preparation of the next-generation sequencing workflow used in this study,
- Curation and management of next-generation sequencing data, epidemiological data and metadata of samples for Figures 1A to 1E, 2C to 2D, 3C to 3D, 4B to 4E, 5A to 5C, S1A to S1E, S3A to S3H, S4G, S5A to S5D,
- Phylodynamic analyses for Figures 1B to 1D, 4B to 4E, 5A to 5B, S3A to S3H, S4G,
- Bioinformatic analysis and interpretation of next-generation sequencing data for Figures 1B to 1D, 2A to 2B, 4B to 4E, S3A to S3H,
- Experimental design and data interpretation together with collaborators for Figures 2C to 2D, 3C to 3E, 4A, 5B to 5D, S4A to S4G, S5A to S5D and
- Assembly of figures and presentation of data for Figures 1 to 5 and S4.

As shared first author, Alexandra Popa agreed to not use any of the content presented here as part of a dissertation or doctoral thesis.

Table of Contents

List of Figures

Introduction

Publication

Abstract

SARS-CoV-2, the causative agent of COVID-19, emerged in December 2019 and started spreading across the globe causing a pandemic with tremendous political and economic impact. The emergence of this new pathogen was met by an unprecedented global research response to develop effective countermeasures, treatments, and vaccines to mitigate its spread, morbidity, and mortality. The work presented in this thesis harnessed deep virus genome sequencing with the aim to investigate the mutational dynamics of SARS-CoV-2 in Austrian superspreader events in the first infection wave between February and May 2020. We applied genomic epidemiology to reconstruct infection clusters and infer transmission chains across borders to extend the limits of national epidemiological contact tracing. Using phylodynamic analysis, we found connections between clusters which eluded detection through traditional epidemiological contact tracing methods. By combining virus sequence information with epidemiological data, we were able to trace the emergence of new mutations from low frequency to fixation in two transmission chains. Moreover, we used transmission bottleneck size estimation in confirmed infector-infectee pairs to calculate the number of virions needed to start a productive infection. Finally, we monitored the occurrence and kinetics of low-frequency variants in longitudinal samples from hospitalized COVID-19 patients. Our study demonstrates the value of genomic epidemiology for pathogen surveillance and shines light on the mutational dynamics of SARS-CoV-2 and its inter- and intrahost genetic diversity. This study presents evidence for adaptation processes of SARS-CoV-2 and provides data resources for further research on the identification of immune evasion variants.

Zusammenfassung

Nach seinem ersten Auftreten im Dezember 2019 führte SARS-CoV-2, der Erreger von COVID-19, zu einer Pandemie mit weltweit weitreichenden politischen und wirtschaftlichen Auswirkungen. Dem Auftreten dieses neuen Virus wurde mit einem massiven Forschungseinsatz begegnet, der auf die Entwicklung wirksamer Maßnahmen, Medikamente und Impfstoffe zur Verringerung seiner Mortalität, Symptomatik und Ausbreitung abzielte. Die hier vorgestellte Studie nutzte Daten von Virusgenomsequenzierungen, um die Mutationsdynamik von SARS-CoV-2 in österreichischen Superspreader-Ereignissen während der ersten Infektionswelle zwischen Februar und Mai 2020 zu untersuchen. Mithilfe der genetischen Epidemiologie rekonstruierten wir Infektionscluster und leiteten grenzüberschreitende Übertragungsketten ab, die nicht von der nationalen epidemiologischen Kontaktverfolgung erfassbar waren. Phylodynamischen Analysen ermöglichten es uns Übertragungsereignisse zwischen epidemiologischen Clustern festzustellen, die durch traditionelle epidemiologische Kontaktverfolgungsmethoden unentdeckt geblieben waren. Die Kombination von Virussequenzinformationen mit epidemiologischen Daten ermöglichte es uns, das Auftreten neuer Mutationen von niedriger Allelfrequenz bis zur fixierten Mutation in zwei Übertragungsketten zu verfolgen. Wir wendeten dann mathematische Modelle zur Abschätzung des Infektionsübertragungsengpasses bei epidemiologisch bestätigten Infektionsketten an und berechneten die Anzahl an Virionen, die für eine erfolgreiche Infektion erforderlich waren. Zuletzt analysierten wir die Kinetik niederfrequenter Varianten während des Infektionsverlaufs in hospitalisierten COVID-19-Patienten. Unsere Studie zeigt den Nutzen der genetischen Epidemiologie für die Überwachung der Ausbreitung von Krankheitserregern und beleuchtet die Mutationsdynamik von SARS-CoV-2 sowie dessen genetische Diversität innerhalb und bei der Übertragung zwischen Wirten. Diese Studie präsentiert erste Ergebnisse zu Adaptionsmechanismen von SARS-CoV-2 und stellt Daten zur Forschung an Immunevasionsvarianten des Virus zur Verfügung.

Publication arising from this thesis

Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2

Science Translational Medicine; 09 Dec 2020; Vol. 12, Issue 573, eabe2555
DOI: 10.1126/scitranslmed.abe2555

Alexandra Popa*, Jakob-Wendelin Genger*, Michael D. Nicholson†, Thomas Penz†, Daniela Schmid†, Stephan W. Aberle†, Benedikt Agerer†, Alexander Lercher†, Lukas Endler , Henrique Colaço , Mark Smyth , Michael Schuster , Miguel L. Grau , Francisco Martínez-Jiménez, Oriol Pich, Wegene Borena, Erich Pawelka, Zsofia Keszei, Martin Senekowitsch, Jan Laine, Judith H. Aberle, Monika Redlberger-Fritz, Mario Karolyi, Alexander Zoufaly, Sabine Maritschnik, Martin Borkovec, Peter Hufnagl, Manfred Nairz, Günter Weiss, Michael T. Wolfinger, Dorothee von Laer, Giulio Superti-Furga, Nuria Lopez-Bigas, Elisabeth Puchhammer-Stöckl, Franz Allerberger, Franziska Michor, Christoph Bock, Andreas Bergthaler

*,† equal contributions

The publication was published as an open-access article and is distributed under a Creative Commons Attribution License 4.0 (CC BY) by the Journal Science Translational Medicine. Reprinted with permission from AAAS (American Association for the Advancement of Science). License number: OP-00145573

Abbreviations

AAV	-	Adeno-associated virus
ACE2	-	Angiotensin-converting enzyme 2
ADAR	-	Adenosine deaminase, RNA specific
ADCC	-	Antibody-dependent cellular cytotoxicity
AGES	-	Agentur für Gesundheit und Ernährungssicherheit
AIDS	-	Acquired immunodeficiency syndrome
APC	-	Antigen-presenting cell
BCV	-	Bovine coronavirus
CCoV	-	Canine coronavirus
cDC	-	Conventional dendritic cell
cGAMP	-	Cyclic GMP-AMP
cGAS	-	Cyclic GMP-AMP synthase
CMV	-	Cytomegalovirus
COVID-19	-	Coronavirus disease 2019
CTL	-	Cytotoxic T lymphocyte
DAMP	-	Damage-associated molecular pattern
DC	-	Dendritic cell
DMV	-	Double-membrane vesicle
EBV	-	Epstein-Barr virus
E protein	-	Envelope protein
ER	-	Endoplasmic reticulum
ERGIC	-	ER-to-Golgi intermediate compartment
ERV	-	Endogenous retrovirus
FCoV	-	Feline coronavirus
GISAID	-	Global initiative on sharing all influenza data
HBoV	-	Human bocavirus
HCV	-	Hepatitis C virus
HIV-1	-	Human Immunodeficiency Virus-1
HCoV	-	Human coronavirus
hMPV	-	Human metapneumovirus
HSV	-	Herpes simplex virus
HPV	-	Human papillomavirus
ICTV	-	International Committee on Taxonomy of Viruses
IBV	-	Infectious bronchitis virus

IFN	-	Interferon
IFNAR1	-	Interferon-alpha/beta receptor alpha chain
IFNAR2	-	Interferon-alpha/beta receptor beta chain
IFNLR1	-	Interferon lambda receptor 1
IL	-	Interleukin
IL10Rβ	-	Interleukin 10 receptor subunit beta
IRAK	-	Interleukin 1-receptor-associated kinase
IRF	-	Interferon regulatory factor
ISG	-	IFN-stimulated gene
ISGF3	-	IFN-stimulated gene factor 3
iSNV	-	Intrahost single-nucleotide variant
ISRE	-	IFN-stimulated response element
JAK1	-	Janus kinase 1
MAF	-	Minor allele frequency
MBL	-	Mannose-binding lectin
MDA5	-	Melanoma differentiation-associated protein 5
MERS-CoV	-	Middle East respiratory syndrome coronavirus
MHC-I	-	Class I major histocompatibility complex
MHV	-	Murine hepatitis virus
MIS-C	-	Multisystem inflammatory syndrome in children
M^{pro}	-	Main protease
M protein	-	Membrane protein
NF-κB	-	Nuclear factor kappa-light-chain-enhancer of activated B cells
NK cell	-	Natural killer cell
N protein	-	Nucleocapsid protein
nsp	-	Non-structural protein
ORF	-	Open reading frame
PAMP	-	Pathogen-associated molecular pattern
PANGOLIN	-	Phylogenetic Assignment of Named Global Outbreak Lineages
PASC	-	Post-acute sequelae of SARS-CoV-2 infection
pDC	-	Plasmacytoid dendritic cell
PHEIC	-	Public health emergency of international concern
PhyCLIP	-	Phylogenetic Clustering by Linear Integer Programming
PIV	-	Parainfluenza virus
PLpro	-	Papain-like protease
PRR	-	Pattern-recognition receptor
RdRp	-	RNA-dependent RNA polymerase

RIG-I	-	Retinoic acid inducible gene I
RIP1	-	Receptor-interacting protein
RSV	-	Respiratory syncytial virus
RT-PCR	-	Real-time polymerase chain reaction
RTC	-	Replication and transcription complex
SARS-CoV	-	Severe acute respiratory syndrome coronavirus
SARS-CoV-2	-	Severe acute respiratory syndrome coronavirus type 2
SIVcpz	-	Simian immunodeficiency virus of chimpanzees
S protein	-	Spike protein
STAT1	-	Signal transducer and activator of transcription 1
STAT2	-	Signal transducer and activator of transcription 2
STING	-	Stimulator of interferon genes
TANK	-	TRAF-family-member-associated NF-kB activator
TBK1	-	TANK-binding kinase 1
$TCID_{50}$	-	Median tissue culture infectious dose
TCV	-	Turkey coronavirus
TCR	-	T cell receptor
T_{FH} cell	-	T follicular helper cell
TGEV	-	Transmissible gastroenteritis coronavirus
T_H cell	-	CD4+ T helper cell
TIR	-	Toll/IL-1R homology
TLR	-	Toll-like receptor
TMPRSS2	-	Transmembrane protease serine 2
TNFα	-	Tumor necrosis factor-α
TRAF	-	TNFR-associated factor
T_{reg} cell	-	Regulatory T cell
TRIF	-	TIR-domain-containing adaptor-inducing interferon-β
TRS	-	Transcription regulatory sequence
TYK2	-	Tyrosine kinase 2
VOC	-	Variant of concern
VOI	-	Variant of interest
WHO	-	World Health Organization
WIV	-	Wuhan Institute for Virology

Acknowledgements

I would like to express my deepest gratitude to my academic supervisor Andreas Bergthaler for the great opportunity to do my PhD in his lab and for his unwavering support and guidance throughout my research journey. His valuable insights and constructive feedback have been invaluable in the work conducted for this thesis and several other projects. I am grateful for having had the privilege to contribute to a wide range of projects in the lab. This helped me to fully explore my potential, gave me the opportunity to expand my skill set with new challenges and fostered tremendous personal and professional growth.

I would like to thank my colleagues in the Bergthaler lab for their insightful discussions, helpful feedback, and support. Each and every one of you has contributed to my growth and development over the years, and I am deeply grateful for the knowledge and experiences gained from our interactions. Although it would be impossible to mention everyone individually, I would like to extend my thanks to Benedikt, Alex, Csilla, Henrique, Laura, Zsofia, Matthew, Lena, Anna S., Anna H., Hatoon, Lindsay and Mark for supporting me in the wet lab and providing support during challenging times. A special thank you to Alexandra, Lukas and Fabian for your expertise and guidance in bioinformatics, which helped me to develop my computational skills. Csilla, I am grateful for the many important tasks you have taken on in the lab and for being a caring presence during times of difficulty. Benedikt, thank you for your friendship and for engaging in our lively discussions about research which created ideas for at least a hundred exciting PhD projects. I am also deeply grateful to my committee members Stefan Kubicek and Dietmar Zehn for their invaluable feedback and constructive criticism throughout the years.

I would also like to thank all people at CeMM for creating such a wonderful and collaborative environment for research. I am pleased to have had the privilege of participating in this community of exceptional individuals and to experience their unwavering dedication to science. I would like to express my sincere appreciation to all students of the 2018 CeMM PhD cohort for their camaraderie, mutual support, and the creation of a strong peer group. Special thanks go to Ben for his valued friendship and stimulating debates on ideas how to transform science and society - some of which I hope will materialize in the future. Thank you to everyone on the 6th floor with special recognition to Sophie, for creating such a pleasant work environment.

I would like to extend my sincere thanks to the co-authors and collaborators of the publication presented in this thesis who generously devoted their time and expertise to this study during a period of instability and uncertainty. Without their participation, this research would not have been possible.

Furthermore, I would like to acknowledge the CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences and the Medical University of Vienna for providing me with the resources and facilities necessary for carrying out this research. I am grateful for the academic and intellectual environment that fostered my growth and development as an early-stage researcher. I am grateful to the Austrian Academy of Sciences ÖAW for awarding me a DOC fellowship.

I would like to express my heartfelt appreciation to my family and friends for their unconditional love and support throughout my academic journey. I would like to express my sincere gratitude to my parents, brothers, grandfather, and grandmother - the first PhD in our family. Their encouragement and belief in me have been a constant source of motivation. Their support has been invaluable in making this thesis possible. Finally, I wish to express my profound gratitude to my first academic supervisor, Reimo Kindler, whose guidance and mentorship paved the way for completing this doctoral thesis 14 years after I started my academic journey as a student.

1. Introduction

In December 2019, several health facilities in Wuhan, China reported clusters of patients with pneumonia of unknown etiology (Wu *et al*, 2020; Zhu *et al*, 2020; Huang *et al*, 2020). The causative agent for this disease was rapidly isolated and identified as a novel coronavirus named Severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2) (Wu *et al*, 2020; Zhu *et al*, 2020; Coronaviridae Study Group of the International Committee on Taxonomy of Viruses *et al*, 2020). On January 31st 2020, the World Health Organization (WHO) recognized the SARS-CoV-2-related disease, termed coronavirus disease 2019 (COVID-19) as a public health emergency of international concern due to its fast spread through the population (World Health Organization, 2020b, 2020c). On March 11th, the WHO made the assessment that COVID-19 could be characterized as a pandemic with high global risk potential after counting 118,000 confirmed cases in 114 countries and over 4,200 global confirmed deaths (World Health Organization, 2020d). This was followed by an unprecedented global response with vast social and economic impact to mitigate the public health risk imposed by SARS-CoV-2. Despite these efforts, more than 676 million cases of SARS-CoV-2 infection and over 6.8 million deaths related to COVID-19 were confirmed as of March 10th, 2023 (Dong *et al*, 2020: COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University).

Starting from a few clusters, SARS-CoV-2 quickly reached global distribution within a few months (Hu *et al*, 2021). The international scientific community countered this outbreak by initiating an unparalleled research response across all disciplines aiming to investigate origin and characteristics of the pathogen and its disease, as well as to develop countermeasures, medical treatments, and vaccines.

This thesis will focus on the spread of SARS-CoV-2 in Austria during the first infection wave in spring 2020 and how complementing epidemiological data with virus genome sequencing data was a powerful approach to determine transmission characteristics of the virus in infection clusters. This approach gave insights into the viral mutational trajectory, the emergence of new SARS-CoV-2 variants and allowed to estimate the number of transmitted viral particles between individuals that led to a productive infection.

1.1. History and characteristics of emerging infectious diseases

From the “Justinianic Plague” to the 1918 Influenza pandemic and Human Immunodeficiency Virus (HIV) – emerging and re-emerging infectious diseases have challenged humankind throughout history (Dobson & Carper, 1996; Taubenberger *et al*, 2019; Morens & Fauci, 2020; van Doorn, 2021; Baker *et al*, 2022). This chapter will give a historical overview over selected emerging infectious diseases. These will serve as examples for how we can gain knowledge about the factors that determine their emergence, spread and control from past events. Moreover, this chapter will introduce the nomenclature of common parameters that are used to describe infectious disease outbreaks.

1.1.1. Traces of ancient emerging diseases today

Infectious microbes and viruses newly emerged already far before the start of historiography and some of them developed into currently existing endemic (prevalent in humans) or enzootic (prevalent in animals) infectious diseases (Dobson & Carper, 1996; Morens & Fauci, 2020). Moreover, endogenous retroviruses (ERVs) in the human genome are remnants of ancient retrovirus outbreaks that reach back thousands or millions of years. Throughout a long-lasting adaptation process, ERVs achieved stable co-existence with their hosts and provided novel genetic elements to host genomes or drove the evolution of host genes that control viral infection (Virgin *et al*, 2009; Johnson, 2019; Morens & Fauci, 2020). Furthermore, some latently infecting viruses like cytomegalovirus (CMV), Epstein-Barr virus (EBV), herpes simplex virus (HSV) or adeno-associated viruses (AAV), are estimated to have infected the vast majority of humans alive today (Virgin *et al*, 2009). They achieved an intricate state of metastable equilibrium with the host and can become a permanent member of the human metagenome through a plethora of adaptation strategies (Virgin *et al*, 2009; Fauci & Morens, 2012; Johnson, 2019). ERVs and the presented latently infecting viruses were emerging infectious diseases a very long time ago but achieved long-term survival by maintaining continued transmission through adaptation to their host’s genetic, cellular and immune mechanisms (Virgin *et al*, 2009; Morens & Fauci, 2020; van Doorn, 2021). However, these examples of virus-host co-existence are thought to be the product of long-lasting adaptation processes that already started before human settlement (Morens & Fauci, 2020; Dobson & Carper, 1996; van Doorn, 2021). It is possible that these pathogens were part of the pathogen collection that humans inherited from higher apes, but it is difficult to reconstruct the emergence, spread and establishment of these ancient infectious diseases (Dobson & Carper, 1996). Nevertheless, they are now an almost ubiquitous part of the human metagenome and

examples for the constant interaction between humans and pathogens throughout evolution (Dobson & Carper, 1996; Virgin *et al*, 2009; van Doorn, 2021).

In fact, the first infectious diseases emerging in humans that later became endemic diseases like measles, smallpox, tuberculosis, rabies, and influenza, are expected to have developed from animal diseases at the time of the establishment of the first larger settlements (Dobson & Carper, 1996; van Doorn, 2021; Baker *et al*, 2022). At that time, close contact to animals due to animal husbandry provided the source for emergence of these new diseases through spillover. The aggregation of humans in the first cities with several thousand inhabitants offered then also sufficient population sizes to cause an outbreak of epidemic dimensions (Dobson & Carper, 1996; Morens & Fauci, 2020; van Doorn, 2021; Baker *et al*, 2022). However, it was the communication and interaction between different human settlements that provided the means for continued spread to maintain survival and avoid de-emergence of the pathogens due to the lack of susceptible individuals. It is estimated that interaction networks between different human settlements were one of the most important factors for the establishment of measles as a common re-occurring disease, for example (Dobson & Carper, 1996; van Doorn, 2021; Baker *et al*, 2022). In addition to these basic requirements, other factors that are crucial for the emergence and establishment of new infectious diseases will be illustrated through the examples of historic disease outbreaks in the following paragraphs.

1.1.2. The “Plague of Athens” – The first historic accounts of an epidemic

One of the first and most comprehensive accounts of an epidemic were written by the Athenian general and historian Thucydides during the Peloponnesian War (431 - 404 BC) when the city of Athens was under siege by Sparta (Morens *et al*, 2008; Littman, 2009). The “Plague of Athens” struck the city-state in 430 BC and lasted for about 5 years without interruption, causing about 100,000 deaths, which was equal to 25% of the population. Thucydides had the intention to share his observations with future generations in case of a re-emergence of the disease. Therefore, he documented the signs and spread of the disease as well as its symptoms, described susceptibility and risk groups and even presented the course of infection of prominent Athenians as case studies (Morens *et al*, 2008; Littman, 2009). His account can be deemed the first clinical-epidemiological characterization of an infectious disease outbreak (Morens *et al*, 2008). However, it does not meet modern standards, for example simply due to the lack of a standardized medical terminology at that time (Littman, 2009). For this reason, the etiology of this disease, characterized by symptoms like pustular rash, high fever and diarrhea with a high mortality rate, is still a subject of studies. Paleopathological and medical history studies suggest more than 20 different pathogens as causative agent for the outbreak

(Morens *et al*, 2008; Littman, 2009). Although Thucydides' reports are still a matter of discussion, they clearly describe some fundamental factors that influence the spread of an emerging infectious disease: (1) the role of international trade routes (the disease is said to have originated from Ethiopia and spread first from the port of Athens), (2) the role of universal susceptibility (the disease affected members of all ages, genders, socioeconomic groups, etc.), (3) the consequences of the war and starvation (e.g. associated refugee movements, crowding, inadequate sanitation and hygienic conditions supported the spread of the disease) and (4) lack/collapse of public health infrastructure promoted disease spread (Morens *et al*, 2008; Littman, 2009).

1.1.3. The "Justinianic Plague" and the "Black Death" – recurring pandemics

The "Justinianic Plague" (first appearance in 541 AD, recurrent until 750/767 AD) and the "Black Death" of the 14th century (1347 AD – 1350 AD) serve as further examples for how connections between countries, e.g. via trade routes, can act as transmission routes for an emerging pathogen (Morens & Fauci, 2020; Barbieri *et al*, 2020; Baker *et al*, 2022). They also show that an infectious disease can emerge, disappear (or de-emerge), and re-emerge several times from reservoirs. The causative agent of the disease in both pandemics was *Yersinia pestis*, a gram-negative bacterium that can be transmitted via infected flea bites but also via aerosols exchanged between individuals (Harbeck *et al*, 2013; Barbieri *et al*, 2020). *Y. pestis* can be found in soil, it resides in protozoa and it infects over 200 species of mammals, of which rodents played the most important role as reservoirs for these two pandemics (Barbieri *et al*, 2020). Contemporary sources estimate that both pandemics have caused up to 50 million deaths across Europe, Asia, and North Africa (Morens *et al*, 2008; Harbeck *et al*, 2013; Morens & Fauci, 2020). However, while the first pandemic put a lot of social and economic pressure on the late Roman Empire, the second large outbreak of *Y. pestis* found a completely different, more fragmented Europe. At that time, the first quarantine measures (derived from the term "quaranta" describing 40 days isolation for arriving ships) were introduced in Venice (Morens *et al*, 2008; Whitby, 2008). Nevertheless, the "Black Death" caused massive death tolls and had an unprecedented impact on society and culture (Morens *et al*, 2008). In summary, these two pandemics serve as examples for (1) possible recurring emergence of an infectious disease pathogen from animal reservoirs and (2) first public health countermeasures like quarantines against infectious disease emergence, which are still part of today's disease control repertoire.

1.1.4. The 1918 Influenza and HIV – pandemics of modern times

The previous examples of infectious disease outbreaks happened before it was even known that microbes can cause disease and how they are transmitted between individuals. The "germ theory", which describes microbes as the causative agent of infectious diseases, and Koch's postulates were just formulated in the second half of the 19th century based on the works of Robert Koch and Louis Pasteur (Porter, 1998; Morens *et al*, 2004). This knowledge existed in the beginning of the 20th century when the 1918 Influenza emerged, but it could not prevent the 1918 Influenza becoming the so far deadliest pandemic in human history with 50 to 100 million deaths. The causative agent of this disease was a new influenza A virus H1N1 that may have emerged from waterfowls (Morens *et al*, 2008; Taubenberger *et al*, 2019; van Doorn, 2021). Influenza A viruses constantly circulate in the human population. Genetic studies have shown that the three subsequent pandemic Influenza A strains H2N2 (1957), H3N2 (1968) and H1N1 (2009) acquired genetic elements from the 1918 Influenza through reassortment and, thus, are discussed to be descendants of the 1918 pandemic virus. In case of the 1918 influenza A pandemic, secondary bacterial pneumonia from gram-positive bacteria in almost all influenza A patients was a major contributor to the high mortality (Taubenberger *et al*, 2019; Morens & Fauci, 2020). Another interesting observation during this pandemic was the rather unusual mortality over different ages groups which depicted a "W-shaped" curve with peaks in the age groups below 5 years, between 20 – 40 years and over 80 years. It is estimated that this resulted from pre-existent immunity in higher age groups from previous exposure to H1 or N1 surface proteins or conserved epitopes during the 19th century (Morens *et al*, 2008; Taubenberger *et al*, 2019; Morens & Fauci, 2020).

Acquired immunodeficiency syndrome (AIDS) was first reported in 1981 in five patients before the etiological agent for the disease, HIV, was discovered in 1983 (Fauci & Morens, 2012; Barré-Sinoussi *et al*, 2013). The disease did not receive much public attention and was stigmatized as supposedly only affecting certain risk groups, but HIV/AIDS developed into a still ongoing pandemic and is affecting individuals throughout the whole population (De Cock *et al*, 2011; Barré-Sinoussi *et al*, 2013). A study in 1986 found the earliest serum sample that tested positive for HIV-1 reactive antibodies. It was collected in the Democratic Republic of Congo in 1959. Phylogenetic analyses suggest that HIV-1 originated from simian immunodeficiency virus of chimpanzees (SIVcpz) upon cross-species transmission, which is estimated to have happened in the beginning of the 20th century (De Cock *et al*, 2011; Fauci & Morens, 2012). The pathogenesis of HIV is characterized by a progressive depletion of $CD4^+$ T cells, the primary host cells of the virus, which results in an acquired immunodeficiency that renders the patient more susceptible to a wide range of immune-system controlled diseases

and leads to more severe disease outcomes. Sexual transmission and direct blood-to-blood contact are the main routes of transmission (Dobson & Carper, 1996; Barré-Sinoussi *et al*, 2013; German Advisory Committee Blood (Arbeitskreis Blut), 2016). Although HIV/AIDS is in terms of transmission, course of infection, and symptoms very different compared to the infectious diseases described above, it serves as an example for some factors involved in infectious disease emergence that were not covered yet. It was ultimately not international trade, commerce, war, the collapse of public health infrastructure or famines that played a major role in the spread of this virus. On the contrary, factors that contributed to the establishment of the disease were (1) the course of infection that is characterized by a long incubation period of ~11 years between HIV infection and symptomatic AIDS in adults, as well as (2) human demographics and behavior, social inequality and stigma, and (3) a lack of political will (De Cock *et al*, 2011; Barré-Sinoussi *et al*, 2013). On the other hand, the HIV/AIDS pandemic was the first pandemic in human history, where humanity was equipped with comparably advanced tools to counter this outbreak with scientific advancements. The first serological tests for HIV were established in 1985. The development of vaccines and medical treatments started soon after identification of the virus. However, several vaccine trials have failed so far, and the development of an effective vaccine remains elusive (Barré-Sinoussi *et al*, 2013). Nevertheless, major progress was made in the development and refinement of anti-retroviral drugs, and current drug regimens allow complete control of the infection and blockade of sexual transmission (treatment as prevention) or mother-to-child transmission for the majority of patients with access to these treatment options (Barré-Sinoussi *et al*, 2013; German Advisory Committee Blood (Arbeitskreis Blut), 2016).

1.1.5. Nomenclature of infectious disease outbreaks

These historic examples of emerging infectious diseases showcase the footprint that pathogens left in human history (Dobson & Carper, 1996; Morens *et al*, 2008; Baker *et al*, 2022). They also demonstrate factors that influence the emergence and propagation of a newly emerging pathogen. The following chapter will introduce basic principles and some nomenclature of infectious disease outbreaks and their epidemiology.

The field of epidemiology is a science of public health and investigates the causality, characteristics, and distribution of diseases in the population (Rothman *et al*, 2008; Frérot *et al*, 2018). In the context of infectious diseases, epidemiologists investigate for example transmission chains and collect clinical and diagnostic data to determine parameters like incidence rate, attack rate and case fatality rates. One of the most popular and publicly communicated epidemiological metrics during the COVID-19 pandemic was the incidence

rate, which describes the number of new cases in a defined time period (Rothman *et al*, 2008; Frérot *et al*, 2018). Another measure of occurrence that rather represents a snapshot in time is the prevalence of a disease, which describes the proportion of a population that is affected by a disease or medical condition at a specific timepoint. A parameter that plays an important role for the spread of infectious diseases is the risk of infection among potentially susceptible exposed individuals – the attack rate. Based on the incidence, the case fatality rate incorporates the clinical outcome death and represents the proportion of those that succumb to the disease or medical condition (Rothman *et al*, 2008).

One of the most fundamental classifications of an infectious disease outbreak is the distinction between epidemic and pandemic. Epidemics are characterized by high incidence rates of the disease (Morens & Fauci, 2020). A disease outbreak can be classified as pandemic if the disease spreads over the globe or large geographical regions like multiple continents. A similar terminology exists for diseases in animals that are classified as epizootic and panzootic. Infectious diseases with high prevalence in the human population or animals are termed endemic diseases or enzootic diseases, respectively (Rothman *et al*, 2008; Morens & Fauci, 2020). Many of today's endemic infectious diseases are expected to have emerged around the time when human settlements developed into cities and provided larger population densities in close contact with animals. These were fertile grounds for animal pathogens to infect human hosts with the opportunity for effective person-to-person transmission in a larger human population. These emerging infectious diseases that originated from animal hosts and lead to dead-end infections or productive transmission between humans are called zoonotic diseases (Dobson & Carper, 1996; Virgin *et al*, 2009; van Doorn, 2021).

Past infectious disease outbreaks such as the ones introduced above have shown that the emergence of an infectious disease and its ability to follow a trajectory towards an endemic state depend on a variety of factors. These factors can be determined by the agent, the host or the environment (see Figure 1) (Morens & Fauci, 2020; van Doorn, 2021; Baker *et al*, 2022). One of these is increased mobility of individuals via international traffic and trade. Furthermore, population migration events in consequence of wars or famines facilitate the geographical distribution of infectious diseases. Different individual characteristics of the host defined by their social environment and lifestyle like human behavior, susceptibility to infection, poverty and social inequality can have positive or negative impact on the establishment of a new pathogen diseases. Moreover, the political and economic framework of a society plays a decisive role and can affect the emergence and establishment of diseases in a variety of ways (Morens *et al*, 2008; Morens & Fauci, 2020; van Doorn, 2021; Baker *et al*, 2022). On one hand, the availability of technological advancements and a medical infrastructure that allow the rapid

identification of new pathogens, the development of molecular tests, vaccines, and treatments, are powerful tools to react to a new disease early during its emergence phase. On the other hand, the effectiveness of these measures will be determined by other factors like the political will to act on the situation, public reception of mitigation strategies like restricting outdoor activities and gatherings (known as “lockdowns”), school closures, mask mandates or in extreme cases the breakdown of public healthcare due to exhaustion of capacities in the public health sector (Morens *et al*, 2008). Finally, different factors can affect the fitness and adaptability of the pathogen, like genetic instability, changes in the molecular characteristics of the microbes like drug resistance or immune evasion, weather and changing climate (Morens *et al*, 2008; van Doorn, 2021; Baker *et al*, 2022). In RNA viruses like HIV, SARS-CoV-2 and influenza, different mechanisms result in higher genetic instability and ability for adaptation to selection pressures, such as inherently error-prone polymerases or the transmission of a cloud of genetic variants (quasispecies transmission) (Morens & Fauci, 2020; van Doorn, 2021; Villa *et al*, 2021).

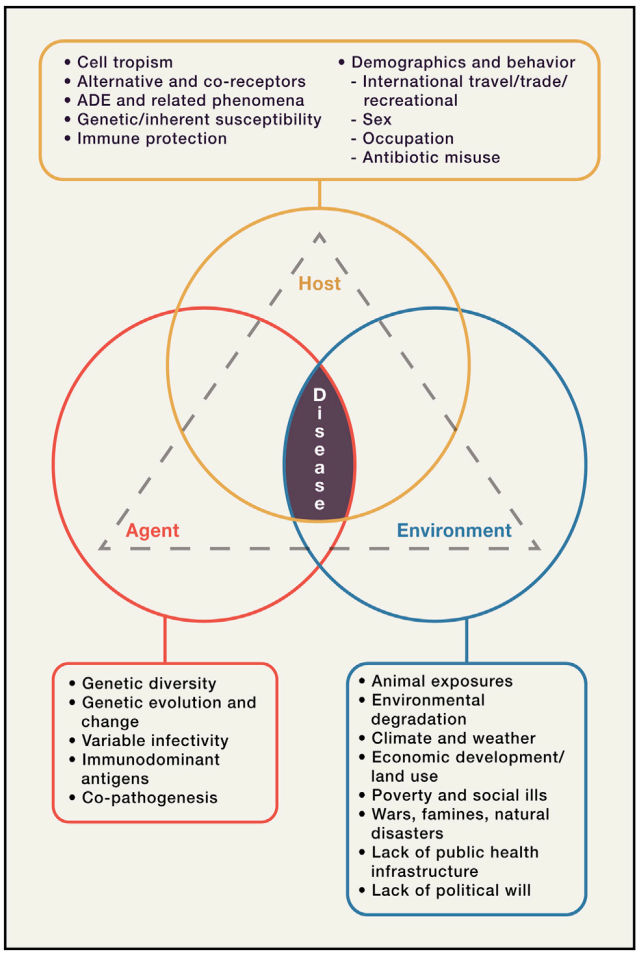


Figure 1: Factors that influence the emergence or re-emergence of an infectious disease and its establishment towards an endemic state. Different factors of the host, the pathogenic agent and the environment influence the emergence of new infectious diseases and play a decisive role for their path towards establishment. Figure obtained from (Morens & Fauci, 2020). Reprinted with permission from Elsevier. License number: 5518180087346

Infectious diseases can be assigned to five major categories based on their origin and mode of emergence. Newly emerging infectious diseases like HIV/AIDS and SARS-CoV-2 were recognized for the first time in the human host (Fauci & Morens, 2012; Morens & Fauci, 2020; van Doorn, 2021). Re-emerging infectious diseases are known pathogens that continue to re-emerge with new characteristics or at different geographical locations like *Y. pestis* described above or Mpox as a contemporary example (Barbieri *et al*, 2020; Kmiec & Kirchhoff, 2022; Ranjan & Biswal, 2022). Established or endemic infectious diseases were newly emerging diseases in the past but managed to sustain in the population, for example through host adaptation, and are at present prevalent in the human population as described above. Two rather new classes are deliberately emerging infectious diseases that are released by humans with harmful purpose (e.g. bioweapons) and accidentally emerging infectious diseases which are unintentionally spread pathogens or the product of medical treatments (e.g. of transmissible vaccine-derived viruses) (Fauci & Morens, 2012; Morens & Fauci, 2020; van Doorn, 2021).

Infectious disease transmission models were developed to describe and predict the propagation of an infectious agent in a population by integrating independent parameters of both the population and the pathogen (Rothman *et al*, 2008). Two important parameters can be modelled to describe the dynamics of a pathogen in the host population: The basic reproductive ratio (R_0) and the threshold of establishment (H_T). The basic reproductive ratio assumes a completely susceptible starting population and describes the number of new cases caused by an average infectious host at the start of the disease outbreak (Dobson & Carper, 1996; Rothman *et al*, 2008). To persist in the host population, a new pathogen must reach an R_0 of at least 1, meaning each host transmits the disease to at least one other individual. An extension of the basic reproduction number R_0 is the effective reproduction number R_t that describes how many successful transmission events occur from each infectious case on average at any time point (t) once the pathogen has spread among the population (Anderson *et al*, 2004; Rothman *et al*, 2008). If R_0 or R_t drop below 1, the infectious disease will de-emerge.

As discussed above, size and density of the host population are critical factors for the emergence of a new pathogen. The threshold of establishment H_T describes the population size that is necessary for a pathogen with a given reproductive ratio to sustain itself in that population (Dobson & Carper, 1996; Rothman *et al*, 2008). Kermack and McKendrick postulated in 1927 a first model to determine the threshold of establishment H_T (or epidemic threshold) of a pathogen and their work still serves as a template for infectious disease models (Dobson & Carper, 1996; Rothman *et al*, 2008; Diekmann *et al*, 2021). In short, a pathogen

with a high virulence will need larger population sizes to be able to continuously infect new hosts that have not succumbed to the disease or are immune to this pathogen (Dobson & Carper, 1996; Rothman *et al*, 2008). Therefore, H_T is inversely correlated to R_0/R_t.

Besides the threshold of establishment, other transmission models were created, for example to predict the course of an epidemic/pandemic, and to estimate the required public healthcare capacities and the effectiveness of interventions like social distancing, school closures or travel restrictions. Many of these models divide individuals of the population into transmission-related states like “susceptible”, “infectious”, “recovered” or “deceased” and accurate prediction of transition probabilities depends on the quality of the available epidemiological data (Rothman *et al*, 2008).

1.2. Epidemiology and clinical characterization of SARS-CoV-2 and COVID-19

In December 2019, a new respiratory disease called COVID-19 caused several patient clusters of viral pneumonia in the city of Wuhan in the Hubei province in China (Zhu *et al*, 2020; Wu *et al*, 2020; Zhu *et al*, 2020; Morens & Fauci, 2020; Hu *et al*, 2021). The novel coronavirus identified as causative agent for the disease was first termed 2019-nCoV and later SARS-CoV-2 (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses *et al*, 2020). Within only 3 months, SARS-CoV-2 spread across the globe and developed into a pandemic (World Health Organization, 2020d).

1.2.1. History of coronaviruses

Coronaviruses are positive-sense single-stranded RNA viruses that cause disease in humans and animals (Cui *et al*, 2019; V'kovski *et al*, 2021). They are members of the order of *Nidovirales*, the family of *Coronavirinae* and the subfamily of *Coronaviridae.* The subfamily of *Coronaviridae* is further divided into four genera based on phylogeny and genome organisation: *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* and *Deltacoronavirus.* All alpha- and betacoronaviruses that were identified as of today infect exclusively mammals, while gamma- and deltacoronaviruses have a wider host range (Cui *et al*, 2019; V'kovski *et al*, 2021). The four common human coronaviruses HCoV-NL63, HCoV-229E, HCoV-OC43 and HCoV-HKU1 are considered part of the seasonal respiratory viruses that cause the common cold and spread during winter months (see Figure 2) (Cui *et al*, 2019; Moriyama *et al*, 2020; V'kovski *et al*, 2021).

The first coronavirus was found in newborn chicks as a disease that caused an acute, fatal respiratory disease (Schalk & Hawn, 1931). Five years later, the causative agent for this disease was isolated, identified and named infectious bronchitis virus (IBV, later Avian coronavirus) (Beach & Schalm, 1936). In the following two years, more coronaviruses were identified in other animals, with a different tissue tropism causing other disease phenotypes than respiratory diseases. A coronavirus that causes gastrointestinal disease in pigs, the transmissible gastroenteritis coronavirus (TGEV), was identified in 1946 (Doyle & Hutchings, 1946). In fact, gastrointestinal symptoms are the typical result of most coronavirus infections in different animals. Other examples for coronaviruses that cause gastrointestinal symptoms like enteritis are bovine coronavirus (BCV), feline coronavirus (FCoV), canine coronavirus (CCoV) and turkey coronavirus (TCV) (Binn *et al*, 1974; Bridger *et al*, 1978; Pedersen *et al*, 1984; Ismail *et al*, 2003). Another coronavirus deviating from the general observation of

gastrointestinal symptoms is murine hepatitis virus (MHV) which causes viral hepatitis in mice and was first described in 1949 (Cheever *et al*, 1949; Gledhill & Andrewes, 1951). It took until 1965 that the first human coronaviruses were identified in a young patient with respiratory illness and in medical students (Tyrrell & Bynoe, 1965; Hamre & Procknow, 1966). These viruses were named B814 and 229E. The latter was later renamed to HCoV-229E (Chazal, 2021). However, at the time of identification, these viruses were not considered to be related. Transmission electron microscopy studies on these viruses revealed their structural similarity and that they share a crown-like structure around the virus envelope as a common feature constituted by the projection of the spike protein on the surface of the virus (Almeida & Tyrrell, 1967). The name coronavirus was derived from this unique feature ("corona" in latin: crown) (Almeida & Tyrrell, 1967; Chazal, 2021). In recent decades, the identification of new coronaviruses continued with the discovery of HCoV-NL63 in 2004 and HCoV-HKU1 in 2005 (van der Hoek *et al*, 2004; Woo *et al*, 2005).

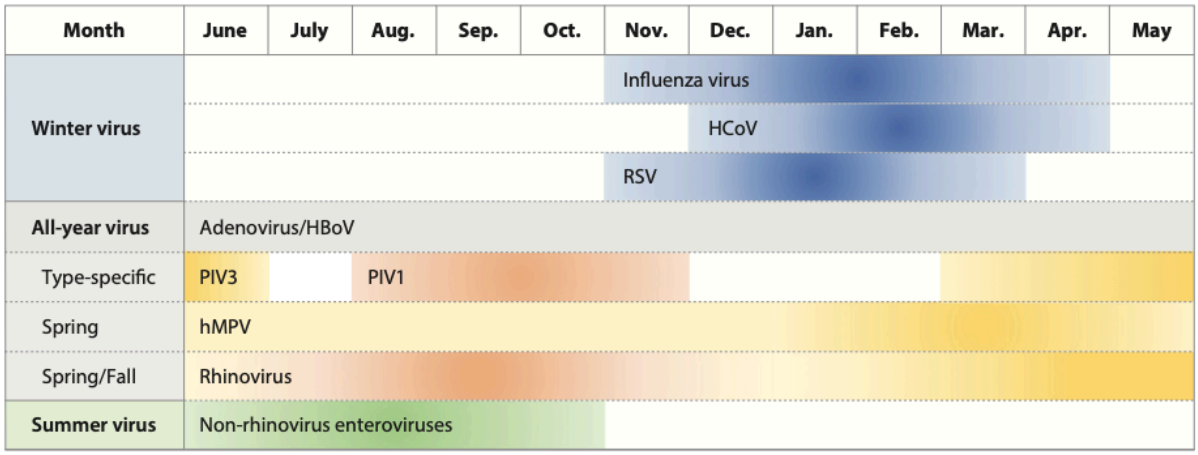
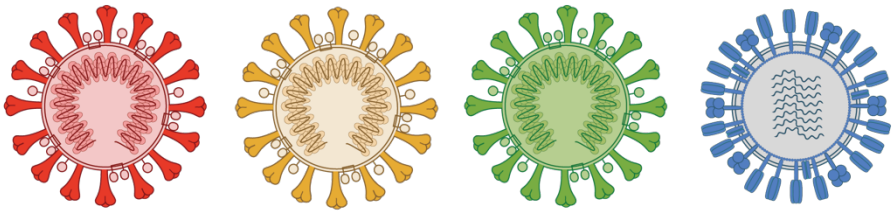


Figure 2: Schematic representation of the seasonality of common respiratory viruses in regions with temperate climate in the northern hemisphere. Viruses causing the common cold like Influenza viruses, common human coronaviruses (HCoV) and respiratory syncytial virus (RSV) are depicted as "winter viruses" and show the opposite seasonality of "summer viruses" like non-rhinovirus enteroviruses. Viruses like adenovirus and human bocavirus (HBoV) have no seasonality and appear all year. Human parainfluenza virus (PIV) strains, human metapneumovirus (hMPV) and rhinovirus are also depicted as all-year viruses, however, with peaks in their appearance. Figure used with permission of Annual Reviews, Inc., obtained from (Moriyama *et al*, 2020).

In immunocompetent humans, infections with coronaviruses mainly involve the upper respiratory and gastrointestinal tract and are considered a variant of the common cold with a mild, self-limiting course of infection. Severe cases in very young, elderly, or immunocompromised patients can involve bronchitis and pneumonia with renal symptoms (van der Hoek, 2007; Su *et al*, 2016). However, the notion that coronaviruses cause rather mild disease in average individuals changed when the two novel coronaviruses Severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome

coronavirus (MERS-CoV) emerged in 2002 and 2012, respectively (see Figure 3) (Cui *et al*, 2019).



	SARS-CoV-2	SARS-CoV	MERS-CoV	1918 Influenza
Genus	Betacoronavirus	Betacoronavirus	Betacoronavirus	Orthomyxoviridae
Reproductive Number (R_0)	2.5 (08/2020)	2.4	0.9	2
Incubation period (days)	4 - 12	2 - 7	5.2[1]	Unknown
Hospitalization rate	20% (08/2020)	>70%	Most[1]	Few
Case fatality rate	~1% (03/2023)[2]	9.7%	34%	Unknown
Total deaths	>6.8 Million (03/2023)[2]	774[3]	858[3]	50-100 Million[4]

Figure 3: Comparison of characteristics of highly pathogenic coronaviruses and 1918 Influenza. Clinical characteristics of the respiratory viruses were obtained from (Petersen *et al*, 2020) if not indicated otherwise. [1](Zumla *et al*, 2015), [2](Dong *et al*, 2020), [3](Deng & Peng, 2020), [4](Morens *et al*, 2008)

In November 2002, workers of a restaurant that processed wild animals for exotic food in the Chinese Guangdong province presented to healthcare facilities with high fever and mild respiratory symptoms that rapidly developed into atypical pneumonia (Zhong *et al*, 2003; Guan *et al*, 2003). A new coronavirus SARS-CoV was quickly identified as the causative agent for this disease (Peiris *et al*, 2003b; Ksiazek *et al*, 2003; Drosten *et al*, 2003). Infected individuals had an incubation period of about 5 days and developed symptoms within 13 days of exposure (Leung *et al*, 2004). SARS-CoV infection caused symptoms commonly reported for respiratory infections like fever, chills, coughing, myalgia, and headache but also gastrointestinal symptoms like diarrhea, vomiting and nausea (Peiris *et al*, 2003a; Lee *et al*, 2003; Peiris *et al*, 2003b; Hsu *et al*, 2003). In case of a fatal course of infection, the mean time between onset of symptoms and death was about 23 days. About half of all cases were the result of nosocomial transmission in hospitals, clinics, and nursing homes (Leung *et al*, 2004). About 20 to 30% of all patients required intensive care and mechanical ventilation (Lee *et al*, 2003; Peiris *et al*, 2003a). In the following months, this emerging infectious disease spread across China and subsequently to Hong Kong, Vietnam and more than 30 other countries (Guan *et al*, 2003; Petersen *et al*, 2020; Baker *et al*, 2022).

A fundamental part of the research response focused on identifying the natural and intermediate hosts of SARS-CoV that led to the zoonotic transmission of the virus from animals to the first human cases. Since the first cases were restaurant workers and associated with a live-animal market that sells wild animals for culinary purposes, the search for the origin of SARS-CoV concentrated on the animals available at this market. Virus isolates closely related to the human SARS-CoV isolates were obtained from a variety of healthy-presenting animals, including Himalayan palm civets and raccoon dogs marketed at the Guangdong live-animal market (Guan *et al*, 2003). A broader search for animals harboring these viruses yielded that these SARS-CoV-like viruses could not be found in farmed animals or wild civets (Kan *et al*, 2005). The analysis of circulating coronaviruses in wild animal populations showed that SARS-CoV was likely transmitted to these market animals from other wild animals, suggesting that the animals at the Guangdong market were only intermediate hosts for the zoonotic transmission. This led to massive culling of palm civets as suspected intermediate host for SARS-CoV transmission in order to prevent the reintroduction of the virus into the human population (Kan *et al*, 2005; Cui *et al*, 2019). However, the exact origin of SARS-CoV remained elusive. Surprisingly, genome sequencing of this novel coronavirus showed that SARS-CoV had only moderate similarity to other human coronaviruses like HCoV-OC43 and HCoV-229E and that it is rather distantly related to all other groups of known coronaviruses (Marra *et al*, 2003). This led to the suggestion that SARS-CoV constitutes a new group of coronaviruses. Viral sequencing in bat populations revealed later that bats are natural reservoirs for SARS-CoV-like viruses. The coronaviruses identified in these bat populations were phylogenetically closely related to those coronaviruses that were found in humans and civets at the time of the SARS-CoV outbreak, indicating that bats could have been the origin of the SARS-CoV outbreak in the Guangdong province (Li *et al*, 2005; Cui *et al*, 2019; Morens & Fauci, 2020; Baker *et al*, 2022).

The response to the outbreak of SARS-CoV was organized by the WHO and comprised five major aims (Anderson *et al*, 2004; Baker *et al*, 2022). First, the identification of the etiological agent for the disease and its intermediate and natural hosts that caused the first zoonotic transmissions to humans had priority. Second, another key achievement to gain control over the SARS-CoV outbreak was the development of serological and real-time polymerase chain reaction (RT-PCR) tests for early and reliable detection of the virus in patients. Third, effort was put towards the clinical characterization of the virus to reduce morbidity and mortality through specified medical treatment protocols. Fourth, key parameters of the epidemiology of SARS-CoV were determined to understand how the virus spreads. Finally, public health measures were designed based on knowledge from the epidemiological data in order to contain the disease. The success of these efforts was supported by the fact that individuals

infected with SARS-CoV only became infectious a few days after initial signs of symptoms (Peiris *et al*, 2003a; Anderson *et al*, 2004; Leung *et al*, 2004; de Wit *et al*, 2016). This allowed early isolation of infected individuals and a reduction of the effective reproduction number. Due to public health measures and mitigation strategies, the SARS-CoV epidemic ceased in June 2003 after 8098 global confirmed cases and 774 deaths with a mortality of 9.7% (Petersen *et al*, 2020; Anderson *et al*, 2004).

In April 2012, another coronavirus, MERS-CoV, emerged on the Arabian Peninsula and in Northern Africa (Zaki *et al*, 2012; Hijawi *et al*, 2013; Azhar *et al*, 2014; de Wit *et al*, 2016). Infected patients presented with disease symptoms in the lower respiratory tract, rhinorrhea, cough, and severe bilateral pneumonia, but they could also remain asymptomatic (Zaki *et al*, 2012; Azhar *et al*, 2014; Haagmans *et al*, 2014; Alfaraj *et al*, 2019). One of the first cases of MERS-CoV succumbed to the disease due to acute pneumonia and kidney failure (Zaki *et al*, 2012). As the virus spread more, it became clearer that the clinical characteristics of this new disease in terms of incubation period and symptoms were very similar to those of SARS-CoV (Zumla *et al*, 2015; Korea Centers for Disease Control & Prevention, 2015; Choi *et al*, 2016). However, the severity of the disease was increased to 50 to 89 % of patients who required intensive care with or without subsequent mechanical ventilation (Assiri *et al*, 2013).

The closest although distant phylogenetic relatives of MERS-CoV were found to be the bat coronaviruses HKU4 and HKU5 (Zaki *et al*, 2012). However, genetic analysis of samples from patients and dromedary camels revealed that MERS-CoV was the product of direct zoonotic transmission from camels to humans and that there was very little genetic difference between the viruses found in camels and humans (Azhar *et al*, 2014; Perera *et al*, 2013; Chu *et al*, 2018). In fact, many cases were related to direct zoonotic transmission from camels with evidence for multiple introduction events into the human population or nosocomial infections (Hijawi *et al*, 2013; Perera *et al*, 2013; Azhar *et al*, 2014; de Wit *et al*, 2016). Serological analyses showed that dromedary camels are a natural reservoir for MERS-CoV-like viruses, and these are expected to have already circulated since 1983 in these animal populations (Müller *et al*, 2014). This suggests that MERS-CoV-like viruses crossed the species barrier between bats and dromedary camels already a long time ago and thus created a virus reservoir from which MERS-CoV could emerge later and again cross the species barrier to humans.

Through travel, MERS-CoV spread to at least 27 countries and led, depending on the methodology of surveillance, to between 1,728 to 2,494 confirmed cases with 624 or 858

deaths, respectively (de Wit *et al*, 2016; Alfaraj *et al*, 2019; Chazal, 2021). Therefore, with 35% MERS-CoV had an even higher case fatality rate than SARS-CoV previously.

In summary, coronaviruses are considered part of the viral diseases causing the common cold in humans during winter months. They are viruses that are endemic in the human population and enzootic in the animal population (Cui *et al*, 2019; Moriyama *et al*, 2020). However, SARS-CoV and MERS-CoV are two examples of new coronaviruses that managed to cross the species barrier between their animal hosts and humans and caused disease epidemics with much more severe courses of infection than the common human coronaviruses. The past epidemics have shown that rapid research responses in science and clinics and concerted international public health efforts could successfully contain the outbreaks. Moreover, much could be learned about the clinical treatment of severe coronavirus diseases during these epidemics. In lack of standardized treatment plans at the time of the outbreaks, a variety of treatments, from host and viral protease inhibitors to interferons, ribavirin and lopinavir, were tested to reduce morbidity and mortality in patients (Al-Tawfiq *et al*, 2014; Choi *et al*, 2016; de Wit *et al*, 2016). Moreover, the development of vaccines for SARS-CoV and MERS-CoV was quickly started after their emergence. However, due to the short time spent on development and successful limitation of the epidemics, these vaccines did not reach the clinical testing phase or approval by medical agencies (Roper & Rehm, 2009; de Wit *et al*, 2016; Cui *et al*, 2019).

1.2.2. Epidemiology of SARS-CoV-2 and COVID-19

On December 31st, 2019, the Wuhan Municipal Health Commission reported several cases of pneumonia of unknown etiology. Patients infected with the novel coronavirus SARS-CoV-2 presented to hospitals in the Chinese city Wuhan (Zhu *et al*, 2020; Huang *et al*, 2020; Wu *et al*, 2020). Within a short time, several patient clusters emerged and could be epidemiologically traced back to the Huanan wholesale seafood market that also sells live animals (Zhu *et al*, 2020; Huang *et al*, 2020; Wu *et al*, 2020; Zhou *et al*, 2021). Not just direct exposure at the seafood market caused patient clusters, but in a next stage person-to-person transmission, e.g., in family clusters and via nosocomial transmission, played a role in the distribution of the virus (Chen *et al*, 2020; Chan *et al*, 2020; Deng & Peng, 2020). The Chinese government rapidly responded to contain the virus with epidemiological measures including contact tracing and strict containment measures like putting the entire city of Wuhan under quarantine to shut down all travel from and to the city by January 23rd 2020.However, increased travel activity between cities due to the approaching Chinese lunar New Year facilitated further spreading of the virus into other Chinese cities, so that all Chinese provinces confirmed cases of SARS-

CoV-2 infection within one month after first reports of the disease (Hu *et al*, 2021). Following these events and the fast spread of the virus through the population, the WHO recognized the outbreak of SARS-CoV-2 as a public health emergency of international concern (PHEIC) by January 31st, 2020 (World Health Organization, 2020d). The Chinese government reacted to the extended spread of COVID-19 with stricter measures like “lockdowns” (Hu *et al*, 2021).

There is evidence that the lockdown measures introduced by the Chinese government reduced the number of new infections and that measures like social distancing and mask mandates were adequate to reduce the risk of person-to-person transmission (Chu *et al*, 2020; Hu *et al*, 2021). Nevertheless, in the months following first recognition, human travel spread SARS-CoV-2 over the globe (see Figure 4).

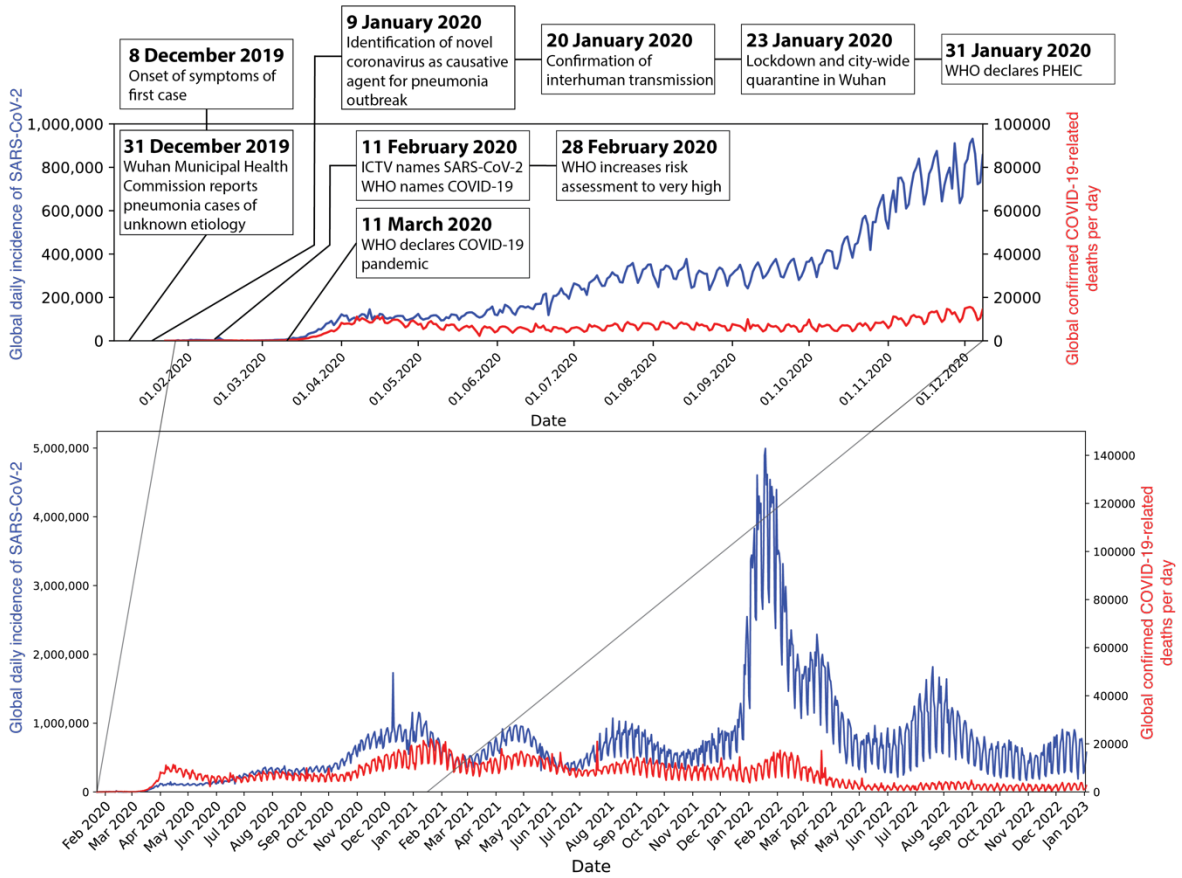


Figure 4: Timeline and key events of COVID-19 pandemic. Graphs show the global number of confirmed daily cases of SARS-CoV-2 (blue) and COVID-19-related deaths (red) between 08.12.2019 and 10.12.2020 (top) or between 22.01.2020 and 01.01.2023 (bottom). Graphs were generated with data obtained from (Dong *et al*, 2020; accessed: 21.02.2023).The timeline of key events in the early stages of the SARS-CoV-2 outbreak were adapted from (Hu *et al*, 2021).

Several studies that investigated local and national outbreaks highlighted the importance of global travel activity for the spread of SARS-CoV-2: One of the first introduction events to Europe occurred early during the pandemic on January 20th, 2020, and created a small cluster

in Bavaria, Germany which was extensively studied (Böhmer *et al*, 2020; Rothe *et al*, 2020). This introduction event was followed by several further events of diverse origins that created clusters in other European countries like France, Italy and Denmark (Bernard Stoecklin *et al*, 2020; Bluhm *et al*, 2020; Tuite *et al*, 2020). At this time, the pandemic reached a state where international transmission took over and introduction events could not be directly linked to travel from China anymore. Using a combination of viral genome sequencing and epidemiology, several studies from the United States showed that the spread of SARS-CoV-2, e.g., to New York, Northern California and other US states could be linked to European and other international clusters (Gonzalez-Reiche *et al*, 2020; Deng *et al*, 2020; Zeller *et al*, 2021). Based on epidemiology and mutation profiles in the viral genome, introduction events to several countries on the African continent were also shown to be linked to European clusters (Wilkinson *et al*, 2021). Upon these first intercontinental introduction events, SARS-CoV-2 continued to transmit further across the continents, e.g., between African countries or US states via domestic travel with air travel being a major driver of disease spread (Wilkinson *et al*, 2021; Fauver *et al*, 2020; Zeller *et al*, 2021). Compared to SARS-CoV, SARS-CoV-2 showed increased person-to-person transmission and can be transmitted before the onset of symptoms and by asymptomatic individuals, which causes difficulties for epidemiological contact tracing (Rothe *et al*, 2020; Baker *et al*, 2022). The outbreak of SARS-CoV-2 across the globe could not be stopped, although many countries imposed unprecedented countermeasures like strict limitations of social life, contact even between family members, stay-at-home orders, and social and economic lockdown strategies.

In the beginning of 2020, several introduction events brought SARS-CoV-2 to Austria. The first documented case of SARS-CoV-2 was connected to business travel from China, where the patient got infected between January 20th to 22nd (Kreidl *et al*, 2020). Through thorough epidemiological analysis, further cases between January and March 2020 in Vienna and a ski resort in Tyrol could be linked to international touristic travel (Kreidl *et al*, 2020; Popa *et al*, 2020). The Austrian efforts in diagnostic testing and epidemiology via contact tracing and viral genome sequencing did not just allow monitoring of the number of infected individuals among the population, but also to efficiently dissect and distinguish infection clusters (Popa *et al*, 2020; Kreidl *et al*, 2020; Leber *et al*, 2021). The lack of international institutional collaboration and international standards in acquiring epidemiological data was an obstacle for the identification of international transmission routes of SARS-CoV-2. However, with vast diagnostic testing, comprehensive epidemiological datasets, and genomic data of virus samples from identified infected individuals, these cross-border tracings worked in some cases. Based on this data, Icelandic clusters could be clearly linked to an outbreak in a ski

resort in the Austrian Alps which later turned out to have contributed significantly to the spread of SARS-CoV-2 across Europe (Gudbjartsson *et al*, 2020; Popa *et al*, 2020).

An advantage for monitoring the spread of SARS-CoV-2, compared to the previous outbreaks of SARS-CoV and MERS-CoV, was the broader availability of viral genome sequencing. Sequencing of patient samples allowed to connect clusters in different countries, where epidemiological data from contact tracing would not be enough. Upon sequencing of the first viral genome of SARS-CoV-2 in record time from the first cases in Wuhan, the wild-type sequence of the virus was uploaded to the database of the global initiative on sharing all influenza data (GISAID) (Shu & McCauley, 2017; Wu *et al*, 2020). Following that, SARS-CoV-2 international sequencing programs submitted viral genome sequences mainly from patients but also from other sources like animals and environmental samples to contribute to the genomic epidemiology of SARS-CoV-2 (Shu & McCauley, 2017; The Lancet, 2021; Burki, 2021). As of April 2023, this amounted to a total of over 15 million publicly available SARS-CoV-2 genome sequences (Shu & McCauley, 2017). The availability of SARS-CoV-2 genome sequences from all over the world allowed tracing the spread of the virus based on matching mutational profiles in the viral genome, but also to identify links between clusters in different countries, a method that is called genomic epidemiology (Popa *et al*, 2020; The Lancet, 2021; Burki, 2021; Baker *et al*, 2022). Examples that demonstrate the power of genomic epidemiology in dissecting and distinguishing infection clusters will be presented in the results section of this thesis. Specifically, various infection clusters in Austria at the onset of the pandemic in early 2020 will be examined. A variety of other examples from other countries also illustrate how viral sequence comparisons helped to trace the origin of introduction events. For example, the D614G mutation appeared in the first European infection clusters, like the outbreak in Bavaria, and it was rapidly established as the dominant variant in Europe, meaning that introduction events, e.g. to North America could be traced back to European origin (Böhmer *et al*, 2020; Korber *et al*, 2020; Gonzalez-Reiche *et al*, 2020; Plante *et al*, 2021). Tracing virus spread through mutation profiles poses a challenge since the same mutations in the viral genome may arise independently at different times, rather than in a sequential manner where one “founder” becomes the root for the emergence of new subvariants. This is particularly true for mutations that provide evolutionary selection advantages (DeGrace *et al*, 2022; Carabelli *et al*, 2023; Escalera *et al*, 2022).

1.2.3. Variants of SARS-CoV-2

Viral genome sequencing was not just a powerful tool for genomic epidemiology. Moreover, it developed into a cornerstone for monitoring the mutational trajectory of SARS-CoV-2 during

its sweep through the population. This allowed rapid creation of evidence for the identification of new virus variants with advanced characteristics compared to the wild type. Furthermore, mutations in the viral genome can change antigens of the virus. Thereby, the virus might evade the humoral and cellular response of acquired immunity in previously infected or vaccinated individuals, a mechanism called immune evasion. Consequently, new variants have the potential to diminish the efficacy of vaccines and drug treatments (Gupta, 2021; Escalera *et al*, 2022; DeGrace *et al*, 2022; Kent *et al*, 2022; Carabelli *et al*, 2023). The titer of neutralizing antibodies is still discussed as best candiate for the correlate of protection, meaning higher antibody titers are suggested to be the best predictor for protection against symptomatic and fatal courses of infection. As a result, mutations in epitopes that permit the virus to evade the antibody response are considered to pose a potential hazard to the public health response to SARS-CoV-2 (Khoury *et al*, 2021; Earle *et al*, 2021; Feng *et al*, 2021; Gupta, 2021; DeGrace *et al*, 2022; Escalera *et al*, 2022; Carabelli *et al*, 2023). It was also shown that altered antigenicity of SARS-CoV-2 variants with nonsynonymous mutations in MHC-I-restricted $CD8^+$ T cell epitopes resulted in decreased proliferation and effector function of epitope-specific cytotoxic T cells (Agerer *et al*, 2021; Kent *et al*, 2022).

Phylogenetic reconstruction based on genetic diversity of an organism or virus is a commonly used tool to infer relationships between different species and model their ancestry. These aspects of SARS-CoV-2 will be discussed in a later subchapter about its putative origin and phylogeny. However, standardized approaches to classify genetic diversity of viruses below the virus species level are lacking (Rambaut *et al*, 2020; Alm *et al*, 2020). Loosely defined terms like “variant”, “strain”, “subtypes”, “genotypes” and “groups” are generally used to refer to viruses with specific mutations in the viral genome that distinguish them from the wild type. These terms can also refer to “clades” that describe a monophyletic group on a phylogenetic tree (Rambaut *et al*, 2020; International Committee on Taxonomy of Viruses (ICTV), 2022). Clades are defined as groups of taxa that originated from a common ancestor. Monophyletic groups are defined by multiple specific properties that distinguish them from other organisms (International Committee on Taxonomy of Viruses (ICTV), 2022). Therefore, the main aim for detecting new variants of SARS-CoV-2 was early recognition of viral variants with mutations that confer either a selection advantage over the wild type or previous ancestral variants in terms of transmissibility, infectivity, viral replication, etc., or show changes in the clinical characteristics of the disease they cause (Thomson *et al*, 2021; Escalera *et al*, 2022; Liu *et al*, 2022; DeGrace *et al*, 2022; Carabelli *et al*, 2023).

An international convention on the nomenclature for the genetic diversity of SARS-CoV-2 did not exist in the beginning of the pandemic. This led to the establishment of different

nomenclatures that were curated in parallel (Alm *et al*, 2020). One of the first nomenclatures was introduced by GISAID and based on the Phylogenetic Clustering by Linear Integer Programming (PhyCLIP) algorithm (Han *et al*, 2019; Global Initiative on Sharing All Influenza Data (GISAID), 2021). Another nomenclature was established by Nextstrain, a platform for monitoring and visualization of geographic spread and genetic diversity of pathogens (Hadfield *et al*, 2018). The Nextstrain nomenclature combined manual curation and a set of simpler rules that were steadily adjusted to label genetically well-defined clades with significant frequency and geographic spread, and to provide memorable names for discussion of outbreak events (Hodcroft *et al*, 2021). These two nomenclature models are more focused on criteria for minimum prevalence and persistence of variants with genetic diversity to define large “clades”. A computational tool for automatic SARS-CoV-2 nomenclature that became very popular in the literature was the Phylogenetic Assignment of Named Global Outbreak Lineages (PANGOLIN) (O’Toole *et al*, 2021). This tool implements the dynamic Pango nomenclature definition that was tailored to incorporate epidemiological evidence of geographical outbreaks and biological/evolutional evidence from a rapidly growing dataset of available genome sequences (Rambaut *et al*, 2020). This allowed a detailed description of large clades and small subclades at the same time and, thus, established the common variant nomenclature that was also adapted by the WHO and other international health institutions to communicate the emergence and spread of SARS-CoV-2 variants (European Centre for Disease Prevention and Control, 2020; World Health Organization, 2022).

One of the first variants that emerged during the pandemic was characterized by the D614G mutation in the spike protein and outperformed the wild type virus through higher viral fitness with enhanced infectivity and viral replication (Plante *et al*, 2021; Korber *et al*, 2020). Since the emergence of the D614G variant, several major variants have been identified worldwide. Some of them showed new characteristics so that they were categorized as variants of interest (VOI) or variants of concern (VOC) (European Centre for Disease Prevention and Control, 2020; World Health Organization, 2022). For public communication, the WHO implemented new labels for VOCs based on the Greek alphabet. The emergence of these virus variants had a tremendous impact on the characteristics of the virus and afforded adaptation of the countermeasures to contain the disease spread (Gupta, 2021; Carabelli *et al*, 2023).

The first VOC was recognized in September 2020 in the United Kingdom and was names “Alpha” (Pangolin lineage B.1.1.7) (Davies *et al*, 2021). It outcompeted previous SARS-CoV-2 variants through a higher reproduction number. The “Alpha” variant showed reduced susceptibility to interferon-inducible innate immune responses, another mechanism of immune evasion (Thorne *et al*, 2022; Liu *et al*, 2022). At the same time, “Beta” (Pangolin lineage

B.1.351) was first found in South Africa with a higher transmissibility and severity based on epidemiological and clinical data (Funk *et al*, 2021). These variants were quickly outperformed by the variant “Delta” (Pangolin lineage B.1.617.2) that was first described in India in December 2020 and spread to 43 countries on six continents in less than six months (Lopez Bernal *et al*, 2021). The emergence of “Delta” was accompanied by concerns about its altered antigenicity, which resulted in a reduced susceptibility to neutralizing antibodies and reduced effectiveness of common vaccines like the BNT162b2 and ChAdOx1 (Planas *et al*, 2021; Lopez Bernal *et al*, 2021; Carabelli *et al*, 2023). This raised the concern that the emergence of “Delta” could jeopardize the success of mass vaccine rollouts in many countries at that time (Gupta, 2021; Carabelli *et al*, 2023). Confirming these concerns, several studies showed later that different SARS-CoV-2 variants are less sensitive to neutralization by antibodies and can subvert the cell-mediated immune response (Gupta, 2021; DeGrace *et al*, 2022; Carabelli *et al*, 2023).

In November 2021, the variant “Omicron” was first detected in South Africa and Botswana and spread with an unprecedented pace within three weeks to over 87 countries, again outcompeting previous variants (Viana *et al*, 2022). Viral genome sequencing data from infected individuals showed that this variant carried over 30 mutations in the spike protein and suggested altered transmissibility and susceptibility to antibody neutralization (Viana *et al*, 2022; Carabelli *et al*, 2023). Nevertheless, although its reduced susceptibility to neutralization was shown, enough T cell epitopes without mutations remained to maintain functional T cell responses in vaccinated and previously infected individuals (Keeton *et al*, 2022). In late 2022, “Omicron” and its sub-lineages still caused high case numbers across many countries (DeGrace *et al*, 2022; Carabelli *et al*, 2023).

1.2.4. Clinical characterization of COVID-19

COVID-19 is an acute respiratory disease caused by SARS-CoV-2 infection. Reports of the symptomatic characteristics of the disease from the initial outbreaks in China range from asymptomatic courses of infection to severe pneumonia and death with 13.8% severe cases and 6.1% critical cases (respiratory failure, septic shock and multiple organ dysfunction or failure) (World Health Organization, 2020d; Hu *et al*, 2021). Typical signs and symptoms of COVID-19 develop on average about 5 days after infection and include fever, dry cough, fatigue, shortness of breath, sore throat, olfactory and taste disorders, and gastrointestinal symptoms like diarrhea (World Health Organization, 2020d; Huang *et al*, 2020; Chen *et al*, 2020; Wang *et al*, 2020; Petersen *et al*, 2020). Its clinical characteristics differ between age groups while individuals of all age groups are susceptible to the disease. An asymptomatic or

mild course of infection is more likely in individuals below the age of 50 (Hu *et al*, 2021; Wang *et al*, 2020; Chen *et al*, 2020; Huang *et al*, 2020; Petersen *et al*, 2020). About 40% of cases present an asymptomatic course of infection (Sette & Crotty, 2021). However, individuals over 60 years of age, especially in patients with co-morbidities like diabetes, are more prone to develop a severe or critical course of infection with dyspnea, acute respiratory distress syndrome, acute cardiac injury and multiple organ failure that requires hospitalization in intensive care units (Hu *et al*, 2021). Ground-glass opacity in radiological chest images is a common feature in patients presenting to the hospital. Another common immunological characteristic of COVID-19 is marked lymphopenia, especially in severe and critical cases that often develop more severe lymphopenia over time. Moreover, patients with severe and critical courses of infection show higher levels of plasma cytokines with immunopathology caused by cytokine storm (Wang *et al*, 2020; Huang *et al*, 2020; Hu *et al*, 2021).

The development of several vaccines and medical treatments started early during the pandemic (Krammer, 2020). The nucleoside-modified mRNA vaccine BNT162b2 (Comirnaty) showed an efficacy of over 95% in preventing COVID-19 and substantially decreased its mortality among age groups (Polack *et al*, 2020; Arbel *et al*, 2021; Bar-On *et al*, 2021). Another vaccine, based on an adenoviral vector, ChAdOx1 (Vaxzevria) was also used for broad vaccination of the population and showed an efficacy of 70.4% in preventing symptomatic COVID-19 in randomized controlled clinical trials (Voysey *et al*, 2021). Like most other SARS-CoV-2 vaccines, both target the viral S protein against which efficient neutralizing antibodies have been observed upon seroconversion following natural infection (Suthar *et al*, 2020; Piccoli *et al*, 2020; Earle *et al*, 2021; Khoury *et al*, 2021).

Different therapeutics were developed or adapted for the acute treatment of COVID-19. Among these are inhibitors of virus entry like recombinant angiotensin-converting enzyme 2 (ACE2), neutralizing antibodies against SARS-CoV-2 and inhibitors of the viral RNA-dependent RNA polymerase (RdRp) like ribavirin and remdesivir to suppress viral replication (Hu *et al*, 2021; Wang *et al*, 2020). The corticosteroid dexamethasone turned out to be an effective treatment by inhibiting the excessive inflammatory response and thereby alleviated the course of infection and reduced immunopathology (Hu *et al*, 2021; The COVID STEROID 2 Trial Group *et al*, 2021; The WHO Rapid Evidence Appraisal for COVID-19 Therapies (REACT) Working Group *et al*, 2020). With the population-wide vaccine programs and the emergence of new variants of SARS-CoV-2, the clinical characteristics of the disease changed, for example towards lower pathogenicity and higher transmissibility with the “Omicron” variant (Vihta *et al*, 2022; Viana *et al*, 2022). Vaccines and treatments largely maintained their overall, however slightly reduced efficacy in protecting from severe disease

but conferred only marginal protection to infection (Vihta *et al*, 2022; Wolter *et al*, 2022; Viana *et al*, 2022).

1.2.5. Post-acute COVID-19 symptoms

After clearance of the virus and recovery from acute COVID-19, 13% to 76% of patients can suffer from post-acute symptoms (Nalbandian *et al*, 2021; Sudre *et al*, 2021; Huang *et al*, 2020). This clinical manifestation was named post-acute COVID-19 syndrome, post-acute sequelae of SARS-CoV-2 infection (PASC), or generally termed “Long COVID”. Post-acute COVID-19 syndrome is characterized by a patient showing diverse symptoms like dyspnea and fatigue lasting for at least two, usually three months from the onset of COVID-19 (Sudre *et al*, 2021; Nalbandian *et al*, 2021; Merad *et al*, 2022). Moreover, multiple organ systems can be affected and extend the set of symptoms to cardiovascular, hematologic, renal, endocrine, gastrointestinal, and dermatologic symptoms with other more common symptoms being chest pain, cognitive disturbances, depression, anxiety and arthralgia (Nalbandian *et al*, 2021; Sudre *et al*, 2021). The pathophysiological mechanisms of post-acute COVID syndrome are still understudied but were found to be similar to long-term effects observed in survivors of SARS, MERS and H1N1 influenza (Liu *et al*, 2015; Suthar *et al*, 2020; Nalbandian *et al*, 2021). According to clinical studies, there is a correlation between the severity of acute COVID-19 and the development of post-acute COVID-19 syndrome, with 79% of individuals who experienced post-acute symptoms having been previously hospitalized during their acute infection (Merad *et al*, 2022). Additionally, patients over the age of 50 are at higher risk to develop long-lasting symptoms (Sudre *et al*, 2021). Different hypotheses for the causes of post-acute COVID-19 are currently under investigation. Leading study subjects are for example persistent virus infection or presence of viral antigens and RNA that could drive chronic inflammation, autoimmune reactions, dysbiosis of the microbiome, or uncleared remains and unrepaired tissue damage. It was found that pro-inflammatory cytokines like IL-6 (interleukin 6) and tumor necrosis factor-α (TNFα), but also interferons (IFNs) were elevated in patients with post-acute COVID-19 symptoms. Autoantibodies were not just found to be a factor involved in the severity of COVID-19 but correlated with the appearance of post-acute COVID-19 symptoms. Importantly, studies found that vaccination with two doses reduced the risk of developing Long COVID (Sudre *et al*, 2021; Merad *et al*, 2022). However, how vaccines protect from the establishment of this condition as well as its cause, are still understudied.

While healthy children were thought to have in general only mild courses of SARS-CoV-2 infection, they can be affected by the multisystem inflammatory syndrome in children (MIS-C) (Henderson & Yeung, 2021). It is characterized by prolonged inflammation and fever, and can

affect multiple organs like the skin, mucous membranes, the gastrointestinal tract, the respiratory and neurological system, and the heart (Henderson & Yeung, 2021; Merad *et al*, 2022). MIS-C is different from post-acute COVID syndrome and described as a post-infectious process with symptoms peaking between 3-6 weeks after the highest viral load during SARS-CoV-2 infection. One cause for the development of MIS-C could be that patients develop appropriate antibody responses to SARS-CoV-2 but show an enrichment of IgG and IgA autoantibodies directed towards antigens expressed on different tissues and immune mediators (Henderson & Yeung, 2021). More research on the long-term effects of COVID-19 but also other viral diseases is needed to address this public health issue and increase the quality of life of affected patients.

1.2.6. Antiviral immune responses against SARS-CoV-2

Higher organisms developed several defense mechanisms against invading microbes. Immune responses are coordinated reactions against molecular structures that are recognized as non-self, microbial, or malignant, and they consist of two main parts. The innate immune system provides a rapid first line of defense against pathogens, but its recognition of pathogen-specific structures is less flexible compared to the adaptive immune system. The adaptive immune system establishes a powerful pathogen-specific response with the establishment of an immunological memory (also called acquired immunity). Main components of the innate immune system are physical and chemical barriers like the skin and mucosal epithelia, blood proteins like the complement system, natural antibodies and soluble proteins that recognize glycans on cell surfaces. In addition, various cell types like phagocytes, dendritic cells and natural killer cells act in the innate immune response. The specificity of the innate immune system is limited to the recognition of general molecular patterns characteristic for microbes or damaged host cells (Akira *et al*, 2006; Vabret *et al*, 2020; Tay *et al*, 2020; Boechat *et al*, 2021; Sette & Crotty, 2021). However, it initiates a quick first stage immune reaction to combat a pathogen or malignant cell and prepares the environment for the adaptive immune response.

Main components of the adaptive immune system are antibodies, T lymphocytes and B lymphocytes (Tay *et al*, 2020; Sette & Crotty, 2021; Merad *et al*, 2022). The adaptive immune system is specialized in adapting to different antigens. These are diverse molecular structures presented by a particular pathogen or malignant cell with high specificity. Adaptive immune responses increase in magnitude and specificity with every successive challenge by an infectious agent and they allow the host to eradicate even those pathogens that evolved to circumvent innate immunity. Immune cells exert systemic immune surveillance by circulating through the organism. Several feedback loops of positive and negative regulation control the

strength of the immune reaction to keep the balance between successful clearance of the pathogen or malignant cell and harmful reactions that cause “immunopathology” (Vabret *et al*, 2020; Tay *et al*, 2020; Boechat *et al*, 2021; Sette & Crotty, 2021; Merad *et al*, 2022). In the following chapter, the general principles of innate and adaptive immunity against viruses and the mechanisms that are relevant for SARS-CoV-2 and the severity of COVID-19 will be introduced.

Innate immune response against SARS-CoV-2

The innate immune system is equipped with a variety of pattern-recognition receptors (PRRs) that detect pathogen-associated molecular patterns (PAMPs) of RNA viruses like coronaviruses. Moreover, cytopathic viruses like SARS-CoV-2 induce cell death and tissue damage by pyroptosis. The lysis of respiratory epithelial cells because of SARS-CoV-2 infection leads to the release of molecules that can be detected as damage-associated molecular patterns (DAMPs) by PRRs (Akira *et al*, 2006; Tay *et al*, 2020; Boechat *et al*, 2021; Merad *et al*, 2022).

PRRs are expressed across different cell types and allow target cells to detect intrinsic infection or specialized immune cells to facilitate immune surveillance of the environment (Akira *et al*, 2006; Tay *et al*, 2020; Madden & Diamond, 2022). Alveolar macrophages are specialized cells of the innate immune system that survey the lumen of the respiratory tract for viral PAMPs and DAMPs and play an important role during SARS-CoV-2 infection (Boechat *et al*, 2021; Merad *et al*, 2022). Patients with severe disease symptoms were found to have reduced numbers of alveolar macrophages in their respiratory tract. Dendritic cells (DCs) are a central cell type of the innate immune system. They reside in tissues or circulate to facilitate systemic immune surveillance and express the largest arsenal of PRRs among all cell types. Upon recognition of microbes, different DC subsets will secrete factors to prepare the environment and they will act as antigen-presenting cells (APCs) facilitating a connection between innate and adaptive immune response. Conventional DCs (cDCs) secrete pro-inflammatory cytokines and capture antigens of the invading microbes to stimulate T cell responses. Plasmacytoid DCs (pDCs) are more specialized for antiviral immune. pDCs show high expression levels of antiviral PRRs, they are specialized in stimulating antiviral T cell responses and are the main producer of type I IFNs (Akira *et al*, 2006; Boechat *et al*, 2021; Merad *et al*, 2022). Besides phagocytes and DCs, natural killer cells (NK cells) play an important role in the recognition and killing of infected cells. Infected host cells can become susceptible to NK cell-mediated cytotoxicity by a variety of mechanisms like expression of ligands for activating NK cell receptors due to the cellular stress response (Björkström *et al*, 2022; Tay *et al*, 2020). Moreover, reduced class I major histocompatibility complex (MHC-I)

expression, as a result of a viral escape mechanism from adaptive immune responses, can lead to NK cell-mediated cytotoxicity (Björkström *et al*, 2022).

Toll-like receptors (TLRs) are one class of membrane-bound PRRs that are situated on the cell surface (TLR1,2,4-6 and 11) or in intracellular vesicles like endosomes (TLR3, 7-10). Among these 12 types of TLRs, six are relevant to detect viral PAMPs. TLR3 (dsRNA), TLR7 and TLR8 (guanosine/uridine-rich ssRNA) and TLR9 (DNA) detect viral genomes or their intermediates during replication (Akira *et al*, 2006; Madden & Diamond, 2022). pDCs have higher expression levels of TLR7 and TLR9 (McNab *et al*, 2015). With focus on SARS-CoV-2, it was shown that TLR7 and TLR8 recognize motifs in the SARS-CoV-2 S protein gene and are expressed in myeloid cells like alveolar macrophages that perform immune surveillance in the lumen of the respiratory tract. Moreover, TLR2 and TLR4 recognize viral glycoproteins like the SARS-CoV-2 envelope protein (Maris *et al*, 2006; Akira *et al*, 2006; Moreno-Eutimio *et al*, 2020; Zheng *et al*, 2021; Madden & Diamond, 2022; Merad *et al*, 2022).

As described above, the expression and secretion of pro-inflammatory cytokines and type I/III IFNs can be induced by TLR engagement. However, the type of activated signaling pathway depends on the TLR type (see Figure 5). PAMP detection leads to dimerization of TLRs and recruitment of adaptor proteins to their Toll/IL-1R homology (TIR) domain (Akira *et al*, 2006; Madden & Diamond, 2022). This process initiates a downstream signaling cascade that leads to the expression of pro-inflammatory cytokines and IFNs. Endosomal TLR3 acts via two distinct signaling pathways that start with the recruitment of TIR-domain-containing adaptor-inducing interferon-β (TRIF) to TLR3. Then, TRIF recruits TNFR-associated factor 6 (TRAF6) and receptor-interacting protein 1 (RIP1), which leads to the activation of nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB). NF-κB then translocates into the nucleus and acts as an inducer of pro-inflammatory cytokine expression. Moreover, TRIF also initiates the activation of a complex of TRAF-family-member-associated NF-kB activator (TANK)-binding kinase 1 (TBK1) and Inducible I kappa-B kinase (IKK-*i*), which will phosphorylate interferon regulatory factor 3 (IRF3) and 7 (IRF7). These also translocate into the nucleus, where IRF3 will induce expression of pro-inflammatory genes and both will induce type I and type III IFNs (Akira *et al*, 2006; Kim & Shin, 2021; Madden & Diamond, 2022).

TLR7 and TLR9 act via a different signaling pathway, which starts with the recruitment of myeloid differentiation primary response 88 (MyD88) to the TLR and subsequent formation of a protein complex comprising interleukin 1-receptor-associated kinase (IRAK) 1, IRAK4, TRAF6 and IRF7 (Akira *et al*, 2006). This complex induces phosphorylation of IRF7 and

translocation of NF-kB and phosphorylated IRF7 which leads to the induction of pro-inflammatory cytokines and IFNs (Akira *et al*, 2006).

Besides TLRs, a variety of other PRRs exist that can detect RNA viruses. For example, retinoic acid inducible gene I (RIG-I)-like receptors such as RIG-I and melanoma differentiation-associated protein 5 (MDA5) are cytosolic RNA sensors that recognize dsRNA or uncapped phosphorylated ssRNA during the replication of coronavirus genomes. Activation of RIG-I-like receptors induces signaling cascades that converge on the previously introduced pathways and lead to expression of type I/III IFNs and pro-inflammatory cytokines (Vabret *et al*, 2020; Merad *et al*, 2022; Madden & Diamond, 2022). However, for SARS-CoV-2 it was shown that RIG-I directly interfered with viral genome replication in lung epithelial cells without triggering cytokine production and type I/III IFN responses (Yamada *et al*, 2021).

Finally, the complex of cyclic GMP-AMP synthase (cGAS) and stimulator of interferon genes (STING) also induces production of IFNs upon recognition of cytosolic DNA, either from a pathogen, the nucleus, or mitochondria. The cGAS-STING pathway has two main steps. First, cGAS detects DNA and converts it into cyclic GMP-AMP (cGAMP). Second, the mitochondria-associated receptor STING detects cGAMP and initiates pro-inflammatory cytokine production and expression of IFNs via activation of TBK1 and IRF3 as described above. Arguably, as an RNA virus, SARS-CoV-2 neither carries DNA nor does it have cytosolic DNA intermediates. Therefore, it does not trigger the cGAS-STING pathway directly but can induce mitochondrial damage and, thus, a release of DNA into the cytosol (Lokugamage *et al*, 2020; Madden & Diamond, 2022). cGAS-STING agonists were found to be promising options for therapeutic strategies against SARS-CoV-2 infection for their ability to induce type I IFN responses (Humphries *et al*, 2021; Li *et al*, 2021).

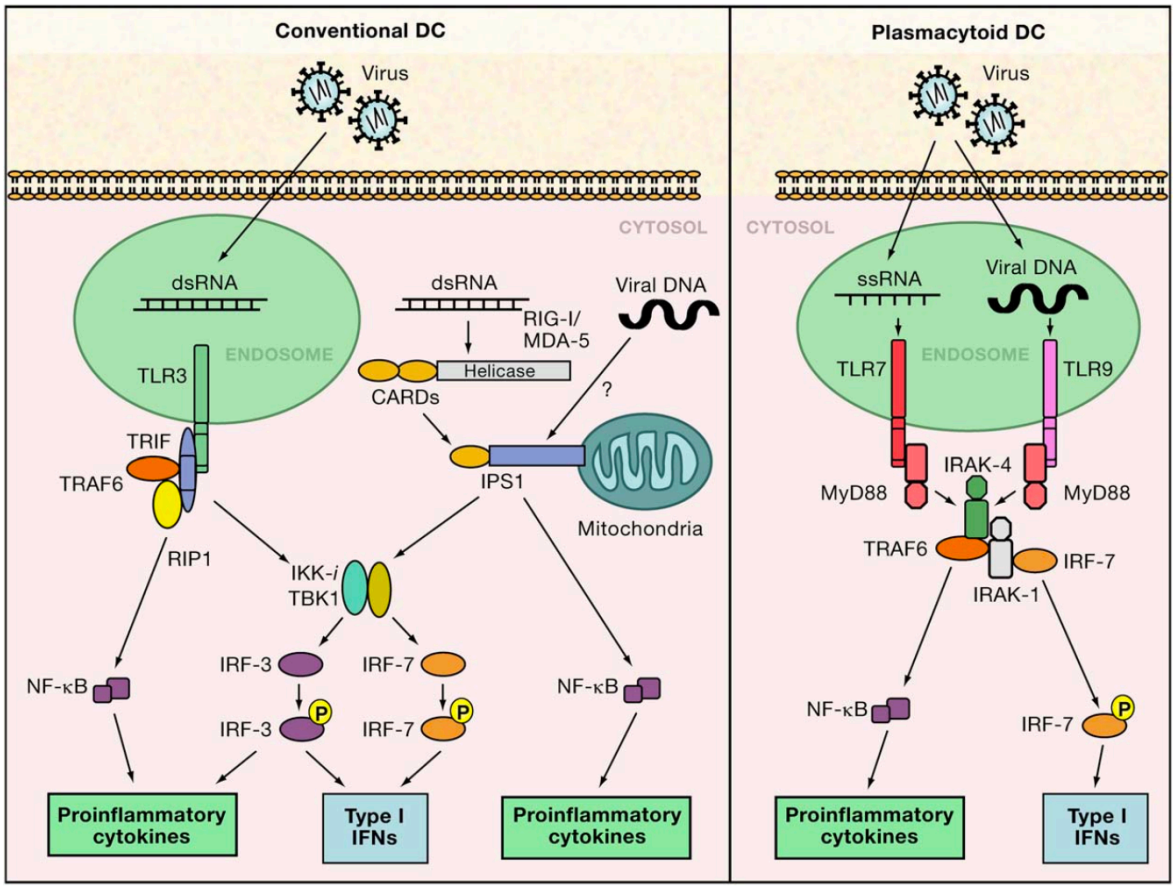


Figure 5: Pathways of PRR-mediated recognition of viral infection. Endosomal TLR3, cytoplasmic RIG-1 or MDA5 recognize genetic material of viruses or their intermediates in conventional DCs and induce expression of proinflammatory cytokines and type I IFNs via the RIP1/TRAF6-NF-kB and the TBK1/IKK-i-IRF-3/IRF-7 pathways (left). Plasmacytoid DCs detect viral ssRNA or DNA via endosomal TLR7 or TLR9 which trigger then complex formation of MyD88, IRAK-4, TRAF6, IRAK-1 and IRF-7 to induce expression of proinflammatory cytokines and type I IFNs via NF-kB and IRF-7. Figure obtained from (Akira *et al*, 2006). Reprinted with permission from Elsevier. License number: 5518190799690

The complement system comprises blood proteins, some of which act as soluble PRRs, initiate pro-inflammatory responses and induce phagocytosis upon encounter of PAMPs (Boechat *et al*, 2021). Mannose-binding lectin (MBL) was shown to bind to mannose-rich glycan structures on SARS-CoV virions and thereby limited its infectivity in experimental models (Zhou *et al*, 2010). Additionally, lower serum levels of MBL could be linked to higher susceptibility to SARS-CoV (Zhang *et al*, 2005). Site-specific glycan analyses in SARS-CoV-2 suggest that MBL could also bind to SARS-CoV-2 virions and interfere in the interaction between the viral spike protein and the target receptor ACE2 (Watanabe *et al*, 2020).

As described above, activation of endosomal or cytosolic PRRs like TLRs, RIG-like receptors or the STING pathway in host or immune cells trigger different signaling cascades that lead to the release of type I and type III IFNs. IFNs are cytokines that induce an antiviral state in both infected cells and uninfected bystander cells in an autocrine and paracrine fashion. They act on the target cell via induction of a transcriptional program with increased expression of IFN-

stimulated genes (ISGs) which interfere with viral replication at various stages (McNab *et al*, 2015; Mesev *et al*, 2019). Type III IFNs rather act locally on epithelial cells, e.g. of mucosal epithelia like the respiratory tract, and are thereby important antagonists of respiratory viruses like coronaviruses and influenza. On the other hand, type I IFNs rather act systemically on many different cell types (McNab *et al*, 2015; Mesev *et al*, 2019; Kim & Shin, 2021). These differences are defined by the receptor tropism. Type I IFNs signal via Interferon-alpha/beta receptor alpha chain and beta chain (IFNAR1 and IFNAR2) and type III IFNs are recognized by a complex of interferon lambda receptor 1 (IFNLR1) and interleukin 10 receptor subunit beta (IL-10Rβ). Despite employing distinct receptors, signaling of both types of IFNs converge on the same intracellular signaling pathway. Receptor engagement activates the receptor-associated tyrosine kinases janus kinase 1 (JAK1) and tyrosine kinase 2 (TYK2) which phosphorylate signal transducer and activator of transcription 1 and 2 (STAT1 and STAT2). These form the protein complex IFN-stimulated gene factor 3 (ISGF3) together with IRF9 and translocate into the nucleus where it binds to IFN-stimulated response elements (ISREs) to induce transcription of hundreds of ISGs. These ISGs mediate the inhibition of viral protein synthesis, degradation of viral RNA and inhibition of viral gene expression and virion assembly (Kim & Shin, 2021; Mesev *et al*, 2019).

Several proteins of SARS-CoV-2 interact with factors in the signaling cascades described above to evade immune detection and prevent expression of IFNs to which SARS-CoV-2 is susceptible (Xia *et al*, 2020; Lokugamage *et al*, 2020). For example, different viral proteins prevent the phosphorylation of TBK1, IRF3, STAT1 and STAT2 or the translocation of IRF3 and NF-kB to the nucleus (Xia *et al*, 2020; Vazquez *et al*, 2021; Hayn *et al*, 2021). Other strategies for interference with innate immune signaling comprise inhibition of immune receptor signaling, signaling via type I and type III IFN receptors and suppression of ISG function (Kim & Shin, 2021). Moreover, in a clinical study about 10% of patients with a life-threatening course of COVID-19 but none of the asymptomatic patients were found to have autoantibodies against IFNs (Bastard *et al*, 2020).

In conclusion, the innate immune system recognizes viral PAMPs or DAMPs via PRRs and initiates the expression of type I/III IFNs and pro-inflammatory cytokines. Pro-inflammatory cytokines like IL-1 and TNFα then prepare the endothelia for inflammation and stimulate chemokine production to attract immune cells to the site of inflammation. Together with IL-6, they mediate systemic effects like production of acute phase proteins and induction of fever. Type I/III IFN signaling also leads to the induction of antiviral transcription programs with the expression of ISGs that interfere at different steps with viral replication (Lazear *et al*, 2019; Mesev *et al*, 2019; Merad *et al*, 2022). However, SARS-CoV-2 developed several strategies

to evade the innate immune response by interfering with pathogen recognition and innate immune signaling. In fact, clinical studies have shown that patients with severe COVID-19 disease often showed elevated pro-inflammatory cytokines and an impairment in type I and type III IFN expression and signaling (Merad *et al*, 2022; Sette & Crotty, 2021). The virus is effective at delaying type I/III IFN responses which allows it to replicate without much interference in the early phase of infection (Sette & Crotty, 2021). Additionally, impaired pathogen recognition due to a depletion of alveolar macrophages in the respiratory tract and inefficient antigen presentation of APCs increase the likelihood for a more severe course of infection (Boechat *et al*, 2021; Merad *et al*, 2022). Such dysregulation in the innate immune response to SARS-CoV-2 can even enable virus persistence since the innate immune response builds the foundation for a powerful and well-balanced adaptive immune response that leads to successful viral clearance (Merad *et al*, 2022).

Adaptive immune response against SARS-CoV-2

Components of the innate immune system detect general pathogen- and damage-associated molecular structures and rapidly establish a first line of defense. However, many pathogens will not be fully eradicated by the innate immune response as they evolved mechanisms of innate immune evasion as described above for SARS-CoV-2 (Sette & Crotty, 2021). Moreover, innate immunity has only limited memory function that would allow a more efficient immune recall upon re-challenge with the same pathogen. Therefore, APCs of the innate immune system like dendritic cells take up molecular structures of the invading pathogens and present these to cells of the adaptive immune system which will develop an antigen-specific response to these pathogens. Moreover, the adaptive immune system has the capacity to create an antigen-specific immunological memory (Hilligan & Ronchese, 2020). This is crucial for long-lasting protection against many pathogens and is the molecular foundation for the efficacy of vaccinations (Qi *et al*, 2022). The adaptive immune response against viruses is based mainly on two components: a humoral response and a cellular response. Based on different mechanisms, both arms of the adaptive immune response generate long-lasting pathogen-specific immunity with memory function (Sette & Crotty, 2021). The following paragraphs will introduce the adaptive immune system with specific focus on the relevant components for the antiviral immune response to SARS-CoV-2 and severity of COVID-19.

Antibodies are part of the humoral arm of the adaptive immune response which is initiated by B lymphocytes, also called B cells. Upon antigen encounter, activation, and maturation in the draining lymph nodes, B cells differentiate into plasma cells and produce pathogen-specific antibodies. This process can be split into two phases that yield antibodies with different grades of affinity to the pathogen antigen and different compositions of antibody isotypes (Qi *et al*,

2022; Sette & Crotty, 2021). During the extrafollicular phase that takes less than a week, B cells quickly differentiate into plasma cells that provide mainly IgM type antibodies with high enough affinity and avidity to neutralize the virus. However, these plasma cells are short-lived. Therefore, during the germinal center phase that can take up to a month, somatic hypermutation and affinity-based maturation will give rise to a long-living bone marrow-located compartment of plasma cells with higher affinity to the virus. Both phases generate memory B cells that will persist long after primary challenge with the pathogen. Antibodies confer protection against pathogens via opsonization, neutralization, initiation of complement-mediated lysis or antibody-dependent cellular cytotoxicity (ADCC) (Sette & Crotty, 2021; Qi *et al*, 2022). Among the forementioned, antibody-mediated neutralization plays one of the most important roles in antiviral immunity (Khoury *et al*, 2021; DeGrace *et al*, 2022).

Neutralizing antibodies for SARS-CoV-2 mainly target the N and S protein and for the latter especially the receptor-binding domain (Khoury *et al*, 2021; Merad *et al*, 2022). The presence of neutralizing antibodies for SARS-CoV-2 is an important predictor for the severity of the course of infection with severe COVID-19 cases presenting with higher antibody titers than those with mild or asymptomatic symptoms (Dan *et al*, 2021; Chen *et al*, 2020; Long *et al*, 2020; Khoury *et al*, 2021). These vaccine- or infection-induced neutralizing antibodies confer protection against reinfection against the same variant of the virus, but the protective effect partially extends to other variants of the virus (Khoury *et al*, 2021). The majority of COVID-19 patients seroconvert within 2 weeks upon infection, meaning antibodies can be detected in their serum, and the concentration of neutralizing antibodies peaks three to four weeks after onset of symptoms and are detectable for more than eight months (Qi *et al*, 2022; Dan *et al*, 2021). However, asymptomatic cases became more often (~40% of cases) seronegative in the early convalescent phase than symptomatic cases (~13% of cases) (Long *et al*, 2020). Especially important to prevent re-infection with a respiratory virus that enters via mucosal epithelia of the respiratory tract is that IgA type antibodies can be detected for example in nasal fluids (Qi *et al*, 2022). The humoral response does not just contribute to long-lasting protection and prevention of reinfection. It was also found that patients with sustained antibody production and higher antibody titers recovered much faster from COVID-19 symptoms than those individuals with lower antibody titers (Chen *et al*, 2020).

T Lymphocytes, also called T cells, are the core of the cell-mediated immune response and they combat intra- and extracellular pathogens as well as tumor cells. T cells can be divided into the two major classes $CD4^+$ T helper cells (T_H cells) and $CD8^+$ T cells. T_H cells can be further divided into subsets like regulatory T cells (T_{reg} cells), $T_{H1,}$ T_{H2} and T_{H17} cells and T follicular helper cells (T_{FH} cells) depending on their specific function and cytokine profile

(Hilligan & Ronchese, 2020). In general, $CD4^+$ T cells interact with other immune cells to induce, refine, and control immune responses. For example, T_{FH} cells react to peptides of foreign antigens (but can also contribute to autoimmunity when they react to self-antigens) and induce antibody production from B cells. T_{H1} cells are specialized for intracellular pathogens like viruses and upon activation they produce the type II IFN IFNγ. Both, T_{FH} and T_{H1} can contribute to autoimmunity when they react to self-antigens. On the other hand, T_{reg} cells regulate the activity of other immune cells like $CD8^+$ T cells and pro-inflammatory $CD4^+$ T cells. They act in an anti-inflammatory fashion to establish tolerance to antigens of non-hostile origin and prevent excessive immune responses and immunopathology. In summary, the subsets of T_H cells confer host defense against different types of pathogens by secreting distinct combinations of cytokines and interacting with other cell types (Hilligan & Ronchese, 2020; Sette & Crotty, 2021).

Activated $CD8^+$ T cells interact directly with infected or malignant host cells. They release cytokines and kill the infected or malignant cells which is why they are also referred to as cytotoxic T lymphocytes (CTLs) or T killer cells (Zinkernagel & Doherty, 1974). Infected cells can present short peptide fragments of intracellular pathogens, called epitopes, on MHC-I molecules on their surface (Zinkernagel & Doherty, 1974; Kaech & Cui, 2012; Toor *et al*, 2021). Recognition of a complex of epitope and MHC-I molecule by a CTL will trigger release of cytotoxic molecules, cytokines and supramolecular attack particles and engagement of the Fas-FasL apoptosis pathway to induce cell death of the infected cell. Perforin and granzymes are cytotoxic molecules released by CTLs; subsequently, perforin forms pores in the target cell membrane through which granzymes enter the cytoplasm of the target. Granzymes then activate the apoptosis pathways of the infected target cell (Voskoboinik *et al*, 2015; Tian *et al*, 2022). CTLs also secrete pro-inflammatory cytokines like IFNγ and TNFα (Slifka & Whitton, 2000; Tian *et al*, 2022). These play a role in the activation of antiviral transcriptional programs in infected cells and uninfected bystander cells, activation of immune cells like macrophages and systemic effects like the activation of acute phase protein expression in the liver and induction of fever (Slifka & Whitton, 2000; Merad *et al*, 2022).

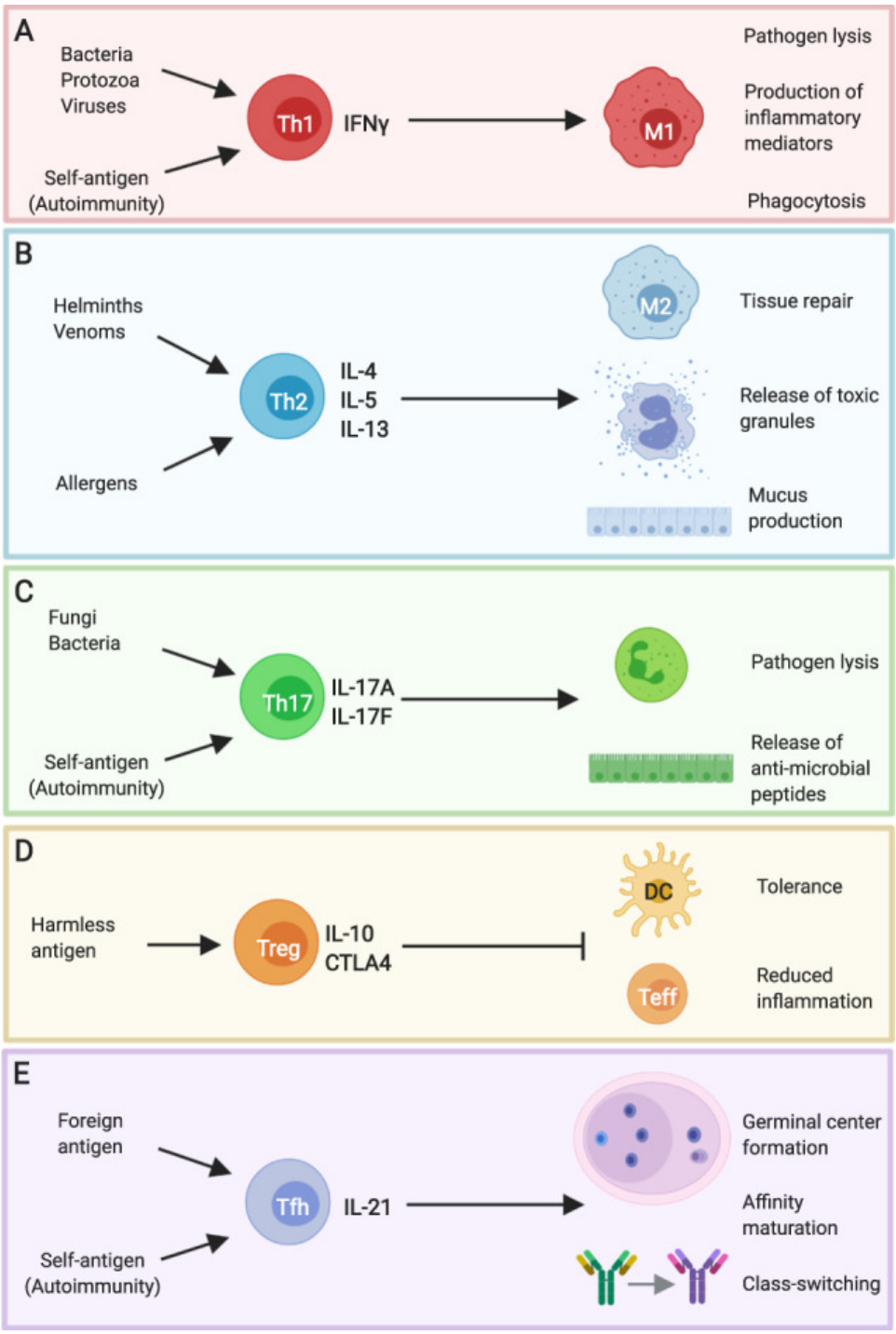


Figure 6: Overview of subsets of CD4⁺ T cells. A) T_{H1} cells play an important role in the defense against intracellular bacterial, protozoa, and viral pathogens. They produce IFNγ and act on macrophages to induce intracellular pathogen lysis, production of inflammatory mediators and phagocytosis. B) T_{H2} cells produce IL-4, IL-5 and IL-13 in response to Helminths, venoms and allergens and act on macrophages to engage in tissue repair, mast cells to release toxic mediators and mucus production by the epithelium. C) Extracellular bacteria and fungi activate T_{H17} cells that produce IL-17A and IL-17F which induce pathogen lysis by neutrophils and macrophages and release of antimicrobial peptide from the epithelium. D) T_{reg} cells downregulate the adaptive immune responses and react to harmless antigens by inducing tolerance in DCs by producing IL-10 and CTLA4. E) T_{FH} cells produce IL-21 upon encounter of foreign antigens and induce the formation of germinal centers, antibody affinity maturation and class switching. Figure obtained from (Hilligan & Ronchese, 2020). Reproduced with permission from Springer Nature. License Number 5518191108665

T cell responses are initiated in the secondary lymphoid organs where activated DCs present antigens to naïve $CD4^+$ and $CD8^+$ T cells (von Andrian & Mempel, 2003). The secondary lymphoid organs spleen and lymph nodes serve as dedicated places of antigen presentation to increase the probability for APCs to interact with naïve T cells with a matching T cell receptor (TCR). DCs are the connection between the innate and adaptive immune response. They surveille their local tissue environments, capture antigens from damaged and infected cells and pathogens from the site of pathogen entry and migrate to lymph nodes where they present these antigens to circulating naïve T cells. DCs are the most efficient APCs for the initiation of T cell responses. However, macrophages and B cells can also act as APCs. In general, T cells can respond to cell-associated short peptide antigens in contrast to B cells which can recognize peptides, intact folded proteins, nucleic acids, carbohydrates, lipids and small chemicals. This is due to the molecular characteristics of the TCR and the mechanism how T cells are trained (Hilligan & Ronchese, 2020; von Andrian & Mempel, 2003). For intracellular pathogens like viruses, proteins of pathogens that infected DCs can be processed in the proteasome to display short peptides in complex with MHC molecules on the cell surface. However, not all pathogens infect DCs. Therefore, DCs can also capture antigens of these pathogens through a process called antigen cross-presentation. DCs can take up molecular structures of damaged and infected cells, process them via the proteasome and display these on the cell surface in a complex with MHC molecules (Hilligan & Ronchese, 2020; Joffre *et al*, 2012)h. MHC-I molecules are ubiquitously expressed to present short peptide antigens of intracellular pathogens, called epitopes, to cytotoxic T cells which will kill the infected cells if they detect a pathogenic antigen with their TCR (Zinkernagel & Doherty, 1974; Kaech & Cui, 2012). On the other hand, MHC-II molecules are expressed specifically by APCs to display extracellular antigens to $CD4^+$ T cells (Kindred & Shreffler, 1972; Hilligan & Ronchese, 2020). Upon antigen capture, DCs will migrate into the next lymph nodes and present the captured antigens to naïve T cells which scan the lymph node for matching antigens (Hilligan & Ronchese, 2020). Naïve T cells will be activated when their TCR interacts with a loaded MHC complex on the surface of a DC. In addition to the antigen-specific activation signal, naïve T cells need co-stimulatory signals like membrane-bound CD80/86 and cytokine signals from DCs for appropriate activation (von Andrian & Mempel, 2003; Kaech & Cui, 2012). Some co-stimulatory receptors are only expressed by DCs upon pathogen encounter which for example triggers TLRs (Hilligan & Ronchese, 2020; Akira *et al*, 2006). This mechanism ensures that T cell responses are only staged against pathogens. Upon successful interaction with a naïve T cell, DCs will switch from antigen capture mode to enhanced antigen presentation due to feedback signals from the activated T cells. Upon activation, T cells undergo clonal expansion and differentiate into different T cell subsets based on specific signals and cytokine environment. Then, they will egress from the secondary lymphoid organs into the circulation

and migrate to the infected tissue guided by chemokine signals (Hilligan & Ronchese, 2020; Kaech & Cui, 2012).

Among T cells, T_{H1}, T_{FH} cells and cytotoxic T lymphocytes are most important in antiviral immune responses for their function in B cell activation and killing of infected host cells. Robust antigen-specific $CD4^+$ and $CD8^+$ T cell responses and antibody responses against SARS-CoV-2 have been described by several clinical studies and are associated with successful resolution of most COVID-19 cases. Early mounting of robust SARS-CoV-2-specific T cell responses was found to be an important factor for successful virus control and resolution for mild COVID-19 cases. $CD8^+$ T cell responses are observed less frequently than $CD4^+$ T cell responses but are correlated with a better disease outcome (Sette & Crotty, 2021). Finally, studies in animal models and clinical studies in humans showed that a combination of antibody and $CD8^+$ T cell responses is necessary for robust protection against SARS-CoV-2 (Brown, 2020; McMahan *et al*, 2021; Earle *et al*, 2021; Feng *et al*, 2021).

As described above, adaptive immunity comprises a variety of humoral and cellular responses to pathogen-specific antigens. However, RNA viruses like coronaviruses are characterized by a certain degree of genetic instability and adaptability (Villa *et al*, 2021). This enables the virus to escape the epitope-specific adaptive immune response. This means that neutralizing antibodies and cytotoxic T cells might not be able to opsonize and neutralize the virus or recognize and eradicate infected cells anymore. This phenomenon is called immune evasion and it was described as one of the major challenges in the pandemic response which has the potential to jeopardize the effectiveness of vaccine programs for herd immunity (Gupta, 2021; Villa *et al*, 2021; DeGrace *et al*, 2022; Kent *et al*, 2022; Escalera *et al*, 2022; Liu *et al*, 2022). Examples for such evasion strategies were shown in a study from Austria where nonsynonymous mutations in MHC-1-restricted $CD8^+$ T cell epitopes led to an altered antigenicity and a loss of antigen recognition by CD8+ T cells (Agerer *et al*, 2021). It highlights the importance of genomic epidemiology and pathogen surveillance for detection and risk assessment of emerging viral variants which might have the potential to evade current vaccines, acquired immunity in hosts, or reduce efficacy of drug regimens.

1.2.7. Phylogeny and origin of SARS-CoV-2

The emergence of SARS-CoV-2 had a tremendous impact on the world economy and social life. This directed a lot of attention to the question where and how the virus entered the population, but determining its origin was also a key matter of coordinating countermeasures to prevent independent reintroductions from animals into the human population. Like the

human coronaviruses HCoV-OC43, HCoV-HKU1, SARS-CoV and MERS-CoV, SARS-CoV-2 belongs to the subfamily of betacoronaviruses (Cui *et al*, 2019; Hu *et al*, 2021; V'kovski *et al*, 2021). Among betacoronaviruses, SARS-CoV, MERS-CoV and SARS-CoV-2 cluster in two adjacent phylogenetic groups, the sarbecoviruses (SARS-like viruses) and merbecoviruses (MERS-like viruses). As described above, all previously described coronaviruses were found to have zoonotic origin (Cui *et al*, 2019; Morens & Fauci, 2020).The outbreak of SARS-CoV-2 shows similarity to SARS-CoV (Xu *et al*, 2004). Both outbreaks of novel coronaviruses were epidemiologically associated with animal markets that sold a variety of live animals like civets and racoon dogs (Holmes *et al*, 2021). The closest phylogenetic relatives for both viruses are among the plethora of alpha- and betacoronaviruses circulating in large bat populations. The viruses that were found to be genetically most closely related to SARS-CoV-2 were documented in populations of bats and pangolins in different countries in South-East Asia (Morens & Fauci, 2020; Holmes *et al*, 2021). Based on genetic distance, the coronavirus RaTG13 from *Rhinolophus affinis* bats was identified as the closest relative with 4% difference in the nucleotide sequence which equals about 1,150 mutations compared to the first sequenced strain Wuhan-Hu-1 (Boni *et al*, 2020; Holmes *et al*, 2021). However, based on genome organization, three other coronaviruses from bats, RmYN02, RpYN06 and PrC31, are closer related to SARS-CoV-2 and, thus, rather expected to share a common ancestor with SARS-CoV-2 than RaTG13 (Zhou *et al*, 2021; Lytras *et al*, 2022). Until today, the zoonotic origin of SARS-CoV-2 or its intermediate hosts for animal-to-human transmission are still matter of discussion.

Epidemiological tracing of the first cases in Wuhan during December 2019 revealed that more than half of the infected individuals had exposure at the Huanan food market or other markets situated in Wuhan. Finally, the Huanan food market was identified as the main location around which the majority of the first cases clustered (Holmes *et al*, 2021). Early during the pandemic theories arose about a possible escape of SARS-CoV-2 from the Wuhan Institute for Virology (WIV), due to missing clear evidence for the zoonotic origin of the virus. In the past, comparable scenarios have been described for example for laboratory-related outbreaks of SARS-CoV (Senio, 2003; Lim *et al*, 2004; Parry, 2004). Nevertheless, despite extensive epidemiological tracing there were no laboratory-related cases reported for staff at the WIV and laboratory staff was tested seronegative in a WHO investigation in 2020 (World Health Organization, 2020a; Holmes *et al*, 2021). Since different animal populations harbor a diversity of coronaviruses, more research is necessary to document and monitor these coronavirus populations, especially in bats. This could help to broaden the body of evidence leading to the origin that facilitated the transmission of SARS-CoV-2 from animals into the human population.

1.3. Virology of SARS-CoV-2

In general, viruses can be divided into two main classes based on the biochemical properties of their genome: DNA and RNA viruses. Among RNA viruses, there are double-stranded RNA viruses and single-stranded RNA viruses either with negative-sense or positive-sense RNA genome. RNA viruses require an RNA-dependent RNA polymerase (RdRp) to replicate their genome. In general, these enzymes express only low fidelity proof-reading mechanisms which leads to a certain degree of genetic instability in RNA viruses (Alcami & Koszinowski, 2000). This feature drives the rise of genetic variants that can display different characteristics (Alcami & Koszinowski, 2000; Villa *et al*, 2021).

1.3.1 General virology of SARS-CoV-2

The virion of SARS-CoV-2 consists of the ~30 kilobases long single-stranded positive-sense RNA genome and four structural proteins: envelope (E) protein, membrane (M) protein, nucleocapsid (N) protein and spike (S) protein (see Figure 7a) (Kim *et al*, 2020; V'kovski *et al*, 2021).

The viral genome is flanked by 5' and 3' untranslated regions that establish secondary RNA structures and play a role in the synthesis of the RNA genome as *cis*-acting sites. About 21 kb at the 5' end of the capped polyadenylated viral genome code for the two large open reading frames (ORFs) ORF1a and ORF1b (Kim *et al*, 2020; V'kovski *et al*, 2021). During the viral life cycle, these two ORFs are translated into two large polyproteins, pp1a and pp1ab which will be co-translationally and post-translationally cleaved into 16 non-structural proteins (nsps) essential for different aspects of the viral life cycle (V'kovski *et al*, 2021; Lou & Rao, 2022). Coronaviruses are part of the order of *Nidovirales.* This class of RNA viruses is characterized by a discontinued transcription process of the viral RNA genome that produces nested mRNA molecules with identical 3' ends during the viral life cycle (see Figure 7b) (Sawicki *et al*, 2007; Sola *et al*, 2015). This process affects several genes that encode the abovementioned structural proteins and at least six accessory proteins (3a, 6, 7a, 7b, 8 and 10) at the 3' end of the viral genome (see Figure 7b).The exact number of accessory proteins is still discussed and needs further experimental evidence (Kim *et al*, 2020; V'kovski *et al*, 2021). However, it is known from other coronaviruses that these proteins are not essential for effective viral replication in cell culture, show higher variability than the structural proteins and are thought to be involved in modulating host immune responses and viral pathogenicity (Yount *et al*, 2005; Perlman & Netland, 2009).

1.3.2. Viral Lifecycle

Coronavirus virions present homotrimeric class I fusion glycoproteins, termed spike proteins, on the surface that are functionally divided into the surface-exposed S1 domain and the transmembrane S2 domain. The S1 domain contains the receptor-binding domain (RBD) that mediates host cell receptor binding and determines the viral cell tropism. The S2 domain contains two heptad repeat regions and a fusion peptide which facilitates fusion of viral and host cell membranes for uncoating and release of the viral genome into the cytoplasm (Tortorici & Veesler, 2019; V'kovski *et al*, 2021). The infection cycle of host cells starts with attachment of the viral S protein to ACE2 expressed on the surface of host cells (see Figure 7b) (Letko *et al*, 2020; Hoffmann *et al*, 2020; Zhou *et al*, 2021). The S proteins of SARS-CoV-2 and SARS-CoV share a high degree of homology in sequence and structure, and both were found to use ACE2 as entry receptor (Li *et al*, 2003; Hoffmann *et al*, 2020; Letko *et al*, 2020). The expression profile of ACE2 can therefore provide some evidence to explain some of the clinical manifestations of SARS-CoV-2 and SARS-CoV. It was found that ACE2 mRNA is almost ubiquitously expressed in human tissues, but only certain cell types express the protein above detection limit. Among other organs, ACE2 is most abundantly expressed in the lung epithelium and the epithelium of the small intestine, corresponding well to the respiratory and gastrointestinal symptoms of SARS-CoV-2 (Hamming *et al*, 2004). Moreover, immunohistochemistry and *in situ* hybridization studies showed that SARS-CoV, could not just be detected in the lung and small intestine but also in the stomach, pancreas, liver, cerebrum, and other organs (Ding *et al*, 2004). Upon binding to the host cell entry receptor ACE2, SARS-CoV-2 and SARS-CoV require cleavage of the S protein by the transmembrane protease serine 2 (TMPRSS2) provided by the host cell to induce virus-plasma membrane fusion (Matsuyama *et al*, 2010; Hoffmann *et al*, 2020). However, it was shown that the endosomal cysteine proteases cathepsin B and L can also facilitate this cleavage and assist in viral entry (Simmons *et al*, 2005). The existence of a polybasic cleavage site (amino acid motif PRRAR) between S1 and S2 domain is another noteworthy feature of the structural composition of the SARS-CoV-2 S protein which distinguishes it from other members of the *Sarbecovirus* genus (V'kovski *et al*, 2021). This motif represents a binding site for cleavage by the enzyme furin which was shown to be required for successful infection of host cells but also contributes to expanded cell tropism and the potential to cross species barriers (Tortorici & Veesler, 2019; Walls *et al*, 2020; Hoffmann *et al*, 2020; Peacock *et al*, 2021).

Upon cellular entry and uncoating, the positive-strand RNA genome of SARS-CoV-2 is released into the cytoplasm of the host cell which leads to the initiation of a complex spatio-temporally controlled viral gene expression pattern (see Figure 7b). The gene expression

program starts with the translation of ORF1a and ORF1b which gives rise to the two polyproteins pp1a and pp1ab. Subsequently, these polyproteins will be co- and post-translationally cleaved into sixteen non-structural proteins by the two viral cysteine proteases nsp3 (papin-like protease; PLpro) and nsp5 (chymotrypsin-line protease or main protease; M^{pro}) (V'kovski *et al*, 2021; Lou & Rao, 2022). This process yields the proteins nsp1 to nsp11 from pp1a, as well as nsp1 to nsp10 and nsp12 to nsp16 from pp1ab (V'kovski *et al*, 2021). The virulence factor nsp1 will be rapidly released from the polypeptide and interfere with the host transcriptional machinery in order to inhibit the expression of host factors that mediate cellular antiviral defense mechanisms (Thoms *et al*, 2020). The remaining non-structural proteins nsp2 to nsp16 form the viral replication and transcription complex (RTC) at defined subcellular locations (Snijder *et al*, 2016). Nsp12 to nsp16 form the core of the RTC and nsp2 to nsp11 provide supporting functions like interaction with host proteins, modulation of host membranes and host immune evasion to facilitate replication and transcription of the viral genome. Nsp12 conducts replication of the viral genome as the viral RdRP together with nsp7 and nsp8 that contribute primase and 3'-terminal adenylyltransferase activity (Perlman & Netland, 2009; Snijder *et al*, 2016; V'kovski *et al*, 2021). With their large genome size, coronaviruses require proofreading during RNA replication which is provided by the exonuclease nsp14 (Eckerle *et al*, 2007; Snijder *et al*, 2016; V'kovski *et al*, 2021; Lou & Rao, 2022). Nsp10, nsp13, nsp14 and nsp16 participate in the coronavirus capping machinery as methyltransferases and triphosphatases (Snijder *et al*, 2016). The formation of endoplasmic reticulum (ER)-derived, interconnected double-membrane vesicles (DMVs) postulated to serve as replication organelles is mediated by nsp3, nsp4 and nsp6 by manipulating host intracellular membranes and takes place early during the coronavirus replication cycle (see Figure 7b) (Knoops *et al*, 2008; Snijder *et al*, 2016).

For the replication process, an intermediary full-length negative-strand RNA serves as template to produce the positive-strand RNA genome which will be used for assembly of new virions or translation of more nsps (V'kovski *et al*, 2021). As mentioned before, the transcription process of viruses of the order of *Nidovirales* like SARS-CoV-2 is characterized by the production of nested subgenomic RNAs mediated by transcription regulatory sequences (TRSs) upstream of ORFs (Sawicki & Sawicki, 1995; Sawicki *et al*, 2007). This process plays a role for all genes for non-structural and structural proteins in the 3' one-third of the genome downstream of ORF1a and ORF1b and takes place during the synthesis of the negative-strand RNA genome template. The RTC stalls transcription at TRSs and will subsequently re-initiate synthesis of the negative-strand RNA at a TRS adjacent to a leader sequence close to the 5' end of the genome (V'kovski *et al*, 2021). This nested transcription

creates subgenomic RNAs with shared 3' and 5' ends which are used subsequently to produce nested positive-strand RNAs that code for structural and accessory proteins (Kim *et al*, 2020).

New virions will then be assembled with the structural proteins S, M, E and N and the viral positive-strand RNA genome. These proteins translocate into the ER, travel further to the ER-to-Golgi intermediate compartment (ERGIC) where they interact with N-encapsidated viral genomes and finally bud into the lumen of secretory vesicles and leave the infected cell via the lysosomal trafficking pathway (Ghosh *et al*, 2020).

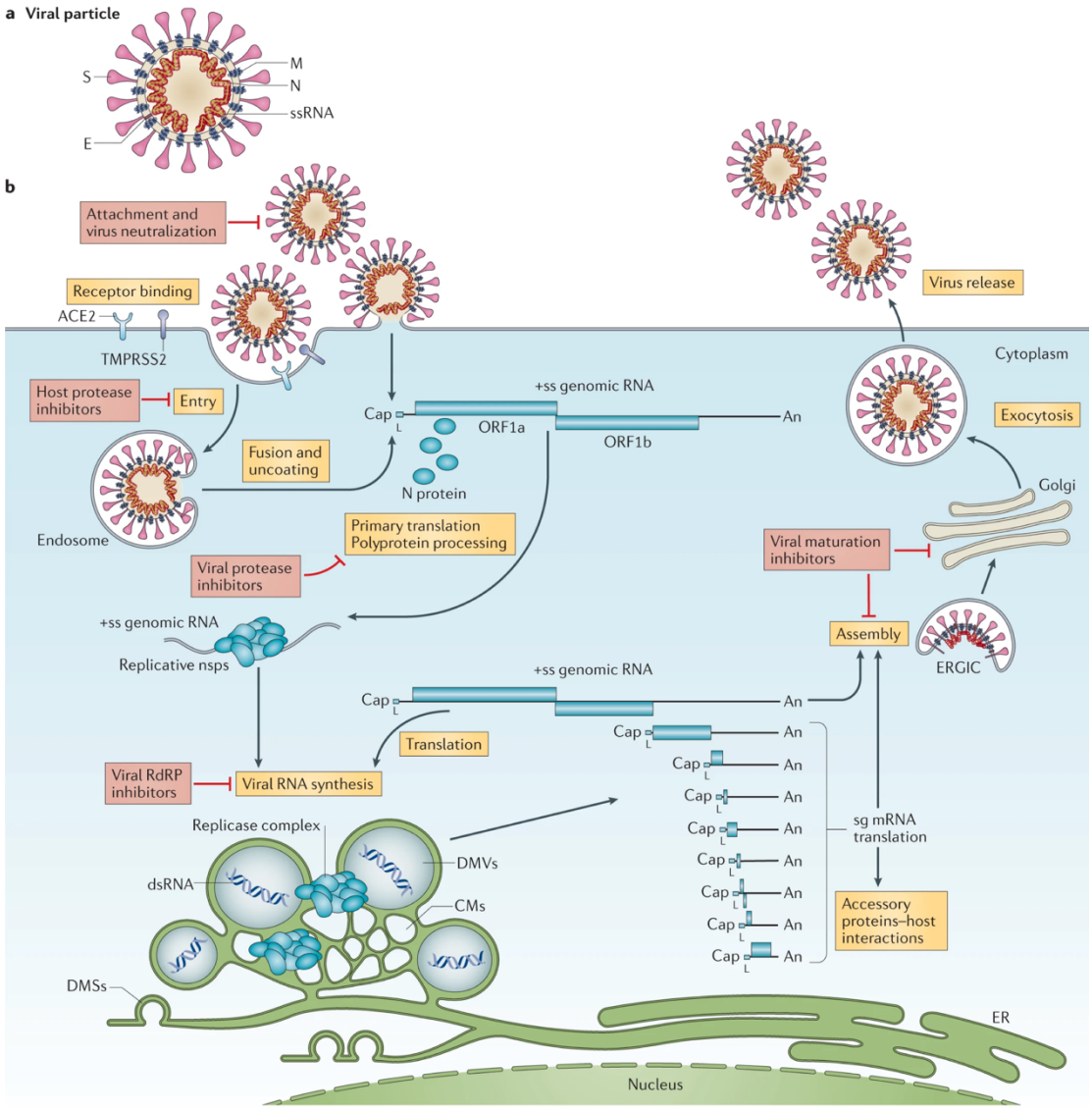


Figure 7: The life cycle of coronaviruses. A) Virions of coronaviruses carry S, E, M and N protein and the ssRNA viral genome. B) The coronavirus life cycle starts by cell entry with the ACE2 and TMPRSS2 receptors and release of the viral genome into the cytoplasm. Then, the translation of ORF1ab is initiated, the resulting polyproteins are cleaved to several non-structural proteins and form the RTC to start replication of the viral genome. Replicated viral genomes are finally assembled to new virions and released from the cell. This figure was taken from (V'kovski *et al*, 2021). Reproduced with permission from Springer Nature. License Number: 5518200077052

1.4. Mutational behavior and host adaptation of viruses

The first chapters of this introduction discussed the necessary and supporting factors for the emergence of a new infectious disease and its establishment in a population. These factors act on different levels: population, individual and pathogen. Pathogens must overcome several obstacles on their path to establishment in the human population, pathogens must overcome several obstacles – the first can be to be adapted enough to two species at the same time in order to cross the species barrier (Parrish *et al*, 2008; Morens & Fauci, 2020; Baker *et al*, 2022).

RNA viruses depend on RdRps for the replication of their genomes (Cui *et al*, 2019; V'kovski *et al*, 2021). In general, these enzymes harbor no or only very low proof-reading capacity which is why a high degree of genetic instability is an inherent characteristic of RNA viruses (Alcami & Koszinowski, 2000). This feature gives RNA viruses the highest mutation rates in nature and confers advanced ability for adaptation (Combe & Sanjuán, 2014; Dolan *et al*, 2018; Villa *et al*, 2021). With different subtypes of Ebola, Influenza A viruses as well as the coronaviruses SARS-CoV, MERS-CoV and SARS-CoV-2, most of the viruses with pandemic potential that emerged during the last century were RNA viruses (Woolhouse *et al*, 2013). This suggests enhanced adaptability through genetic instability is a mechanism that makes RNA viruses generate variants with pandemic potential more often than for example DNA viruses (Alcami & Koszinowski, 2000; Villa *et al*, 2021).

However, this trade-off between stability and adaptability comes at the cost of limited genome size (Alcami & Koszinowski, 2000; Lloyd *et al*, 2014). Excessive genetic instability could result in a decline of viral fitness, particularly for coronaviruses, which carry relatively large viral genomes (~30 kb) compared to other RNA viruses such as HIV and HCV, which have genomes smaller than 10 kb (German Advisory Committee Blood (Arbeitskreis Blut), 2016; Shi & Suzuki, 2018; V'kovski *et al*, 2021). However, coronaviruses encode enzymes with proofreading capacity, such as the viral RdRp nsp12 and the exonuclease nsp14 in contrast to other RNA viruses, such as HIV (Lloyd *et al*, 2014; V'kovski *et al*, 2021). In general, RNA virus genomes leave very limited capacity to evolve genes dedicated only to defending against the host immune response compared to DNA viruses like herpes- and poxviruses whose genomes largely code for genes that facilitate host control. This is the reason why many RNA virus genes are multifunctional and often serve in the viral replication cycle and in suppression of the host immune response as discussed in earlier chapters. This chapter will introduce the general basic concepts of RNA virus evolution and discuss the process of species barrier crossing (Alcami & Koszinowski, 2000).

1.4.1. Mechanisms of RNA virus evolution

RNA viruses show mutation frequencies of 10^{-4} to 10^{-6} substitutions per nucleotide per replication round. Transitions (substitution within same chemical class of nucleotide, e.g. purine base to purine base) are more likely than transversions (substitution across chemical classes of nucleotides, e.g. purine base to pyrimidine base). The mutation frequency varies among different RNA virus families and is also dependent on the host cells, the polarity of the RNA genome and whether the RNA genome is single- or double-stranded (Alcami & Koszinowski, 2000; Combe & Sanjuán, 2014). During or after the replication process, host-derived editing enzymes like adenosine deaminase, RNA specific (ADAR) can edit nucleotide bases on regions of double-stranded RNA by deamination of adenosine to inosine which causes A-to-G mutations (Dolan *et al*, 2018).

In general, evolutionary changes in the population are driven by mutation, genetic drift, and selection. Mutations caused by the error-prone replication machinery of RNA viruses or gene editing enzymes generate phenotypic diversity in the viral population (Holland *et al*, 1982; Eigen, 1993; Domingo & Holland, 1997; Dolan *et al*, 2018). Each mutation can have a mutational fitness effect, meaning it affects the viral replicative capacity, transmissibility, or other parameters of viral fitness. Moreover, the RdRp can mediate recombination between RNA genomes and antigenomes and, thus, create linkage between different mutations in the same population or link mutations from different lineages in co-infected cells. RNA viruses with segmented genomes can exchange these segments between different parental strains during a co-infection in a process called genome reassortment (Dolan *et al*, 2018). Genome reassortment is one of the major drivers of the evolution of influenza viruses (Petrova & Russell, 2018). Besides single mutations, recombination events and genome reassortment, as well as genetic drift influence the establishment of variants with new phenotypic characteristics. Genetic drift describes the stochastic fluctuation of allele frequencies. It can be induced for example when a virus transmits through or between hosts and encounters several bottlenecks that each can stochastically shape the genotypic composition of the virus population (Dolan *et al*, 2018). The estimation of transmission bottleneck sizes for SARS-CoV-2 was part of this thesis and will be discussed later. Briefly, the transmission bottleneck describes the number of viral particles that are on average transmitted between two hosts during a successful infection event (Domingo *et al*, 2012). In general, bottleneck events can adjust the equilibrium of allele frequencies in the virus population. A small transmission bottleneck can lead to large genetic drifts while large transmission bottlenecks (e.g. “en bloc” transmission) can preserve genotypic diversity in the virus population (Domingo *et al*, 2012; Dolan *et al*, 2018).

Upon infection of a new host, the virus will expand from a few transmitted particles to a larger virus population size. Due to its genetic instability, the virus will establish a genotypically diverse population of variants which can be regarded as a “swarm of mutant genotypes” that surrounds one or several modal master sequences (Dolan *et al*, 2018). This mutational behavior of RNA viruses is the foundation for their extraordinary ability for adaptation and was described by Eigen and Schuster as “viral quasispecies evolution” (Eigen & Schuster, 1977; Eigen, 1993; Domingo *et al*, 2012; Andino & Domingo, 2015). RNA virus evolution will act on the level of quasispecies represent (Eigen & Schuster, 1977; Eigen, 1993; Domingo *et al*, 2012; Andino & Domingo, 2015). Negative or purifying selection will remove alleles with deleterious effect on the viral fitness and positive selection will drive the fixation of advantageous mutations in the virus population (Dolan *et al*, 2018). Studies showed that the genetic instability of RNA viruses can drive rapid host adaptation or evasion of selection pressure and that diminishing genetic instability attenuates some RNA viruses *in vivo* (Vignuzzi *et al*, 2006; Kautz & Forrester, 2018). This led to hypotheses proposing that different genotypes within a virus quasispecies population could interact in a cooperative fashion to invade the host or that dynamic genotype compositions at different stages of the infection are crucial for a successful infection of the host (Vignuzzi *et al*, 2006; Domingo *et al*, 2012; Andino & Domingo, 2015). This means that the swarm of mutant genotypes contains variants that are dedicated to performing specific tasks to drive infection and host invasion. For example, fidelity mutants that mutate slower than the average of the virus population serve as molecular memory to conserve advantageous traits in the virus population. Or fidelity mutants that mutate faster than the average population can increase its overall adaptability to hosts and tissues (Domingo *et al*, 2012; Kautz & Forrester, 2018). These concepts have been observed for a variety of RNA viruses that are known for their ability to rapidly escape from selection pressures like antiviral treatments and immune responses such as Hepatitis C virus (HCV) and Human Immunodeficiency Virus-1 (HIV-1) (Bowen & Walker, 2005; Deng *et al*, 2015). It was also observed for respiratory RNA viruses like influenza (Pircher *et al*, 1990).

1.4.2. Crossing the species barrier

The first step in the emergence of a novel infectious disease is crossing the species barrier to infect a new host, also known as host-switching. As described in the first chapters of this introduction, many infectious diseases that are endemic today are expected to have undergone this process at some point during the last 3,000 years since human settlements reached sufficient population densities to surpass the threshold of establishment. However, the molecular mechanisms and decisive parameters driving this process are still poorly

understood, also due to difficulties reproducing natural-world findings of within-host viral evolution in the lab (Morens & Fauci, 2020).

One of the principles of host-switching is that the pathogen must overcome the “fitness valley” – meaning it has to find sufficient adaptation levels for both species A and species B to facilitate a successful transition (Parrish *et al*, 2008; Araujo *et al*, 2015; Morens & Fauci, 2020; Baker *et al*, 2022). Therefore, the pathogen must be able to replicate in, but should not be too adapted to, species A so that it can still adapt enough to species B to establish a successful infection (Morens & Fauci, 2020; Baker *et al*, 2022). It remains elusive whether the viral quasispecies adaptation mechanisms of RNA viruses increase the ability for host-switching compared to other pathogens (Parrish *et al*, 2008). It is also possible that it could be a concept of Darwinian evolution where the emergence of single novel virus variants facilitates host-switching and starts new quasispecies populations in the next species (Morens & Fauci, 2020). Recent research has shown that host-switching might not depend on the unlikely sudden emergence of a specific variant that manages host-switching. It was postulated that the population densities of two species, the closeness of their interaction and, thus, the number of times the pathogen encounters both species have a positive influence on the probability for host-switching (Morens & Fauci, 2020; Baker *et al*, 2022).

Upon successful infection of a new host, the newly emerging infectious disease has to adapt sufficiently in order to facilitate transmission among individuals of the new species. In terms of a zoonotic disease this means that the pathogen must adapt to be able to facilitate not just animal-to-human but also human-to-human transmission. This requires sufficient mutational adaptation to the molecular characteristics of the new host species (Morens & Fauci, 2020; Baker *et al*, 2022). This thesis investigates the emergence of SARS-CoV-2 in the first stage of the COVID-19 pandemic in 2020 with specific focus on introduction and transmission events in Austria.

1.5. Aims of this thesis

The aims of this thesis were the following:

1) Collect and sequence RNA samples of SARS-CoV-2 cases from Austrian infection clusters and superspreading events during the early phase of the pandemic in 2020.
2) Analyze the dynamics of low-frequency mutations and fixed mutations in these infection clusters.
3) Combine genomic information with epidemiological information to conduct genomic epidemiology on these infection clusters.
4) Investigate the effect of transmission bottlenecks on interhost mutational dynamics of the SARS-CoV-2.
5) Determine the intrahost diversity of low-frequency mutations from longitudinal samplings in SARS-CoV-2 cases.

2. Results

2.1. Prologue

SARS-CoV-2 emerged as a new infectious disease in December 2019 in China. Within the following months, SARS-CoV-2 rapidly established several infection clusters across the globe leading to a pandemic outbreak. The first introduction events to Europe were identified in late January 2020. In this study titled "**Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2**", we combined genome information from SARS-CoV-2 cases with their epidemiological information to perform genomic epidemiology. We sequenced 572 SARS-CoV-2 RNA samples from 449 cases of SARS-CoV-2 infection sampled between February 24th and May 7th, 2020. These samples were sequenced at a sequencing depth sufficient to conduct mutational analysis of fixed and low-frequency mutations. These data were used to investigate the dynamics of low-frequency mutations and establishment of new fixed mutations in superspreading events in infection clusters in Austria. Moreover, we analyzed the intra- and inter-host mutational dynamics of the virus in longitudinal measurements of single cases of COVID-19.

2.2. Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2

Science Translational Medicine; 09 Dec 2020; Vol. 12, Issue 573, eabe2555
DOI: 10.1126/scitranslmed.abe2555

Alexandra Popa*, Jakob-Wendelin Genger*, Michael D. Nicholson†, Thomas Penz†, Daniela Schmid†, Stephan W. Aberle†, Benedikt Agerer†, Alexander Lercher†, Lukas Endler , Henrique Colaço , Mark Smyth , Michael Schuster , Miguel L. Grau , Francisco Martínez-Jiménez, Oriol Pich, Wegene Borena, Erich Pawelka, Zsofia Keszei, Martin Senekowitsch, Jan Laine, Judith H. Aberle, Monika Redlberger-Fritz, Mario Karolyi, Alexander Zoufaly, Sabine Maritschnik, Martin Borkovec, Peter Hufnagl, Manfred Nairz, Günter Weiss, Michael T. Wolfinger, Dorothee von Laer, Giulio Superti-Furga, Nuria Lopez-Bigas, Elisabeth Puchhammer-Stöckl, Franz Allerberger, Franziska Michor, Christoph Bock, Andreas Bergthaler
*,† equal contributions

The publication was published as an open-access article and is distributed under a Creative Commons Attribution License 4.0 (CC BY) by the Journal Science Translational Medicine. Reprinted with permission from AAAS (American Association for the Advancement of Science). License number: OP-00145573

2.2.1. Issue Cover

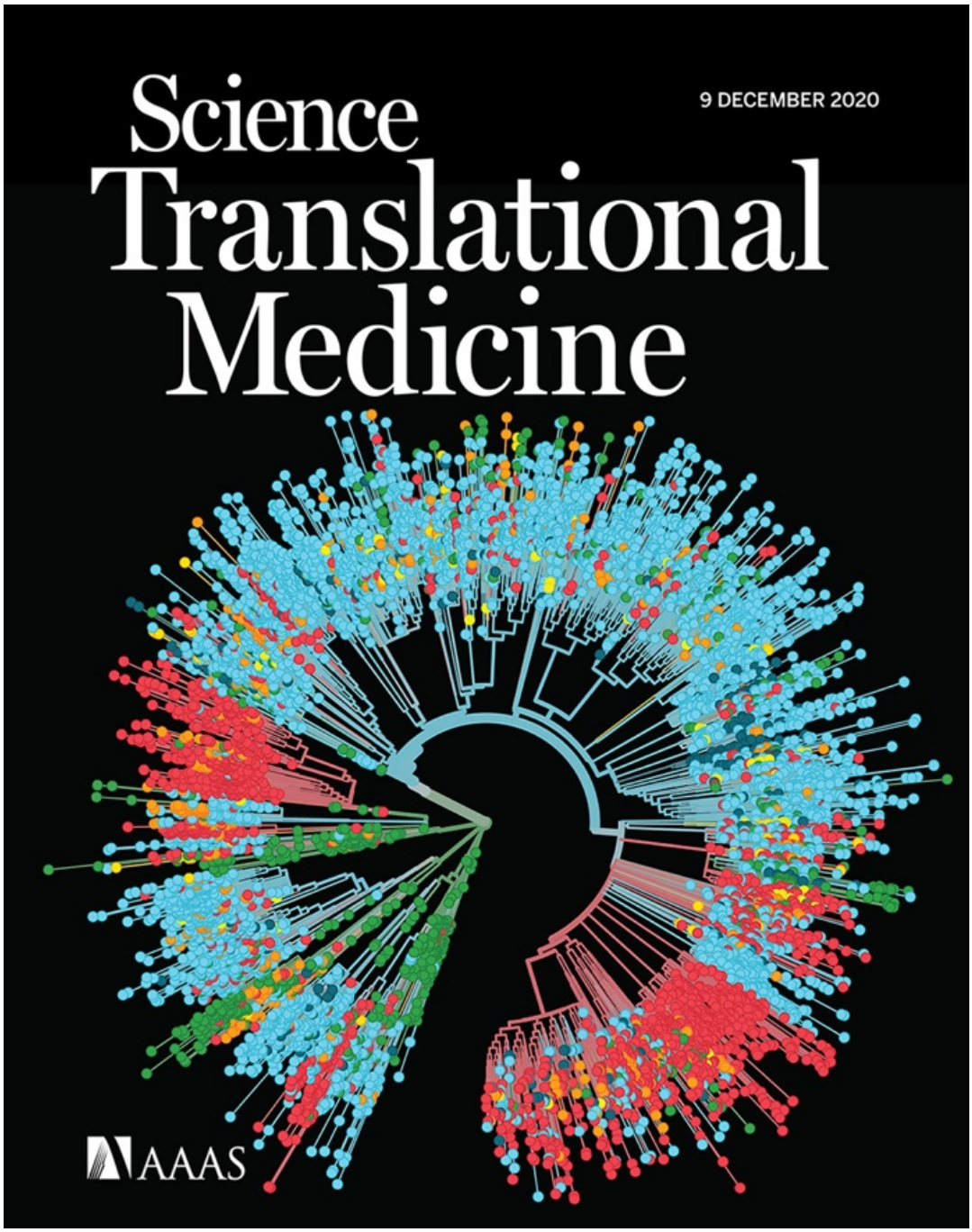


Figure 8: Online cover of Science Translational Medicine Issue 573 (December 9th 2020; Vol. 12) designed by Jakob-Wendelin Genger. The image depicts a circular phylogenetic tree of SARS-CoV-2 genomes used in the publication. Colors indicate continental origin of the samples: Africa (yellow), Asia (green), Austria (dark blue), Europe (light blue), North America (red), South America (orange). Reprinted with permission from AAAS (American Association for the Advancement of Science). License number: OP-00145573

SCIENCE TRANSLATIONAL MEDICINE | RESEARCH ARTICLE

CORONAVIRUS

Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2

Alexandra Popa[1]*, Jakob-Wendelin Genger[1]*, Michael D. Nicholson[2,3,4†], Thomas Penz[1†], Daniela Schmid[5†], Stephan W. Aberle[6†], Benedikt Agerer[1†], Alexander Lercher[1†], Lukas Endler[7], Henrique Colaço[1], Mark Smyth[1], Michael Schuster[1], Miguel L. Grau[8], Francisco Martínez-Jiménez[8], Oriol Pich[8], Wegene Borena[9], Erich Pawelka[10], Zsofia Keszei[1], Martin Senekowitsch[1], Jan Laine[1], Judith H. Aberle[6], Monika Redlberger-Fritz[6], Mario Karolyi[10], Alexander Zoufaly[10], Sabine Maritschnik[5], Martin Borkovec[5], Peter Hufnagl[5], Manfred Nairz[11], Günter Weiss[11], Michael T. Wolfinger[12,13], Dorothee von Laer[9], Giulio Superti-Furga[1,14], Nuria Lopez-Bigas[8,15], Elisabeth Puchhammer-Stöckl[6], Franz Allerberger[5], Franziska Michor[2,3,4,16,17,18], Christoph Bock[1,19], Andreas Bergthaler[1‡]

Copyright © 2020 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY).

Superspreading events shaped the coronavirus disease 2019 (COVID-19) pandemic, and their rapid identification and containment are essential for disease control. Here, we provide a national-scale analysis of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) superspreading during the first wave of infections in Austria, a country that played a major role in initial virus transmissions in Europe. Capitalizing on Austria's well-developed epidemiological surveillance system, we identified major SARS-CoV-2 clusters during the first wave of infections and performed deep whole-genome sequencing of more than 500 virus samples. Phylogenetic-epidemiological analysis enabled the reconstruction of superspreading events and charts a map of tourism-related viral spread originating from Austria in spring 2020. Moreover, we exploited epidemiologically well-defined clusters to quantify SARS-CoV-2 mutational dynamics, including the observation of low-frequency mutations that progressed to fixation within the infection chain. Time-resolved virus sequencing unveiled viral mutation dynamics within individuals with COVID-19, and epidemiologically validated infector-infectee pairs enabled us to determine an average transmission bottleneck size of 10^3 SARS-CoV-2 particles. In conclusion, this study illustrates the power of combining epidemiological analysis with deep viral genome sequencing to unravel the spread of SARS-CoV-2 and to gain fundamental insights into mutational dynamics and transmission properties.

INTRODUCTION

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic has already infected more than 20 million people in 188 countries, causing 737,285 deaths globally as of 11 August 2020 and extraordinary disruptions to daily life and national economies (*1*, *2*). The international research community rapidly defined pathophysiological characteristics of the coronavirus disease 2019 (COVID-19), established diagnostic tools, assessed immunological responses, and identified risk factors for a severe disease course (*3*–*6*). Clustered outbreaks and superspreading events of the SARS-CoV-2 pose a particular challenge to pandemic control (*7*–*10*). However, we still know comparatively little about fundamental properties of SARS-CoV-2 genome evolution and transmission dynamics within the human population.

Acquired fixed mutations enable phylogenetic analyses and have already led to insights into the origins and routes of SARS-CoV-2 spread (*11*–*14*). Conversely, low-frequency mutations and their changes over time within individual patients can provide insights into the dynamics of intrahost evolution. The resulting intrahost viral populations represent groups of variants with different frequencies, whose genetic diversity contributes to fundamental properties of infection and pathogenesis (*15*, *16*).

Austria is located in the center of Europe and has a population of 8.8 million. It operates a highly developed health care system, which includes a national epidemiological surveillance program. As of 7 August 2020, contact tracing had been performed for all 21,821 reported SARS-CoV-2–positive cases. Out of these, 10,385 cases were linked to epidemiological clusters, whereas no infection chains were identified for the remaining cases (*17*). Linked to Austria's prominent role in international winter tourism, the country emerged as a

[1]CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, 1090 Vienna, Austria. [2]Department of Data Sciences, Dana-Farber Cancer Institute, Boston, MA, USA. [3]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [4]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA. [5]Austrian Agency for Health and Food Safety (AGES), 1220 Vienna, Austria. [6]Center for Virology, Medical University of Vienna, 1090 Vienna, Austria. [7]Bioinformatics and Biostatistics Platform, Department of Biomedical Sciences, University of Veterinary Medicine, 1210 Vienna, Austria. [8]Institute for Research in Biomedicine (IRB), 08028 Barcelona, Spain. [9]Institute of Virology, Medical University Innsbruck, 6020 Innsbruck, Austria. [10]Department of Medicine IV, Kaiser Franz Josef Hospital, 1100 Vienna, Austria. [11]Department of Internal Medicine II, Medical University of Innsbruck, 6020 Innsbruck, Austria. [12]Department of Theoretical Chemistry, University of Vienna, 1090 Vienna, Austria. [13]Research Group Bioinformatics and Computational Biology, Faculty of Computer Science, University of Vienna, 1090 Vienna, Austria. [14]Center for Physiology and Pharmacology, Medical University of Vienna, 1090 Vienna, Austria. [15]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain. [16]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [17]Ludwig Center at Harvard, Boston, MA, USA. [18]Center for Cancer Evolution, Dana-Farber Cancer Institute, Boston, MA, USA. [19]Department of Laboratory Medicine, Medical University of Vienna, 1090 Vienna, Austria.
*These authors contributed equally to this work.
†These authors contributed equally to this work.
‡Corresponding author. Email: abergthaler@cemm.oeaw.ac.at

Downloaded from https://www.science.org on August 02, 2022

Reprinted with permission from AAAS. License number: OP-00145573

potential superspreading transmission hub across the European continent in early 2020. During the first phase of the pandemic in Europe (February to May 2020), winter tourism–associated spread of SARS-CoV-2 from Austria may have been responsible for up to half of the imported cases in Denmark and Norway and a considerable share of imported cases in several other countries including Iceland and Germany (*11*, *18*, *19*).

In this study, we reconstructed major SARS-CoV-2 infection clusters in Austria and analyzed their role in international virus spread by combining phylogenetic and epidemiological analyses. Moreover, we analyzed our deep viral genome sequencing data from epidemiologically identified transmission chains and family clusters using biomathematical models, to infer genetic bottlenecks and the mutation dynamics of SARS-CoV-2 genome evolution. Our results provide fully integrated genetic and epidemiological evidence for continental spread of SARS-CoV-2 from Austria and establish fundamental transmission properties in the human population.

RESULTS

Genomic epidemiology reconstruction of SARS-CoV-2 infection clusters in Austria

We selected and analyzed SARS-CoV-2 virus samples from geographical locations across Austria, with a focus on the provinces of Tyrol and Vienna, given that these two regions were initial drivers of the pandemic in Austria (fig. S1A) (*17*). We sequenced 572 SARS-CoV-2 RNA samples from 449 unique SARS-CoV-2 cases spanning a time frame between 24 February and 7 May. This captured both the onset and the peak of the initial COVID-19 outbreak in Austria (Fig. 1A). The selected samples covered multiple epidemiological and clinical parameters including age, sex, and viral load (fig. S1, B and C). Samples from both swabs (nasal and oropharyngeal) and secretions (tracheal and bronchial) were included (fig. S1D) to investigate the evolutionary dynamics not only within the population but also within individuals.

Of the 572 samples, 427 passed our sequencing quality controls (>96% genome coverage, >80% aligned viral reads, and ≤1500 uncalled nucleotides in the consensus sequences), and after the removal of cell culture samples, 420 samples were considered for low-frequency analysis. Of the 420 samples, 345 corresponded to unique SARS-CoV-2 cases and were further integrated in our phylogenetic analyses, as they corresponded to unique patient identifiers with complete sample annotation at the time of the analysis (fig. S1E). For these 345 samples, we assembled SARS-CoV-2 genome sequences, constructed phylogenies, and identified low-frequency mutations based on high-quality sequencing results with >5 million reads per sample and >80% of mapped viral reads (fig. S2, A and B).

To obtain robust quantifications of minor variants in all 420 samples, we validated our sample processing workflow and pipeline with additional experimental controls including synthetic SARS-CoV-2 genome titrations, technical replicates for sample preparation and sequencing runs, and dilution experiments (data file S1). Matched controls were highly consistent with each other, indicative of excellent assay performance and a highly reproducible analysis pipeline (fig. S2, C to F). For an alternative allele frequency of 0.01, we obtained an average accuracy of 90.92% (ranging from 68 to 97%). In addition, the shared percentage of detected variants between control pairs ranged from 50 to 90.97% for a cutoff of 0.02 of the allele frequencies. The high specificity of detection even at low frequencies, as well as the large overlap of detected variants, supported the choice of a 0.02 frequency cutoff for calling high-confidence variants (data file S1).

To investigate the link between local outbreaks in Austria and the global pandemic, we performed phylogenetic analysis of 345 SARS-CoV-2 genomes from Austrian cases and 7666 global genomes from the GISAID (Global Initiative on Sharing All Influenza Data) database (data file S2). Similar mutation profiles, together with information of geographical proximity of the samples and time of infection, are strong indicators of possible transmission links. Therefore, groups of virus sequences were annotated as phylogenetic clusters when they all shared a homogeneous mutation pattern and originated from the same geographical location and time period. Among the distinct phylogenetic clusters identified, six could be linked to specific geographic locations of the probable region of infection (Fig. 1B). Three of these six clusters comprised samples with a geographical location mainly in the Tyrol region (hereafter named Tyrol-1, Tyrol-2, and Tyrol-3), whereas the other three originated in Vienna (hereafter named Vienna-1, Vienna-2, and Vienna-3). These clusters are related to the global clades 19A, 20A, 20B, and 20C of the widely used Nextstrain classification (fig. S3A).

Independently, contact tracing surveillance assigns SARS-CoV-2 cases to epidemiological clusters based on the identification of transmission lines. In Austria, an extensive centralized tracing program was implemented during the COVID-19 outbreak. This program facilitated grouping of positive cases with a common exposure history and a comparable time frame of infection into epidemiological clusters. Integration of the phylogenetic analysis of Austrian SARS-CoV-2 sequences with epidemiological data resulted in strong overlap of these two lines of evidence, with 199 of the 345 sequences (65%) assigned to epidemiological clusters (data file S3). All sequenced samples from epidemiological cluster A mapped to the relatively homogeneous phylogenetic cluster Vienna-1 (Fig. 1C) with an index patient who had returned from Italy.

Our largest phylogenetic cluster, Tyrol-1 (fig. S3B), contained samples originating mainly from Austria's Tyrol region (73 of 90 samples) and overlapped with epidemiological cluster S (44 of 53 epidemiologically annotated samples). This phylogenetic cluster included resident and travel-associated cases to the ski resort Ischgl or the related valley Paznaun (Fig. 1C). Although different SARS-CoV-2 strains circulated in the region of Tyrol, these data suggest that epidemiological cluster S originated from a single strain with a characteristic mutation profile leading to a large outbreak in this region. To elucidate the possible origin of the SARS-CoV-2 strain giving rise to this cluster, we searched for sequences matching the viral mutation profile among global SARS-CoV-2 sequences (Fig. 1, D and E). Using phylogenetic analysis, we found that the mutation profile defining the Tyrol-1 cluster matched the definition of the global clade 20C of the Nextstrain classification (fig. S3C). This clade is predominantly populated by strains from North America.

To reveal possible transmission lines specifically between European countries in February and March 2020, we performed phylogenetic analysis using all 7731 European high-quality SARS-CoV-2 sequences sampled before 31 March that were available in the GISAID database (data file S2). Using this approach, we identified several samples matching the Tyrol-1 cluster mutation profile from a local outbreak in the region Hauts-de-France in the last week of February (*20*). Introduction of this SARS-CoV-2 strain to Iceland by tourists with a travel history to Austria was reported as early as 2 March (Fig. 1E and fig. S3C) (*11*), indicating that viruses with this mutational profile

Downloaded from https://www.science.org on August 02, 2022

Reprinted with permission from AAAS. License number: OP-00145573

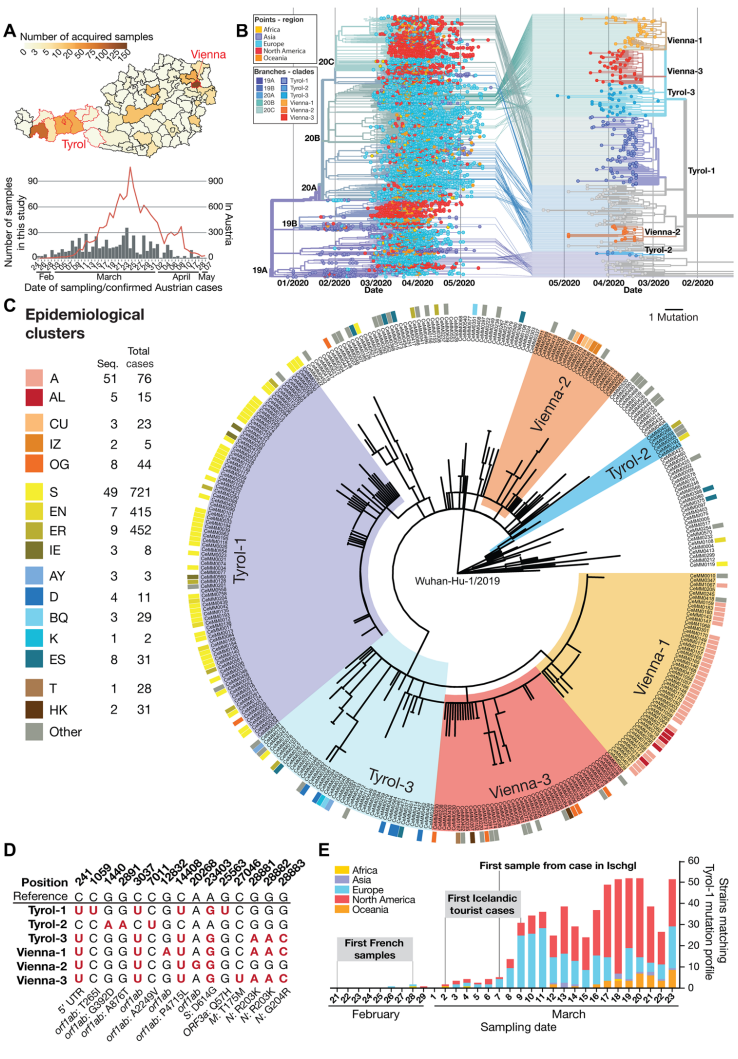


Fig. 1. Phylogenetic-epidemiological reconstruction of SARS-CoV-2 infection clusters in Austria. (**A**) Number of acquired samples per district in Austria (top) and sampling dates of samples that underwent viral genome sequencing in this study (bottom), plotted in the context of all confirmed cases (red line) in Austria. (**B**) Connection of Austrian strains to global clades of SARS-CoV-2. Points indicate the regional origin of a strain in the time-resolved phylogenetic tree from 7666 randomly subsampled sequences obtained from GISAID including 345 Austrian strains sequenced in this study (left). Lines from global phylogenetic tree (left) to phylogenetic tree of all Austrian strains obtained in this study (right) indicate the phylogenetic relation and Nextstrain clade assignment of Austrian strains. Color schemes of branches represent Nextstrain clade assignment (left) or phylogenetic clusters of Austrian strains (right). (**C**) Phylogenetic tree of SARS-CoV-2 strains from Austrian patients with COVID-19 sequenced in this study. Phylogenetic clusters were identified on the basis of characteristic mutation profiles in viral genome sequences of SARS-CoV-2–positive cases in Austria. Cluster names indicate the most abundant location of patients based on epidemiological data. The circular color code indicates the epidemiological cluster assigned to patients based on contact tracing. (**D**) Mutation profiles of phylogenetic clusters identified in this study. Positions with characteristic mutations compared to reference sequence "Wuhan-Hu-1" (GenBank: MN908947.3) are highlighted in red. Details regarding the affected genes or genomic regions and the respective codon and amino acid change are given below the table. (**E**) Timeline of the emergence of strains matching the mutation profile of the Tyrol-1 cluster in the global phylogenetic analysis by geographical distribution with additional information from European phylogenetic reconstruction.

were already present in Ischgl in the last week of February. These findings suggest that the emergence of cluster Tyrol-1 coincided with the local outbreak in France and with the early stages of the severe outbreak in northern Italy (*21*). The viral genomes observed in the Tyrol-1 cluster were closely related to those observed among the Icelandic cases with a travel history to Austria (fig. S3, D and E) (*11*). Vice versa, many of the Icelandic strains with a Tyrol-1 mutation profile had reported an Austrian or Icelandic exposure history (fig. S3F). Together, these observations and epidemiological evidence support the notion that the SARS-CoV-2 outbreak in Austria propagated to Iceland. Moreover, the emergence of these strains coincided with the emergence of the global clade 20C. One week after the occurrence of SARS-CoV-2 strains with this mutation profile in France and Ischgl, an increasing number of related strains based on the same mutation profile could be found across continents (Fig. 1E), for example, in New York City (*12*). As a popular skiing destination attracting thousands of international tourists, Ischgl may have played a critical role as transmission hub for the spread of clade 20C in Europe and beyond (fig. S3, G and H) (*12*). However, because of the lack of global epidemiological surveillance programs, it is rarely possible to infer direct transmission lines between countries.

Downloaded from https://www.science.org on August 02, 2022

Reprinted with permission from AAAS. License number: OP-00145573

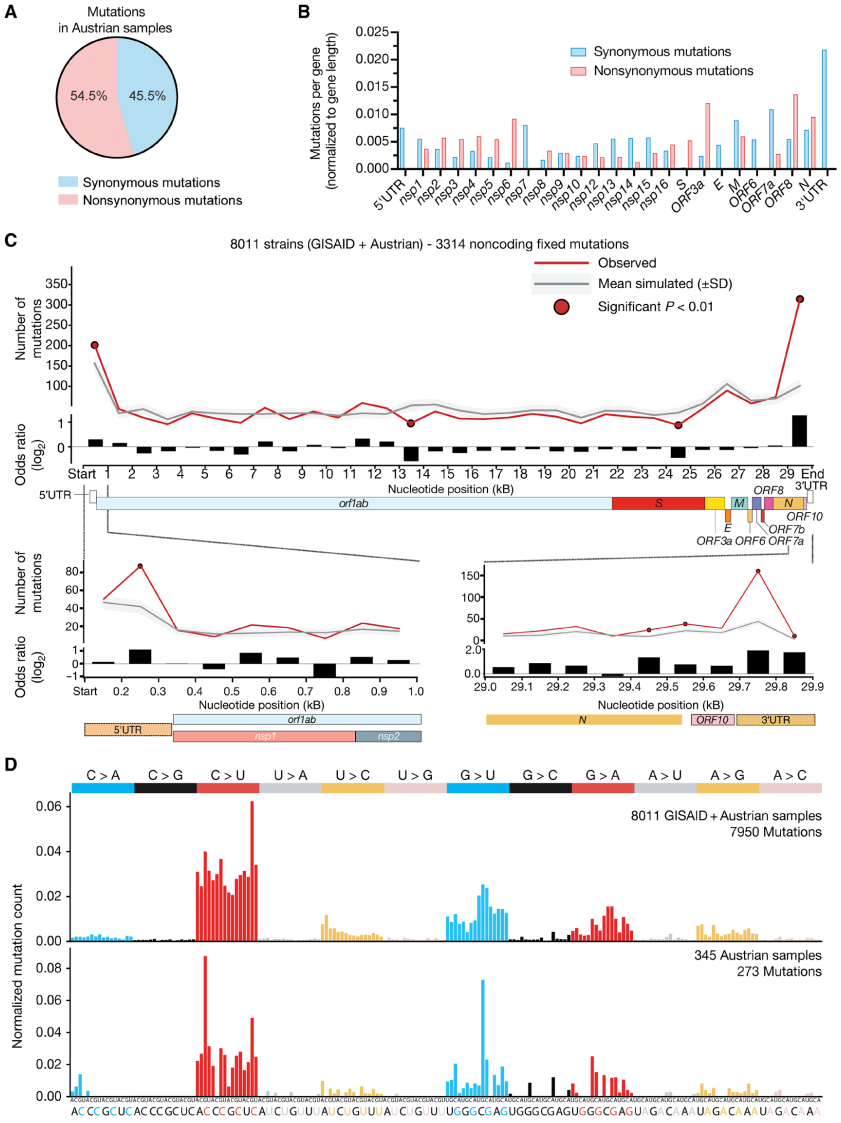


Fig. 2. Mutational analysis of fixed mutations in SARS-CoV-2 sequences. (**A**) Ratio of nonsynonymous to synonymous mutations in unique mutations identified in Austrian SARS-CoV-2 sequences. (**B**) Frequencies of synonymous and nonsynonymous mutations per gene or genomic region normalized to length of the respective gene, genomic region, or gene product (*nsp1-16*). (**C**) Mutational spectra panel. Mutational profile of interhost mutations. Relative probability of each trinucleotide change for mutations across SARS-CoV-2 sequences in 7666 global sequences obtained from GISAID samples plus 345 Austrian samples (top) or 345 SARS-CoV-2 sequences from Austrian patients with COVID-19 (bottom). (**D**) Mutation rate distribution along the SARS-CoV-2 genome. Top: A 1-kb window comparison of the observed number of synonymous mutations across the global subsample of 8011 SARS-CoV-2 sequences from GISAID compared with the expected distribution (based on 10^6 randomizations) according to their trinucleotide context. The gray line indicates the mean number of simulated mutations in the window, the colored background represents the distribution of expected mutations (mean ± SD), and red dots indicate a significant difference (G-test goodness of fit $P < 0.01$). Odds ratio in $\log_2$ scale of the observed compared with the expected number of synonymous mutations across the thirty 1-kb windows of the SARS-CoV-2 genome. Bottom: A zoom-in into the mutation rate across the first (left) and last (right) 1-kb windows. The comparisons were performed using ten 100–base pair windows. Gene annotations for SARS-CoV-2 genome are given below the top panel.

Our results integrating epidemiological and sequencing data emphasize that phylogenetic analyses of SARS-CoV-2 sequences empower robust tracing from interindividual to local and international spreading events (*12*). Both clusters Tyrol-1 and Vienna-1 originated from crowded indoor events (an Apré Ski bar and a sports class, respectively), which are now appreciated as high-risk situations for superspreading events.

Dynamics of low-frequency and fixed mutations in clusters

Next, we sought to uncover the mutational dynamics of SARS-CoV-2 during its transmission through the human population. We investigated the mutation profiles of our samples in terms of both fixed mutations (that drive the phylogenetic analyses) and the pool of low-frequency variants of each one of our samples. More than half of the fixed mutations in the Austrian SARS-CoV-2 genomes were

Downloaded from https://www.science.org on August 02, 2022

Reprinted with permission from AAAS. License number: OP-00145573

nonsynonymous (Fig. 2A), most frequently occurring in nonstructural protein 6 (*nsp6*), open reading frame 3a (*ORF3a*), and *ORF8* (Fig. 2, B and C). An analysis of mutational signatures in the 7666 global strains and the Austrian subset of SARS-CoV-2 isolates showed a heterogeneous mutational pattern dominated by C > U, G > U, and G > A substitutions (Fig. 2D).

We assessed the pool of variants for both low-frequency and fixed mutations (Fig. 3, A and B) and observed similar mutation patterns among these two sets of variants, which supports the accuracy of low-frequency mutation calling (Fig. 3, C and D). However, this pattern was lost for variants with an alternative frequency less than 0.01, which appear prone to false-positive variant calls. These results suggest that the same biological and evolutionary forces are at work for low-frequency and fixed mutations. Although the functional impact of variants across the genomes will need further research, we found that regions such as the 5′ untranslated region (5′UTR), which contains multiple stable RNA secondary structures, were subject to an increased mutation rate (Fig. 3D). Variants in the 5′UTR region are mainly localized along the stem-loop secondary structures (Fig. 3E). We found that 31% of the positions in the genome (9391 total positions) harbored variants (alternative allele frequency, ≥0.02) among the 420 sequenced strains from Austria and identified mutational hotspots for both high-frequency (≥0.5) and low-frequency (<0.5) mutations (Fig. 4A). Among these, 9034 positions exhibited only low-frequency mutations (<0.50), whereas four positions (241, 3037, 14,408, and 23,403) demonstrated fixation of the alternative allele in more than 50% of samples. We also identified 31 positions with alternative alleles being fixed in more than three samples and exhibiting a frequency <0.5 in at least two other samples (for example, 15,380 and 20,457).

On the basis of our phylogenetic analysis, we identified a subcluster inside the phylogenetic Tyrol-1 cluster that was defined by a fixed nonsynonymous G > U mutation at position 15,380 (Fig. 4B). This mutation was absent from all other Austrian cases but was detected at low and intermediary frequencies in other cases of the Tyrol-1 cluster. Around the time of emergence of this mutation, sequences sharing the same mutational profile (Tyrol-1 haplotype and G > U at position 15,380) appeared in other European countries including Denmark and Germany (Fig. 4C). Similarly, a synonymous fixed C > U mutation at position 20,457 defined a subcluster inside the phylogenetic Vienna-1 cluster (Fig. 4D). The cases from this subcluster intersected with members of two families (families 1 and 7) (Fig. 4E). Four members of family 1 tested positive for SARS-CoV-2 on 8 March and were epidemiologically assigned to cluster A. Yet, their viral sequences exhibited a wide range of C > U mutation frequencies at position 20,457 (0.00, 0.036, 0.24, and 1.00, respectively) (Fig. 4, D and E). Conversely, four members of family 7, who tested positive for SARS-CoV-2 between 16 and 22 March, were epidemiologically assigned to cluster AL and harbored viral genomes with a fixed U nucleotide at position 20,457 (Fig. 4, D and E).

Through several telephone interviews, we followed up with the members of both families to reconstruct the timeline of the infection events (data file S4). Both grandparents of family 1 were exposed to infected case CeMM1056 (node N13; sampling date 3 March) during a recreational indoor event on 28 February and subsequently tested positive for SARS-CoV-2 (Figs. 4, D and E, and 5A). The woman, CeMM0176 (node N16; sampling date 8 March), did not present a mutation at position 20,457, whereas her husband, CeMM1057 (node N15; sampling date 6 March), had the U allele at this position with a frequency of 0.036. The chain of transmission continued in family 1 with the infection of the couple CeMM0175 (node N18; sampling date 8 March) and CeMM0177 (node N17; sampling date 8 March), who had the U mutation at frequencies of 0.25 and 1, respectively. All further transmissions from CeMM1057 (node N15) resulted in a fixed mutation at position 20,457. CeMM1058 (node N25; sampling date 8 March) was in contact with CeMM1057 on 2 March and attended a funeral on 5 March with CeMM1059 (node N27; sampling date 11 March). On March 8, multiple persons participated at a birthday party, which included case CeMM1059 together with CeMM1062 (node N29; sampling date 13 March). Case CeMM1062 was part of a choir with multiple members of family 7 [CeMM0218 (node N31), CeMM0219 (node N32), and CeMM0217 (node N33)] on 10 March (Figs. 4, D and E, and 5A). Given our phylogenetic analysis and epidemiological reconstruction of transmission chains, we thus provide strong evidence for the emergence of a fixed mutation within a family and its spreading across previously disconnected epidemiological clusters. Together, these results from two superspreading events (Tyrol-1 and Vienna-1) demonstrate the power of deep viral genome sequencing in combination with detailed epidemiological data for observing viral mutation on their way from emergence at low frequency to fixation.

Impact of transmission bottlenecks and intrahost evolution on SARS-CoV-2 mutational dynamics

The emergence and potential fixation of mutations in the viral populations within a patient depend on interhost bottlenecks and intrahost evolutionary dynamics (*22*, *23*). An examination of the individual contributions of these forces requires pairs of samples from validated transmission events. For this purpose, we combined intrafamily cases, known epidemiological transmission chains, and subsequent telephone investigations to track the index cases as well as the context, date, and nature of each transmission event (Fig. 5A and data file S4) (*22*, *24*). Our set of SARS-CoV-2–positive cases comprised 39 epidemiologically confirmed infector-infectee pairs (Fig. 5A, fig. S4A, and data file S4).

One particularly well-defined network of SARS-CoV-2 transmission events linked cases from epidemiological cluster A and AL (Figs. 4E and 5A). The index case of cluster A is CeMM0003 (node N1), who contracted the virus during a visit to the north of Italy, further infecting his family members and, later, case CeMM0146 (node N3) during a dinner meeting (*17*). Multiple infections were linked to case CeMM0146 through an indoor sports activity. Among these cases was CeMM1056 (node N13), who further transmitted the virus to case CeMM1057 (node N15) as previously described for the 20,457 mutation linking cluster A and AL (Fig. 4E) (*17*). On the basis of these data, we investigated the transmission dynamics between known pairs of infectors and infectees by inferring the number of virions initiating the infection, also known as the genetic bottleneck size (*22*, *24*). The quality of the samples and the underlying low-frequency variants are critical for computing robust bottleneck sizes. In our data, samples with low Ct values (≤28) resulted in the detection of 38.6 variants (cutoff of 0.02) on average. Samples with high Ct values (>28) had on average 109.1 variants. The samples in the transmission chain were of high quality, with an average Ct value of 22.2, and only 9 of the 43 samples were higher than 28 (fig. S4A and data file S4).

Bottleneck size estimates were calculated by comparing the frequency of detected variants in each transmission pair (fig. S4, B to E).

Downloaded from https://www.science.org on August 02, 2022

Reprinted with permission from AAAS. License number: OP-00145573

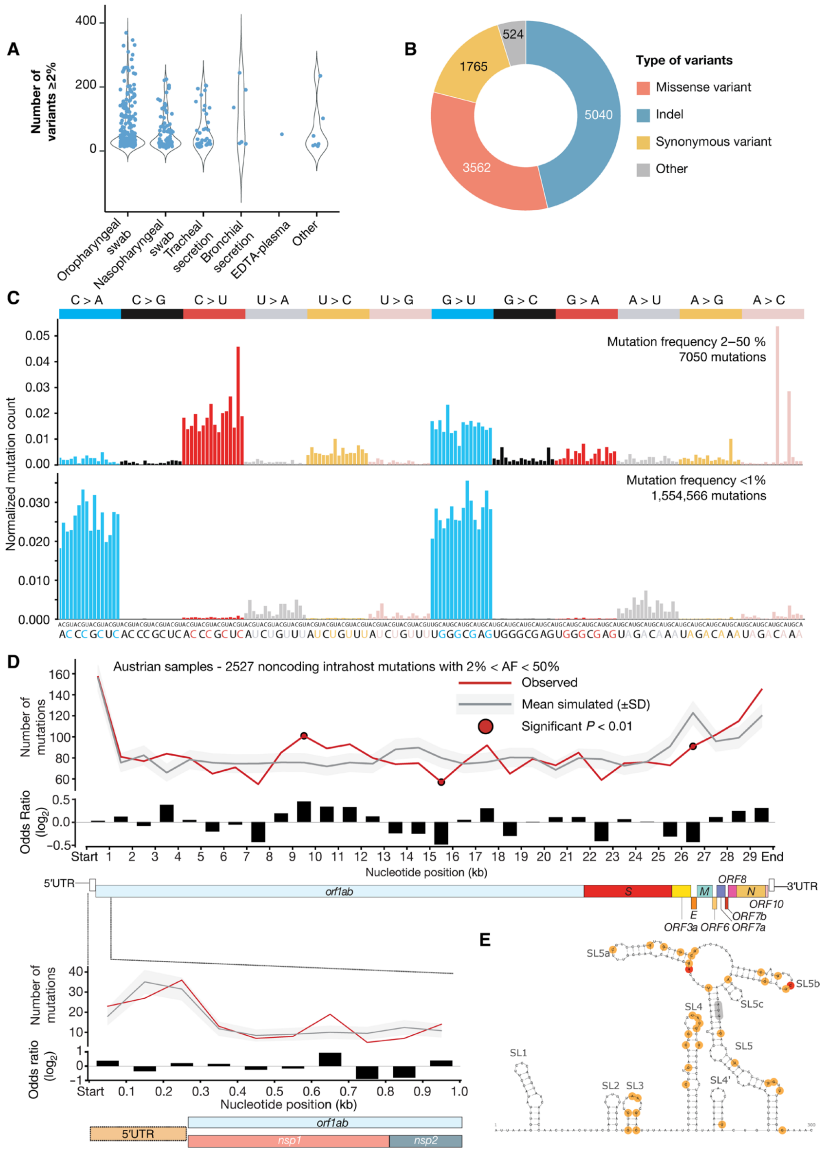


Fig. 3. Analysis of low-frequency mutations. (**A**) Number of variants detected across different sample types. (**B**) Number of variants per variant class. (**C**) Mutational profile (relative probability of each trinucleotide) of 7050 intrahost mutations across Austrian samples (allele frequencies between 0.02 and 0.05) (top). Mutational profile (relative probability of each trinucleotide) of 1,554,566 intrahost mutations across Austrian samples (allele frequencies <0.01) (bottom). (**D**) Analysis of the mutation rate (analogous to the interhost mutation rate panel) across the SARS-CoV-2 genome using 2527 intrahost nonprotein affecting mutations with allele frequencies between 0.02 and 0.5. (**E**) RNA secondary structure prediction of the upstream 300 nucleotides of the SARS-CoV-2 reference genome (NC 045512.2), comprising the complete 5′ untranslated region (UTR) and parts of the nsp1 protein nucleotide sequence. The canonical AUG start codon is located in a stacked region of SL5 (highlighted in gray). Mutational hotspots observed in the Austrian SARS-CoV-2 samples are highlighted: Two fixed mutations at positions 187 and 241, respectively, are marked in red, and low-frequency variants with an abundance between 0.02 and 0.5 in individual samples are shown in orange. Insertion and deletion variants are not shown.

Downloaded from https://www.science.org on August 02, 2022

Reprinted with permission from AAAS. License number: OP-00145573

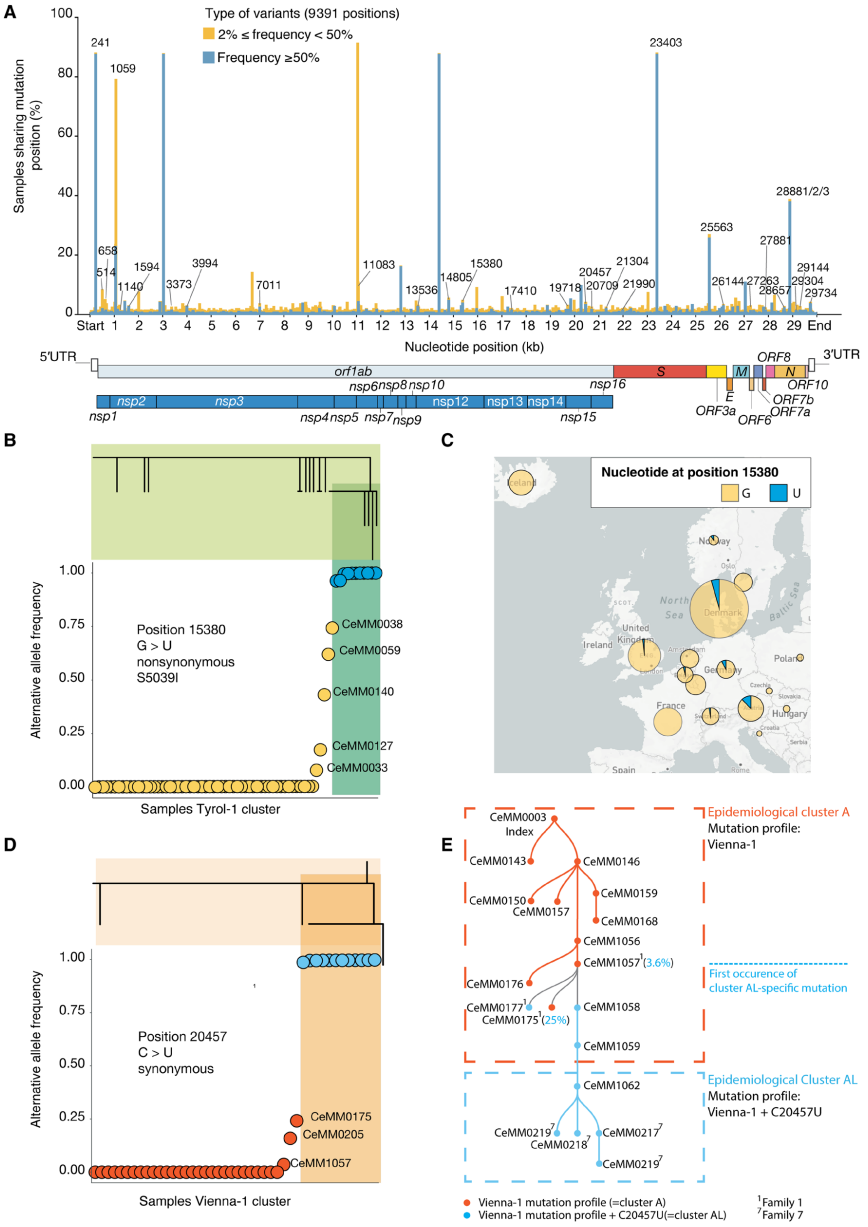


Fig. 4. Dynamics of low-frequency and fixed mutations in superspreading clusters. (**A**) Percentage of samples sharing detected (≥0.02) mutations across genomic positions. For each of the 9391 positions harboring an alternative allele, the percentage of samples with high (≥0.50) or low [0.02, 0.50] frequency are reported in dark blue and orange, respectively. (**B**) Allele frequency of nonsynonymous mutation G > U at position 15,380 across samples in the phylogenetic cluster Tyrol-1. This variant has been observed both as low-frequency variant and as fixed mutation, the latter defining a phylogenetic subcluster (dark green). (**C**) Proportion of European samples with a reference (yellow) or alternative (blue) allele at position 15,380. (**D**) Allele frequency of synonymous mutation C > U at position 20,457 across samples of the Vienna-1 phylogenetic cluster. This variant is fixed and defines a phylogenetic subcluster (dark orange) as part of the broader Vienna-1 cluster. (**E**) Schematic representation of the transmission lines between epidemiological cluster A and cluster AL was reconstructed on the basis of results from deep viral sequencing and case interviews. The transmission scheme is overlaid with epidemiological clusters and family-related information.

In particular, we computed bottleneck size using the beta-binomial method (*24*) and on three sets of alternative frequency cutoffs: [0.01, 0.95], [0.02, 0.95], and [0.03, 0.95] (fig. S4F and data file S4). Although the absolute values of the estimates were influenced by these cutoffs, their underlying average bottleneck sizes were comparable: 1227.59 (25 and 75% quartile: 21 to 2053.5; SD, 1692.235), 1110.513 (25 and 75% quartile: 2.5 to 2115; SD, 1661.183), and 1319.41 (25 and 75% quartile: 3.5 to 1763; SD, 1685.378) for the 0.01, 0.02, and 0.03 cutoffs, respectively (Fig. 5B and fig. S4G). In conclusion,

Downloaded from https://www.science.org on August 02, 2022

Reprinted with permission from AAAS. License number: OP-00145573

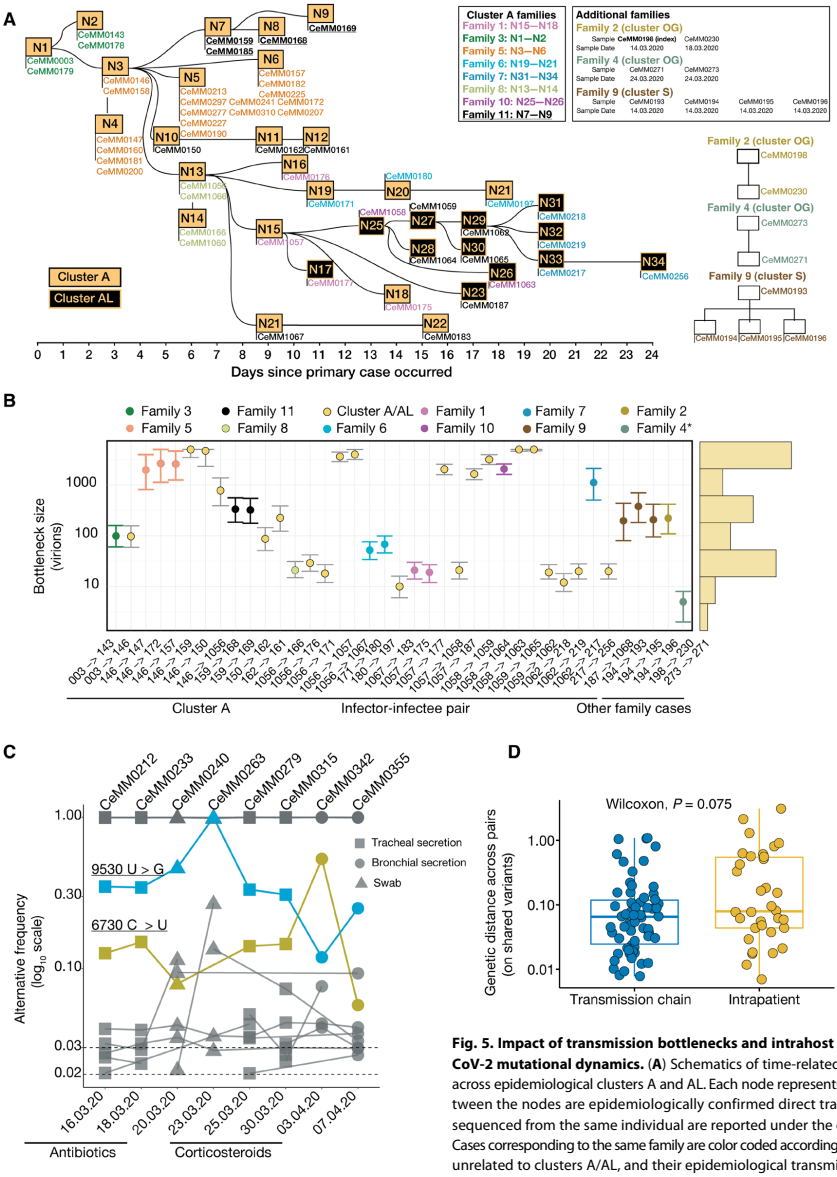


Fig. 5. Impact of transmission bottlenecks and intrahost evolution on SARS-CoV-2 mutational dynamics. (**A**) Schematics of time-related patient interactions across epidemiological clusters A and AL. Each node represents a case, and links between the nodes are epidemiologically confirmed direct transmissions. Samples sequenced from the same individual are reported under the corresponding node. Cases corresponding to the same family are color coded accordingly. Additional families, unrelated to clusters A/AL, and their epidemiological transmission details are also reported. (**B**) Bottleneck size (number of virions that initiate the infection in an infectee) estimation across infector-infectee pairs based on the transmission network depicted in (A), ordered according to the timeline of cluster A for the respective pairs, and with a cutoff of [0.01, 0.95] for alternative allele frequency. For patients with multiple samples, the earliest sample was considered for bottleneck size inference. Centered dots are maximum likelihood estimates, with 95% confidence intervals. A star (*) for family 4 indicates that the transmission line was inferred as detailed in Materials and Methods. The histogram (yellow bars) of all the bottleneck values is provided on the right side of the graph. (**C**) Alternative allele frequency (*y* axis) of mutations across available time points (*x* axis) for patient 5. Only variants with frequencies ≥0.02 and shared between at least two time points are shown. Two mutations increasing in frequency are color coded. (**D**) Genetic distance values of mutation frequencies between infector-infectee pairs (A and B) (transmission chains) and intrapatient consecutive time points [(C) and fig. S5D]. Only variants detected in two same-patient samples were considered.

Downloaded from https://www.science.org on August 02, 2022

Reprinted with permission from AAAS. License number: OP-00145573

taking advantage of a well-described and independently confirmed transmission network with 39 transmission events, we found that the number of viral particles transmitted from one individual to another that contributed productively to the infection was on average higher than 1000.

Last, we investigated the dynamics of intrahost evolution by using time-resolved viral sequences from 31 longitudinally sampled patients. These patients were subject to different medical treatments, and five of them succumbed to COVID-19–related complications (data file S5). To analyze intrahost viral dynamics, we focused on variants observed in at least two samples from the same patient. This approach resulted in a pool of high-confidence mutations (>0.02) with high coverage across same-patient samples (mean, 42,099 reads) (fig. S5A). Same-patient samples shared more variants than unrelated sample pairs (defined as non–same-patient, nor from the transmission chains) (fig. S5B). In addition, variants shared between samples from the same patient were unlikely to be found in unrelated samples (fig. S5C).

We observed diverse mutation patterns across individual patients and over time. Most patient samples showed a small number of stable low-frequency mutations ($\geq$0.02 and $\leq$0.50), whereas cases CeMM0108, CeMM0172, CeMM0251, CeMM0269, CeMM0299, and CeMM0221 exhibited higher variability, including the fixation and loss of individual mutations (Fig. 5C and fig. S5D). The patient-specific dynamics of viral mutation frequencies may reflect the effect of host-intrinsic factors such as immune responses or the patients' overall health, and extrinsic factors such as different treatment protocols. We also examined the genetic distance between samples obtained across infector-infectee pairs and serially acquired patient samples. However, the difference between increased genetic divergence of the virus within individual patients over the course of infection compared with interhost transmission was not significant ($P = 0.075$) (Fig. 5D).

DISCUSSION

Unprecedented global research efforts are underway to counter the COVID-19 pandemic around the globe and its pervasive impact on health and socioeconomics. These efforts include the genetic characterization of SARS-CoV-2 to track viral spread and to investigate the viral genome as it undergoes changes in the human population. Here, we leveraged deep viral genome sequencing in combination with national-scale epidemiological workup to reconstruct Austrian SARS-CoV-2 clusters that played a substantial role in the international spread of the virus. Our study describes how emerging low-frequency mutations of SARS-CoV-2 became fixed in local clusters, followed by viral spread across countries, thus connecting viral mutational dynamics within individuals and across populations. Exploiting our well-defined epidemiological clusters, we determined the interhuman genetic bottleneck size for SARS-CoV-2—which is the number of virions that start the infection and produce progeny in the viral population—at around 10^3. Our estimated bottlenecks are based on a substantial number of defined infector-infectee pairs and in agreement with recent studies implying larger bottleneck sizes for SARS-CoV-2 compared with estimates for the influenza A virus (*22*, *25*–*28*). These bottleneck sizes correlated inversely with higher mutation rates of influenza virus as compared with SARS-CoV-2.

In agreement with our experimentally determined bottleneck sizes, a recent preprint describing a dose-response modeling study estimated 3×10^2 to 2×10^3 SARS-CoV-2 virions necessary to initiate an infection (*29*). The dynamics of superspreading events seem to be driven by the number of interindividual contacts and the quantity of transmitted virus over time (*29*). Accordingly, our relatively large observed bottleneck size could be the result of patient exposure to high virus accumulations in shared and closed space and may have been influenced by a lack of protective measures in the early phase of the first COVID-19 wave in spring 2020. Although we inferred an average bottleneck size of 10^3 viral particles on average, the broad range of these values indicates that lower numbers of transmitted particles may also lead to a successful infection.

Our sequencing approach resulted in high-confidence variant calling and robust genome-wide coverage; hence, it is unlikely that technical limitations constituted a major source of bias. However, estimates of viral bottleneck sizes are likely influenced by many parameters not covered in this study, including virus-specific differences and stochastic evolutionary processes (*28*). Successful viral transmission also depends on other factors including the rate of decay of viral particles, frequency of susceptible cells, the host immune response, and comorbidities (*22*, *30*). The cases we analyzed were subject to different clinical contexts and treatments as well as disease outcomes. To better understand the mechanisms at work during infection, future investigations will need to probe these factors in the context of viral intrahost diversity across body compartments and time (*31*–*34*).

This study underscores the value of combining epidemiological approaches with virus genome sequencing to provide critical information to help public health experts track pathogen spread. Our genomic epidemiology analysis enabled the retrospective identification of SARS-CoV-2 chains of transmission and international hotspots such as the phylogenetic cluster Tyrol-1 (*14*, *35*–*37*). We also found that the Tyrol clusters were heterogeneous with regard to the S protein D614G mutation, which has been reported to contribute to viral transmissibility and fitness (*38*–*41*). Moreover, our phylogenetic analysis of the Vienna-1 cluster demonstrated the practical utility of viral genome sequencing data for uncovering previously unknown links between epidemiological clusters. This result was subsequently confirmed by follow-up contact tracing. We presented this case as an example of how the integration of contact tracing and sequencing information supports tracking the emergence and development of clusters. This demonstrates that deep viral genome sequencing can contribute directly to public health efforts by enhancing epidemiological surveillance.

Since the onset of the SARS-CoV-2 outbreak, many pandemic containment strategies have been implemented across the world. Where effective, these measures led to the reduction in the number of positive cases and limited superspreading events such as those investigated in this study. We found that most of the investigated infections likely involved the effective transmission of at least 1000 viral particles between individuals, suggesting that social distancing and mask wearing may be effective even when they cannot prevent the spread of all viral particles. As a future perspective, our study supports the relevance of investigating viral genome evolution of SARS-CoV-2 to enable informed decision-making by public health authorities (*42*).

MATERIALS AND METHODS

Study design

The goal of this study was to analyze mutational patterns in the SARS-CoV-2 genome to infer transmission in the human population

Downloaded from https://www.science.org on August 02, 2022

Reprinted with permission from AAAS. License number: OP-00145573

from interindividual to global scale. For this purpose, isolated viral RNA from 572 Austrian samples (February to May 2020) was processed for genome consensus sequence reconstruction and variant calling as approved by the ethics committee of the Medical University of Vienna. Additional analyses on subsets of samples consisted of the profiling of the mutational patterns across the genome and bottleneck size estimates based on transmission pairs. Data presented in this study are based on epidemiological and contact tracing data from the Austrian Department of Infection Epidemiology & Surveillance at the Austrian Agency for Health and Food Safety (AGES).

Sample collection and processing

Patient samples were obtained from the Medical Universities of Vienna Institute of Virology, Medical University of Innsbruck Institute of Virology, Medical University of Innsbruck Department of Internal Medicine II, Central Institute for Medical-Chemical Laboratory Diagnostics Innsbruck, Klinikum Wels-Grieskirchen, and AGES. Samples were obtained from suspected or confirmed SARS-CoV-2 cases or contact persons of these. Sample types included oropharyngeal swabs, nasopharyngeal swabs, tracheal secretion, bronchial secretion, serum, plasma, and cell culture supernatants. RNA was extracted using the following commercially available kits by adhering to the manufacturers' instructions: MagMax (Thermo Fisher Scientific), EasyMag (bioMérieux), AltoStar Purification Kit 1.5 (Altona Diagnostics), MagNA Pure LC 2.0 (Roche), MagNA Pure Compact (Roche), and QIAsymphony (Qiagen). Viral RNA was reverse transcribed with Superscript IV Reverse Transcriptase (Thermo Fisher Scientific). The resulting complementary DNA was used to amplify viral sequences with modified primer pools from the Artic Network initiative (*43*). Polymerase chain reactions were pooled and subjected to high-throughput sequencing.

Sample sequencing

Amplicons were cleaned up with AMPure XP beads (Beckman Coulter) with a 1:1 ratio. Amplicon concentrations were quantified with the Qubit Fluorometric Quantitation system (Life Technologies), and the size distribution was assessed using the 2100 Bioanalyzer system (Agilent). Amplicon concentrations were normalized, and sequencing libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs) according to the manufacturer's instructions. Library concentrations again were quantified with the Qubit Fluorometric Quantitation system (Life Technologies), and the size distribution was assessed using the 2100 Bioanalyzer system (Agilent). For sequencing, samples were pooled into equimolar amounts. Amplicon libraries were sequenced on the NovaSeq 6000 platform (Illumina) using S Prime (SP) flowcell with a read length of 2 × 250 base pairs in paired-end mode.

Sequencing data processing and analysis

Following demultiplexing, fastq files containing the raw reads were inspected for quality criteria (base quality, N and GC content, sequence duplication, and overrepresented sequences) using FastQC (v.0.11.8) (*44*). Trimming of adapter sequences was performed with BBDUK from the BBtools suite (http://jgi.doe.gov/data-and-tools/bbtools). Overlapping read sequences within a pair were corrected for using BBMERGE function from BBTools. Read pairs were mapped on the combined Hg38 and SARS-CoV-2 genome (GenBank: MN908947.3, RefSeq: NC_045512.2) using the BWA-MEM software package with a minimal seed length of 17 (v0.7.17) (*45*). BWA-MEM accounts for mismatches, insertions, and deletions in the alignment score and the mapping quality. Only reads mapping uniquely to the SARS-CoV-2 viral genome were retained. Primer sequences were removed after mapping by masking with iVar (*46*). From the viral reads BAM (binary alignment map) file, the consensus FASTA file was generated using Samtools (v1.9) (*47*), mpileup, Bcftools (v 1.9) (*47*), and SEQTK (https://github.com/lh3/seqtk). For calling low-frequency variants, the viral read alignment file was realigned using the Viterbi method provided by LoFreq (v2.1.2) (*48*). After adding InDel qualities, low-frequency variants were called using LoFreq. Variant filtering was performed with LoFreq and Bcftools (v1.9) (*49*). Only variants with a minimum coverage of 75 reads, a minimum phred value of 90, and indels (insertions and deletions) with an HRUN of minimum 4 were considered. All analyses except for the control analysis in Fig. 3C were performed on variants with a minimum alternative frequency of 0.01. The cutoff for the alternative frequency mainly used in this study was set to 0.02, except for Fig. 5B. Annotations of the variants were performed with SnpEff (v4.3) (*50*) and SnpSift (v4.3) (*51*).

Epidemiological analyses and identification of SARS-CoV-2 infection clusters

The investigation of transmission chains (contact tracing) was conducted by the Department of Infection Epidemiology & Surveillance at the AGES. Epidemiological clusters were defined as accumulations of cases within a certain time period in a defined region and with common source of exposure. The required information for cluster annotation and resolution in chains of transmission was collected during the official case contact tracing by the public health authorities, resulting in identification of the most likely source cases and successive cases of the index cases. Contact tracing was performed according to technical guidance relating to this measure produced by the European Centre for Disease Prevention and Control (ECDC) (*52*). For refinement and validation of contact tracing data for cluster A and cluster AL, we contacted 17 cases for 15-minute interviews. The interviews comprised 10 questions concerning the most likely source, time, place, and setting of transmission, contact persons, and the course of disease (start and end of symptoms, kind of symptoms, severity, and hospitalization).

Phylogenetic analysis and inference of transmission lines

Phylogenetic analysis was conducted using the Augur package (version 7.0.2) (*53*). We compiled a randomly subsampled dataset of 7666 full-length viral genomes with high coverage (<1% Ns) that were available from GISAID (https://gisaid.org/, 2 June) and the 345 sequences obtained in this publication. GISAID sequences were filtered for entries from human hosts with complete sampling dates. Metadata information for patient age and sex was excluded from the analysis. Multiple sequence alignments were performed using mafft (*54*). A masking scheme for homoplasic and highly ambiguous sites was applied to avoid bias in the following phylogenetic analysis as discussed elsewhere (*55*). We reconstructed the phylogeny with the augur pipeline using IQ-TREE (*54*) and further processed the resulting trees with treetime to infer ancestral traits of the nodes (*56*). Phylogenetic trees were rooted with the genome of "Wuhan-Hu-1/2019." The same workflow was repeated for phylogenetic reconstruction of all high-quality European strains before 31 March 2020 available in the GISAID database by 7 June 2020 (7731). Clade annotations

Downloaded from https://www.science.org on August 02, 2022

Reprinted with permission from AAAS. License number: OP-00145573

for global trees were adapted from nextstrain.org (https://github.com/nextstrain/ncov/blob/master/defaults/clades.tsv; https://clades.nextstrain.org/); clusters of Austrian strains were identified on the basis of shared mutation profiles and patient location from epidemiological data.

Bottleneck estimation

Our analysis to estimate the transmission bottleneck sizes for each infector-infectee pair was based on the beta-binomial method presented in (*24*). For a given variant present in the infector, this method assumes that the number of transmitted virions carrying the variant is binomially distributed with the bottleneck size as the number of trials and success probability as the variant frequency in the infector. Following transmission, the viral population during early infection is modeled as a linear birth-death process, implying that the proportion of the viral population descended from any virion in the bottleneck population is beta-distributed. Using this model for the change in variant frequencies between infector and infectee pairs and assuming independence of mutations lead to the likelihood model of (*24*). Maximum likelihood analysis then provides the bottleneck statistics. Error bars denote 95% confidence intervals, determined by a likelihood ratio test. This method was applied to variants in the following frequency ranges: [0.01, 0.95], [0.02, 0.95], and [0.03, 0.95]. Because of the high sequencing depth of our study, we used the approximate version of the beta-binomial method.

Intrapatient time series analyses

Among our 420 high-quality SARS-CoV-2–positive samples, we had 31 unique cases with multiple time-point samplings (a total of 106 samples). Nineteen of 31 cases had only two samples per patient. For each of the 31 cases, we only considered variants with an alternative frequency greater than 0.02 and that were shared across at least two of the intrapatient samples. We retrieved the depth of coverage of the selected variants for each sample for each patient. To compare how many variants were shared intrapatient as opposed to unrelated samples, we first identified potentially unrelated cases by eliminating all samples from the same patient, as well as all the samples in the transmission chains in Fig. 5A, resulting in 281 samples hereafter termed "unrelated." We then enumerated all 39,340 unordered pairs of the 281 unrelated samples. Only variants between 0.02 and 0.5 were considered. We computed the percentage of variants shared by each pair out of the total number detected across the two samples. We then compared the percentage of variant sharing between intrapatient and unrelated pairs of samples with a Wilcoxon test. To test how widely the intrapatient variants ([0.02, 0.5]; 173 positions) were detected in other samples, we examined how often they were detected in the pool of 218 unrelated samples.

Genetic distance

For shared mutations with defined infector-to-infectee transmission, we determined those mutations present in both samples and calculated their absolute difference in frequency. Similarly, we performed the same computations between time consecutive pairs for serially sampled patients. If multiple samples were obtained on the same day, the sample with the lowest Ct value was considered. Note that the time-consecutive pairs had a differing number of days between samples. To these genetic distances obtained from the shared variants, we added the sum of the frequencies of the variants detected in only one of the pairs of shared samples; that is, we calculated the l1-norm of the variant frequencies. Statistical difference between the genetic distances from transmission pairs versus consecutive pairs from serially sampled patients was determined by a Wilcoxon (one-sided) rank sum test.

Statistical methods

Control samples were compared with a linear regression method, and the corresponding R^2 was reported. For mutational patterns analyses, a statistical test was devised to compare the deviation of the observed number of mutations from the expected distribution as detailed in Materials and Methods. The frequency of mutations in overlapping windows across the genome was statistically assessed with a log-likelihood test. For bottleneck size computations, a maximum likelihood approach was applied. The comparison of genetic diversity between groups was performed with a standard Wilcoxon test. Significance was inferred for P values ≤ 0.05.

SUPPLEMENTARY MATERIALS

stm.sciencemag.org/cgi/content/full/12/573/eabe2555/DC1
Materials and Methods
Fig. S1. Data overview.
Fig. S2. Technical pipeline and controls.
Fig. S3. Phylogenetic analysis of SARS-CoV-2 sequences from Austrian patients with COVID-19 in global context.
Fig. S4. Bottleneck size estimations.
Fig. S5. Viral intrahost diversity in individual patients.
Data file S1. Sample and sequencing information of the 572 samples and controls.
Data file S2. Acknowledgments for SARS-CoV-2 genome sequences derived from GISAID.
Data file S3. Epidemiological clusters referred to in this study.
Data file S4. Transmission chain and sample information for cluster A/cluster AL and family-related cases.
Data file S5. Clinical information of patients with COVID-19 relating to Fig. 5 and fig S5.
Reference (*57*)

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G.-F. Xiao, Z.-L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
2. E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020).
3. N. Vabret, G. J. Britton, C. Gruber, S. Hegde, J. Kim, M. Kuksin, R. Levantovsky, L. Malle, A. Moreira, M. D. Park, L. Pia, E. Risson, M. Saffern, B. Salomé, M. E. Selvan, M. P. Spindler, J. Tan, V. van der Heide, J. K. Gregory, K. Alexandropoulos, N. Bhardwaj, B. D. Brown, B. Greenbaum, Z. H. Gümüş, D. Homann, A. Horowitz, A. O. Kamphorst, M. A. C. de Lafaille, S. Mehandru, M. Merad, R. M. Samstein; Sinai Immunology Review Project, Immunology of COVID-19: Current state of the science. *Immunity* **52**, 910–941 (2020).
4. D. Mathew, J. R. Giles, A. E. Baxter, D. A. Oldridge, A. R. Greenplate, J. E. Wu, C. Alanio, L. Kuri-Cervantes, M. B. Pampena, K. D'Andrea, S. Manne, Z. Chen, Y. J. Huang, J. P. Reilly, A. R. Weisman, C. A. G. Ittner, O. Kuthuru, J. Dougherty, K. Nzingha, N. Han, J. Kim, A. Pattekar, E. C. Goodwin, E. M. Anderson, M. E. Weirick, S. Gouma, C. P. Arevalo, M. J. Bolton, F. Chen, S. F. Lacey, H. Ramage, S. Cherry, S. E. Hensley, S. A. Apostolidis, A. C. Huang, L. A. Vella; UPenn COVID Processing Unit, M. R. Betts, N. J. Meyer, E. J. Wherry, Deep immune profiling of COVID-19 patients reveals distinct immunotypes with therapeutic implications. *Science* **369**, eabc8511 (2020).
5. X. Zhang, Y. Tan, Y. Ling, G. Lu, F. Liu, Z. Yi, X. Jia, M. Wu, B. Shi, S. Xu, J. Chen, W. Wang, B. Chen, L. Jiang, S. Yu, J. Lu, J. Wang, M. Xu, Z. Yuan, Q. Zhang, X. Zhang, G. Zhao, S. Wang, S. Chen, H. Lu, Viral and host factors related to the clinical outcome of COVID-19. *Nature* **583**, 437–440 (2020).
6. Severe Covid-19 GWAS Group, D. Ellinghaus, F. Degenhardt, L. Bujanda, M. Buti, A. Albillos, P. Invernizzi, J. Fernández, D. Prati, G. Baselli, R. Asselta, M. M. Grimsrud, C. Milani, F. Aziz, J. Kässens, S. May, M. Wendorff, L. Wienbrandt, F. Uellendahl-Werth, T. Zheng, X. Yi, R. de Pablo, A. G. Chercoles, A. Palom, A.-E. Garcia-Fernandez,

Downloaded from https://www.science.org on August 02, 2022

Reprinted with permission from AAAS. License number: OP-00145573

F. Rodriguez-Frias, A. Zanella, A. Bandera, A. Protti, A. Aghemo, A. Lleo, A. Biondi, A. Caballero-Garralda, A. Gori, A. Tanck, A. C. Nolla, A. Latiano, A. L. Fracanzani, A. Peschuck, A. Julià, A. Pesenti, A. Voza, D. Jiménez, B. Mateos, B. N. Jimenez, C. Quereda, C. Paccapelo, C. Gassner, C. Angelini, C. Cea, A. Solier, D. Pestaña, E. Muñiz-Diaz, E. Sandoval, E. M. Paraboschi, E. Navas, F. García Sánchez, F. Ceriotti, F. Martinelli-Boneschi, F. Peyvandi, F. Blasi, L. Téllez, A. Blanco-Grau, G. Hemmrich-Stanisak, G. Grasselli, G. Costantino, G. Cardamone, G. Foti, S. Aneli, H. Kurihara, H. ElAbd, I. My, I. Galván-Femenia, J. Martín, J. Erdmann, J. Ferrusquía-Acosta, K. Garcia-Etxebarria, L. Izquierdo-Sanchez, L. R. Bettini, L. Sumoy, L. Terranova, L. Moreira, L. Santoro, L. Scudeller, F. Mesonero, L. Roade, M. C. Rühlemann, M. Schaefer, M. Carrabba, M. Riveiro-Barciela, M. E. F. Basso, M. G. Valsecchi, M. Hernandez-Tejero, M. Acosta-Herrera, M. D'Angiò, M. Baldini, M. Cazzaniga, M. Schulzky, M. Cecconi, M. Wittig, M. Ciccarelli, M. Rodríguez-Gandía, M. Bocciolone, M. Miozzo, N. Montano, N. Braun, N. Sacchi, N. Martínez, O. Özer, O. Palmieri, P. Faverio, P. Preatoni, P. Bonfanti, P. Omodei, P. Tentorio, P. Castro, P. M. Rodrigues, A. B. Ortiz, R. de Cid, R. Ferrer, R. Gualtierotti, R. Nieto, S. Goerg, S. Badalamenti, S. Marsal, G. Matullo, S. Pelusi, S. Juzenas, S. Aliberti, V. Monzani, V. Moreno, T. Wesse, T. L. Lenz, T. Pumarola, V. Rimoldi, S. Bosari, W. Albrecht, W. Peter, M. Romero-Gómez, M. D'Amato, S. Duga, J. M. Banales, J. R. Hov, T. Folseraas, L. Valenti, A. Franke, T. H. Karlsen, Genomewide association study of severe COVID-19 with respiratory failure. *N. Engl. J. Med.* **383**, 1522–1534 (2020).

7. J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, W. M. Getz, Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359 (2005).
8. T. M. McMichael, D. W. Currie, S. Clark, S. Pogosjans, M. Kay, N. G. Schwartz, J. Lewis, A. Baer, V. Kawakami, M. D. Lukoff, J. Ferro, C. Brostrom-Smith, T. D. Rea, M. R. Sayre, F. X. Riedo, D. Russell, B. Hiatt, P. Montgomery, A. K. Rao, E. J. Chow, F. Tobolowsky, M. J. Hughes, A. C. Bardossy, L. P. Oakley, J. R. Jacobs, N. D. Stone, S. C. Reddy, J. A. Jernigan, M. A. Honein, T. A. Clark, J. S. Duchin; Public Health-Seattle and King County, EvergreenHealth, and CDC COVID-19 Investigation Team, Epidemiology of COVID-19 in a long-term care facility in King County, Washington. *N. Engl. J. Med.* **382**, 2005–2011 (2020).
9. D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, Y. Zhao, Y. Li, X. Wang, Z. Peng, Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA* **323**, 1061–1069 (2020).
10. L. Hamner, P. Dubbel, I. Capron, A. Ross, A. Jordan, J. Lee, J. Lynn, A. Ball, S. Narwal, S. Russell, D. Patrick, H. Leibrand, High SARS-CoV-2 attack rate following exposure at a choir practice—Skagit County, Washington, March 2020. *MMWR. Morb. Mortal. Wkly. Rep.* **69**, 606–610 (2020).
11. D. F. Gudbjartsson, A. Helgason, H. Jonsson, O. T. Magnusson, P. Melsted, G. L. Norddahl, J. Saemundsdottir, A. Sigurdsson, P. Sulem, A. B. Agustsdottir, B. Eiriksdottir, R. Fridriksdottir, E. E. Gardarsdottir, G. Georgsson, O. S. Gretarsdottir, K. R. Gudmundsson, T. R. Gunnarsdottir, A. Gylfason, H. Holm, B. O. Jensson, A. Jonasdottir, F. Jonsson, K. S. Josefsdottir, T. Kristjansson, D. N. Magnusdottir, L. le Roux, G. Sigmundsdottir, G. Sveinbjornsson, K. E. Sveinsdottir, M. Sveinsdottir, E. A. Thorarensen, B. Thorbjornsson, A. Löve, G. Masson, I. Jonsdottir, A. D. Möller, T. Gudnason, K. G. Kristinsson, U. Thorsteinsdottir, K. Stefansson, Spread of SARS-CoV-2 in the Icelandic population. *N. Engl. J. Med.* **382**, 2302–2315 (2020).
12. A. S. Gonzalez-Reiche, M. M. Hernandez, M. J. Sullivan, B. Ciferri, H. Alshammary, A. Obla, S. Fabre, G. Kleiner, J. Polanco, Z. Khan, B. Alburquerque, A. van de Guchte, J. Dutta, N. Francoeur, B. S. Melo, I. Oussenko, G. Deikus, J. Soto, S. H. Sridhar, Y.-C. Wang, K. Twyman, A. Kasarskis, D. R. Altman, M. Smith, R. Sebra, J. Aberg, F. Krammer, A. García-Sastre, M. Luksza, G. Patel, A. Paniz-Mondolfi, M. Gitman, E. M. Sordillo, V. Simon, H. van Bakel, Introductions and early spread of SARS-CoV-2 in the New York City area. *Science* **369**, 297–301 (2020).
13. R. Pung, C. J. Chiew, B. E. Young, S. Chin, M. I.-C. Chen, H. E. Clapham, A. R. Cook, S. Maurer-Stroh, M. P. H. S. Toh, C. Poh, M. Low, J. Lum, V. T. J. Koh, T. M. Mak, L. Cui, R. V. T. P. Lin, D. Heng, Y.-S. Leo, D. C. Lye, V. J. M. Lee; Singapore 2019 Novel Coronavirus Outbreak Research Team, Investigation of three clusters of COVID-19 in Singapore: Implications for surveillance and response measures. *Lancet* **395**, 1039–1046 (2020).
14. X. Deng, W. Gu, S. Federman, L. du Plessis, O. G. Pybus, N. R. Faria, C. Wang, G. Yu, B. Bushnell, C.-Y. Pan, H. Guevara, A. Sotomayor-Gonzalez, K. Zorn, A. Gopez, V. Servellita, E. Hsu, S. Miller, T. Bedford, A. L. Greninger, P. Roychoudhury, L. M. Starita, M. Famulare, H. Y. Chu, J. Shendure, K. R. Jerome, C. Anderson, K. Gangavarapu, M. Zeller, E. Spencer, K. G. Andersen, D. MacCannell, C. R. Paden, Y. Li, J. Zhang, S. Tong, G. Armstrong, S. Morrow, M. Willis, B. T. Matyas, S. Mase, O. Kasirye, M. Park, G. Masinde, C. Chan, A. T. Yu, S. J. Chai, E. Villarino, B. Bonin, D. A. Wadford, C. Y. Chiu, Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California. *Science* **369**, 582–587 (2020).
15. M. Vignuzzi, J. K. Stone, J. J. Arnold, C. E. Cameron, R. Andino, Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* **439**, 344–348 (2006).
16. R. Andino, E. Domingo, Viral quasispecies. *Virology* **479–480**, 46–51 (2015).
17. P. Kreidl, D. Schmid, S. Maritschnik, L. Richter, W. Borena, J.-W. Genger, A. Popa, T. Penz, C. Bock, A. Bergthaler, F. Allerberger, Emergence of coronavirus disease 2019 (COVID-19) in Austria. *Wien. Klin. Wochenschr.* **132**, 645–652 (2020).
18. A. Bluhm, E. Al, M. Christandl, F. Gesmundo, F. R. Klausen, L. Mančinska, V. Steffan, D. S. França, A. H. Werner, SARS-CoV-2 transmission chains from genetic data: A Danish case study. *bioRxiv* 2020.05.29.123612 (2020).
19. C. L. Correa-Martínez, S. Kampmeier, P. Kümpers, V. Schwierzeck, M. Hennies, W. Hafezi, J. Kühn, H. Pavenstädt, S. Ludwig, A. Mellmann, A pandemic in times of global tourism: Superspreading and exportation of COVID-19 cases from a ski area in Austria. *J. Clin. Microbiol.* **58**, e00588-20 (2020).
20. H. Salje, C. Tran Kiem, N. Lefrancq, N. Courtejoie, P. Bosetti, J. Paireau, A. Andronico, N. Hozé, J. Richet, C.-L. Dubost, Y. Le Strat, J. Lessler, D. Levy-Bruhl, A. Fontanet, L. Opatowski, P.-Y. Boelle, S. Cauchemez, Estimating the burden of SARS-CoV-2 in France. *Science* **369**, 208–211 (2020).
21. A. R. Tuite, V. Ng, E. Rees, D. Fisman, Estimation of COVID-19 outbreak size in Italy. *Lancet Infect. Dis.* **20**, 537 (2020).
22. M. P. Zwart, S. F. Elena, Matters of size: Genetic bottlenecks in virus infection and their potential impact on evolution. *Annu. Rev. Virol.* **2**, 161–179 (2015).
23. J. L. Geoghegan, A. M. Senior, E. C. Holmes, Pathogen population bottlenecks and adaptive landscapes: Overcoming the barriers to disease emergence. *Proc. Biol. Sci.* **283**, 20160727 (2016).
24. A. Sobel Leonard, D. B. Weissman, B. Greenbaum, E. Ghedin, K. Koelle, Transmission bottleneck size estimation from pathogen deep-sequencing data, with an application to human influenza A virus. *J. Virol.* **91**, e00171-17 (2017).
25. K. A. Lythgoe, M. Hall, L. Ferretti, M. de Cesare, G. MacIntyre-Cockett, A. Trebes, M. Andersson, N. Otecko, E. L. Wise, N. Moore, J. Lynch, S. Kidd, N. Cortes, M. Mori, A. Justice, A. Green, M. A. Ansari, L. Abeler-Dorner, C. E. Moore, T. E. A. Peto, R. Shaw, P. Simmonds, D. Buck, J. A. Todd; OVSG Analysis Group, D. Bonsall, C. Fraser, T. Golubchik, Shared SARS-CoV-2 diversity suggests localised transmission of minority variants. *bioRxiv* 2020.05.28.118992 (2020).
26. S. Pfefferle, T. Günther, R. Kobbe, M. Czech-Sioli, D. Nörz, R. Santer, J. Oh, S. Kluge, L. Oestereich, K. Peldschus, D. Indenbirken, J. Huang, A. Grundhoff, M. Aepfelbacher, J. K. Knobloch, M. Lütgehetmann, N. Fischer, SARS-CoV-2 variant tracing within the first coronavirus disease 19 clusters in Northern Germany. *Clin. Microbiol. Infect.* j. cmi.2020.09.034 , (2020).
27. L. L. M. Poon, T. Song, R. Rosenfeld, X. Lin, M. B. Rogers, B. Zhou, R. Sebra, R. A. Halpin, Y. Guan, A. Twaddle, J. V. DePasse, T. B. Stockwell, D. E. Wentworth, E. C. Holmes, B. Greenbaum, J. S. M. Peiris, B. J. Cowling, E. Ghedin, Quantifying influenza virus diversity and transmission in humans. *Nat. Genet.* **48**, 195–200 (2016).
28. J. T. McCrone, R. J. Woods, E. T. Martin, R. E. Malosh, A. S. Monto, A. S. Lauring, Stochastic processes constrain the within and between host evolution of influenza virus. *eLife* **7**, e35962 (2018).
29. M. Prentiss, A. Chu, K. K. Berggren, Superspreading events without superspreaders: Using high attack rate events to estimate N_0 for airborne transmission of COVID-19. *medRxiv* 2020.10.21.20216895 , (2020).
30. X. He, E. H. Y. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, X. Mo, Y. Chen, B. Liao, W. Chen, F. Hu, Q. Zhang, M. Zhong, Y. Wu, L. Zhao, F. Zhang, B. J. Cowling, F. Li, G. M. Leung, Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 1491–1493 (2020).
31. R. Wölfel, V. M. Corman, W. Guggemos, M. Seilmaier, S. Zange, M. A. Müller, D. Niemeyer, T. C. Jones, P. Vollmar, C. Rothe, M. Hoelscher, T. Bleicker, S. Brünink, J. Schneider, R. Ehmann, K. Zwirglmaier, C. Drosten, C. Wendtner, Virological assessment of hospitalized patients with COVID-2019. *Nature* **581**, 465–469 (2020).
32. Y. Wang, D. Wang, L. Zhang, W. Sun, Z. Zhang, W. Chen, A. Zhu, Y. Huang, F. Xiao, J. Yao, M. Gan, F. Li, L. Luo, X. Huang, Y. Zhang, S.-s. Wong, X. Cheng, J. Ji, Z. Ou, M. Xiao, M. Li, J. Li, P. Ren, Z. Deng, H. Zhong, H. Yang, J. Wang, X. Xu, T. Song, C. K. P. Mok, M. Peiris, N. Zhong, J. Zhao, Y. Li, J. Li, J. Zhao, Intra-host variation and evolutionary dynamics of SARS-CoV-2 population in COVID-19 patients. *bioRxiv* 2020.05.20.103549 , (2020).
33. S. L. Díaz-Muñoz, R. Sanjuán, S. West, Sociovirology: Conflict, cooperation, and communication among viruses. *Cell Host Microbe* **22**, 437–441 (2017).
34. M. A. Nowak, C. E. Tarnita, T. Antal, Evolutionary dynamics in structured populations. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 19–30 (2010).
35. J. R. Fauver, M. E. Petrone, E. B. Hodcroft, K. Shioda, H. Y. Ehrlich, A. G. Watts, C. B. F. Vogels, A. F. Brito, T. Alpert, A. Muyombwe, J. Razeq, R. Downing, N. R. Cheemarla, A. L. Wyllie, C. C. Kalinich, I. M. Ott, J. Quick, N. J. Loman, K. M. Neugebauer, A. L. Greninger, K. R. Jerome, P. Roychoudhury, H. Xie, L. Shrestha, M.-L. Huang, V. E. Pitzer, A. Iwasaki, S. B. Omer, K. Khan, I. I. Bogoch, R. A. Martinello, E. F. Foxman, M. L. Landry, R. A. Neher, A. I. Ko, N. D. Grubaugh, Coast-to-coast spread of SARS-CoV-2 during the early epidemic in the United States. *Cell* **181**, 990–996.e5 (2020).
36. M. M. Böhmer, U. Buchholz, V. M. Corman, M. Hoch, K. Katz, D. V. Marosevic, S. Böhm, T. Woudenberg, N. Ackermann, R. Konrad, U. Eberle, B. Treis, A. Dangel, K. Bengs,

Downloaded from https://www.science.org on August 02, 2022

Reprinted with permission from AAAS. License number: OP-00145573

V. Fingerle, A. Berger, S. Hörmansdorfer, S. Ippisch, B. Wicklein, A. Grahl, K. Pörtner, N. Muller, N. Zeitlmann, T. S. Boender, W. Cai, A. Reich, M. an der Heiden, U. Rexroth, O. Hamouda, J. Schneider, T. Veith, B. Mühlemann, R. Wölfel, M. Antwerpen, M. Walter, U. Protzer, B. Liebl, W. Haas, A. Sing, C. Drosten, A. Zapf, Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: A case series. *Lancet Infect. Dis.* **20**, 920–928 (2020).
37. J. F.-W. Chan, S. Yuan, K.-H. Kok, K. K.-W. To, H. Chu, J. Yang, F. Xing, J. Liu, C. C.-Y. Yip, R. W.-S. Poon, H.-W. Tsoi, S. K.-F. Lo, K.-H. Chan, V. K.-M. Poon, W.-M. Chan, J. D. Ip, J.-P. Cai, V. C.-C. Cheng, H. Chen, C. K.-M. Hui, K.-Y. Yuen, A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *Lancet* **395**, 514–523 (2020).
38. L. Zhang, C. B. Jackson, H. Mou, A. Ojha, E. S. Rangarajan, T. Izard, M. Farzan, H. Choe, The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. *bioRxiv* 2020.06.12.148726 , (2020).
39. B. Korber, W. M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, N. Hengartner, E. E. Giorgi, T. Bhattacharya, B. Foley, K. M. Hastie, M. D. Parker, D. G. Partridge, C. M. Evans, T. M. Freeman, T. I. de Silva; Sheffield COVID-19 Genomics Group, C. McDanal, L. G. Perez, H. Tang, A. Moon-Walker, S. P. Whelan, C. C. LaBranche, E. O. Saphire, D. C. Montefiori, Tracking changes in SARS-CoV-2 Spike: Evidence that D614G increases infectivity of the COVID-19 virus. *Cell* **182**, 812–827.e19 (2020).
40. Q. Li, J. Wu, J. Nie, L. Zhang, H. Hao, S. Liu, C. Zhao, Q. Zhang, H. Liu, L. Nie, H. Qin, M. Wang, Q. Lu, X. Li, Q. Sun, J. Liu, L. Zhang, X. Li, W. Huang, Y. Wang, The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell* **182**, 1284–1294.e9 (2020).
41. J. A. Plante, Y. Liu, J. Liu, H. Xia, B. A. Johnson, K. G. Lokugamage, X. Zhang, A. E. Muruato, J. Zou, C. R. Fontes-Garfias, D. Mirchandani, D. Scharton, J. P. Bilello, Z. Ku, Z. An, B. Kalveram, A. N. Freiberg, V. D. Menachery, X. Xie, K. S. Plante, S. C. Weaver, P.-Y. Shi, Spike mutation D614G alters SARS-CoV-2 fitness and neutralization susceptibility. *bioRxiv* 2020.09.01.278689 (2020).
42. S. M. Kissler, C. Tedijanto, E. Goldstein, Y. H. Grad, M. Lipsitch, Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* **368**, 860–868 (2020).
43. K. Itokawa, T. Sekizuka, M. Hashino, R. Tanaka, M. Kuroda, Disentangling primer interactions improves SARS-CoV-2 genome sequencing by multiplex tiling PCR. *PLOS ONE* **15**, e0239403 (2020).
44. S. Andrews, FastQC - A quality control tool for high throughput sequence data, http://bioinformatics.babraham.ac.uk/projects/fastqc/, *Babraham Bioinforma*, http://bioinformatics.babraham.ac.uk/projects/ (2010).
45. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
46. N. D. Grubaugh, K. Gangavarapu, J. Quick, N. L. Matteson, J. G. De Jesus, B. J. Main, A. L. Tan, L. M. Paul, D. E. Brackney, S. Grewal, N. Gurfield, K. K. A. Van Rompay, S. Isern, S. F. Michael, L. L. Coffey, N. J. Loman, K. G. Andersen, An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol.* **20**, 8 (2019).
47. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
48. A. Wilm, P. P. K. Aw, D. Bertrand, G. H. T. Yeo, S. H. Ong, C. H. Wong, C. C. Khor, R. Petric, M. L. Hibberd, N. Nagarajan, LoFreq: A sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res.* **40**, 11189–11201 (2012).
49. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
50. P. Cingolani, A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, D. M. Ruden, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)* **6**, 80–92 (2012).
51. P. Cingolani, V. M. Patel, M. Coon, T. Nguyen, S. J. Land, D. M. Ruden, X. Lu, Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Front. Genet.* **3**, 35 (2012).
52. European Centre for Disease Prevention and Control, *Contact Tracing: Public Health Management of Persons, including Healthcare Workers, who have had Contact with COVID-19 Cases in the European Union* (European Centre for Disease Prevention and Control, Stockholm, 2020).
53. J. Hadfield, C. Megill, S. M. Bell, J. Huddleston, B. Potter, C. Callender, P. Sagulenko, T. Bedford, R. A. Neher, J. Kelso, Nextstrain: Real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).
54. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
55. N. De Maio, C. Walker, R. Borges, L. Weilguny, G. Slodkowicz, N. Goldman, Issues with SARS-CoV-2 sequencing data, in *virological.org* (2020); http://virological.org/t/issues-with-sars-cov-2-sequencing-data/473.
56. P. Sagulenko, V. Puller, R. A. Neher, TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol.* **4**, vex042 (2018).
57. R. Lorenz, S. H. Bernhart, C. Höner zu Siederdissen, H. Tafer, C. Flamm, P. F. Stadler, I. L. Hofacker, ViennaRNA Package 2.0. *Algorithm Mol. Biol.* **6**, 26 (2011).

Acknowledgments: We thank the Biomedical Sequencing Facility at CeMM for assistance with next-generation sequencing. We thank P. Obrist, R. Gattringer, C. Paar, and G. Hörmann for providing samples, and T. Pahlke for support with the computing cluster. We thank the tourism office Paznaun-Ischgl for the statistical data. **Funding:** A.L. and M.S. were supported by a DOC fellowship of the Austrian Academy of Sciences. Z.K. was supported by a fellowship of the Marie Skłodowska-Curie Actions (MSCA) Innovative Training Network H2020-MSCA-ITN-2019 (grant agreement no. 813343). B.A. was supported by the Austrian Science Fund (FWF) PhD program in Inflammation and Immunity (FWF1212). C.B. and A.B. were supported by ERC Starting Grants (European Union's Horizon 2020 research and innovation programme, grant agreement numbers 679146 and 677006, respectively). This project was funded, in part, by the Vienna Science and Technology Fund (WWTF) as part of the WWTF COVID-19 Rapid Response Funding 2020 (to A.B.). **Author contributions:** A.P., J.-W.G., C.B., and A.B. designed the study design and wrote the manuscript. A.P., J.-W.G., M.N., D.S., B.A., A.L., L.E., H.C., M. Smyth, M. Schuster, M.L.G., F.M., O.P., Z.K., M. Senekowitsch, S.M., M.B., M.T.W., G.S.-F., N.L.-B., F.A., F.M., C.B., AB., M.D.N., and F.M.-J. performed the data analysis for this study. T.P., B.A., A.L., M. Senekowitsch, and J.L. designed the assays and processed the experimental samples. S.W.A., W.B., E.P., J.H.A., M.R.-F., M.K., A.Z., P.H., M.N., G.W., D.v.L., and E.P.-S. provided the samples and collected the data. A.B. coordinated the project. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data associated with this study are in the paper or the Supplementary Materials. An online repository of all study-related data, results, and the interactive Nextstrain Austria database is provided on the website http://sarscov2-austria.org. Raw BAM files were submitted for inclusion in the COVID-19 Data Portal hosted by the European Bioinformatics Institute under project number PRJEB39849. Virus sequences (data file S2) are deposited in the GISAID database. All phylogenetic trees used in this study are available for visualization under the following URLs: (i) Global build: https://nextstrain.org/community/bergthalerlab/SARS-CoV-2/NextstrainAustria, with raw data available at https://zenodo.org/record/4247401; (ii) Build with European strains before 31 March: https://nextstrain.org/community/bergthalerlab/SARS-CoV-2/EarlyEurope, raw data available at https://zenodo.org/record/4247401; (iii) Build with Austrian strains used for phylogenetic analysis: https://nextstrain.org/community/bergthalerlab/SARS-CoV-2/OnlyAustrian, with raw data available at https://zenodo.org/record/4247401. Code for sample processing and phylogenetic analyses is available at https://zenodo.org/record/4247401. The time-dynamics frequency of variants in each patient is available at https://zenodo.org/record/4247401. The pairwise comparison of variants between pairs of samples in the transmission lines (Fig. 5A) is available at https://zenodo.org/record/4247401. The code to reproduce the mutational profile and genome-wide mutation rate analysis is available at https://zenodo.org/record/4275398.
This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using this material.

Submitted 12 August 2020
Accepted 16 November 2020
Published First Release 23 November 2020
Published 9 December 2020
10.1126/scitranslmed.abe2555

Citation: A. Popa, J.-W. Genger, M. D. Nicholson, T. Penz, D. Schmid, S. W. Aberle, B. Agerer, A. Lercher, L. Endler, H. Colaço, M. Smyth, M. Schuster, M. L. Grau, F. Martínez-Jiménez, O. Pich, W. Borena, E. Pawelka, Z. Keszei, M. Senekowitsch, J. Laine, J. H. Aberle, M. Redlberger-Fritz, M. Karolyi, A. Zoufaly, S. Maritschnik, M. Borkovec, P. Hufnagl, M. Nairz, G. Weiss, M. T. Wolfinger, D. von Laer, G. Superti-Furga, N. Lopez-Bigas, E. Puchhammer-Stöckl, F. Allerberger, F. Michor, C. Bock, A. Bergthaler, Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. *Sci. Transl. Med.* **12**, eabe2555 (2020).

Downloaded from https://www.science.org on August 02, 2022

Reprinted with permission from AAAS. License number: OP-00145573

2.2.3. Supplementary Material

Science Translational Medicine
AAAS

stm.sciencemag.org/cgi/content/full/scitranslmed.abe2555/DC1

Supplementary Materials for

Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2

Alexandra Popa, Jakob-Wendelin Genger, Michael D. Nicholson, Thomas Penz, Daniela Schmid, Stephan W. Aberle, Benedikt Agerer, Alexander Lercher, Lukas Endler, Henrique Colaço, Mark Smyth, Michael Schuster, Miguel L. Grau, Francisco Martínez-Jiménez, Oriol Pich, Wegene Borena, Erich Pawelka, Zsofia Keszei, Martin Senekowitsch, Jan Laine, Judith H. Aberle, Monika Redlberger-Fritz, Mario Karolyi, Alexander Zoufaly, Sabine Maritschnik, Martin Borkovec, Peter Hufnagl, Manfred Nairz, Günter Weiss, Michael T. Wolfinger, Dorothee von Laer, Giulio Superti-Furga, Nuria Lopez-Bigas, Elisabeth Puchhammer-Stöck, Franz Allerberger, Franziska Michor, Christoph Bock, Andreas Bergthaler*

*Corresponding author. Email: abergthaler@cemm.oeaw.ac.at

Published First Release 23 November 2020, *Sci. Transl. Med.*
DOI: 10.1126/scitranslmed.abe2555

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at stm.sciencemag.org/cgi/content/full/scitranslmed.abe2555/DC1)

Reprinted with permission from AAAS. License number: OP-00145573

Materials and Methods

Technical controls

Two synthetic SARS-CoV-2 RNA genomes consisting of 5 kilobase (Kb) fragments and containing well-characterized fixed mutation differences were used to test the reproducibility and sensitivity of our protocol. Twist-1 (MT007544.1, #102019, Twist Bioscience) contains 7 fixed mutations compared to Twist-2 (MN908947.3, #102024, Twist Bioscience) the reference SARS-CoV-2 genome. Twist-1 was titrated in Twist-2 in increasing ratios (0.1%, 1%, 5%, 10%, 90%, 100%), in duplicates, and subjected to cDNA synthesis and PCR amplification as described (**data file S1**). These controls are important for assessing the limit of low frequency detection across samples.

One sample, CeMM0001 was sequenced in replicates across all our runs to test for possible sequencing biases. In addition, RNA was extracted and processed independently from this sample to serve as a technical control for PCR processing. Amplicons from a second sample, CeMM0008, were sequenced twice in order to assess the potential biases introduced by the sequencing step. To test the impact of different initial viral loads on the variant calling, we have performed a 1:100 dilution experiment of sample CeMM0001. As additional control, we also sequenced two swab-derived samples that were tested negative for SARS-CoV-2 and, as expected, did not obtain any viral reads.

Mutational profiles

Inter-host mutations were reconstructed using the augur pipeline to infer nucleotide changes at the internal nodes *(51)*. Positions reported as highly homoplasic were masked, including the first 55 and the last 100 nucleotides [N. De Maio, C. Walker, R. Borges, L. Weilguny, G. Slodkowicz, and N. Goldman, "Issues with SARS-CoV-2 sequencing data," *virological.org*.]. The consequence type of the mutations was annotated using a customized implementation of the Ensembl Variant Effect Predictor (VEP version 92) using the first SARS-CoV-2 sequenced genome (NCBI ID: NC_045512v2) as a reference. The mutational profile was obtained as the normalized count of the number of mutations in each of the 192 trinucleotide changes. To account for the genomic composition of the SARS-CoV-2 virus we also divided each triplet probability by the total number

Reprinted with permission from AAAS. License number: OP-00145573

of available triples in the SARS-CoV-2 reference genome. For the intra-host analysis, the process to obtain the mutational spectra panels was the same as intra-host but using the low frequency variant calling output (3136 mutations across 420 Austrian samples with alleles frequencies between 0.05 and 0.5). The mutational profile was computed following the same rationale as for the inter-host variants.

Genome-wide mutation rate analysis

We aimed to assess whether the variation in the rate of single nucleotide substitution along the SARS-CoV-2 genome can be solely explained by its tri-nucleotide composition. We devised a statistical test performing local estimations of the deviation from the expectation of the observed number of mutations with respect to the expected based on the tri-nucleotide composition of a particular region of the genome. We first counted the total number of non-protein affecting mutations (that is, synonymous variants and upstream/downstream gene variants) that has been observed across sequenced viral genomes of infected individuals. The focus on non-protein affecting mutations aims to lessen the potential positive selection bias derived from their effect into the coding parts of the viral genome. We did not consider mutations in masked sites (see filtering of mutations for further information about masked sites). We next assigned to each reference nucleoside a probability of mutation of the three alternatives based on its tri-nucleotide context (5' and 3' nucleosides) and the relative probability of mutation derived from the 7,666 samples from GSAID. Then we performed N (N=106) randomizations of the same number of observed mutations distributing them along the SARS-CoV-2 genome according to their mutational probability. Protein-affecting mutations were not randomized, and masked sites were not available to the randomization. We then divided the 29,903 base-pairs (bp) of the viral genome into 10 windows of 1 Kb (except the last window with 903 bps). Analogously, in the zoom-in analysis, we divided the first and last 1kb window of the viral genome into 10 windows of 100 bp. For each window we estimated the mean and standard deviation number of simulated mutations within the window. Last, for each window we estimated the deviation from the expectation using a log-likelihood test (G-test goodness of fit), where we compared the observed number of mutations in the window versus the mean simulated number.

Reprinted with permission from AAAS. License number: OP-00145573

RNA secondary structure prediction

To address the question whether mutations that have been observed in the Austrian SARS- CoV-2 samples have an influence on the RNA structure of the virus we performed computational predictions at the secondary structure level with the ViennaRNA package *(54)*. We started with characterizing locally stable RNA structures in the SARS-CoV-2 reference genome NC 045512.2 with RNALfold. We required that the underlying sequences were not longer than 150 nt and we targeted thermodynamic stability by selecting only regions whose free energy z score was at least -3 among 1000 dinucleotide shuffled sequences of the same sequence composition. We performed single sequence minimum free energy (MFE) structure predictions for both the reference and the mutation variants. In addition, we assessed for each region the level of structural conservation within a set of phylogenetically related viruses. Here we were particularly interested in finding evidence for covariation in stacked helices. Typical covariation patterns are compensatory mutations, i.e. cases where a mutation in one nucleotide is compensated by a second mutation of its pairing partner, such as a GC base-pair being replaced by an AU pair. Likewise, consistent mutations comprise cases where only one nucleotide is exchanged, thereby maintaining the base-pair, for example GC to GU. We characterized orthologous regions in selected Sarbecovirus species with Infernal *(55)*, produced structural multiple sequence nucleotide alignments with locARNA *(56)*and computed consensus structures with RNAalifold *(57)*. In addition, each block was analyzed for structural conservation by RNAz *(58)*.

This PDF file includes:

Figs. S1 to S5

Data files S1 to S5

Reprinted with permission from AAAS. License number: OP-00145573

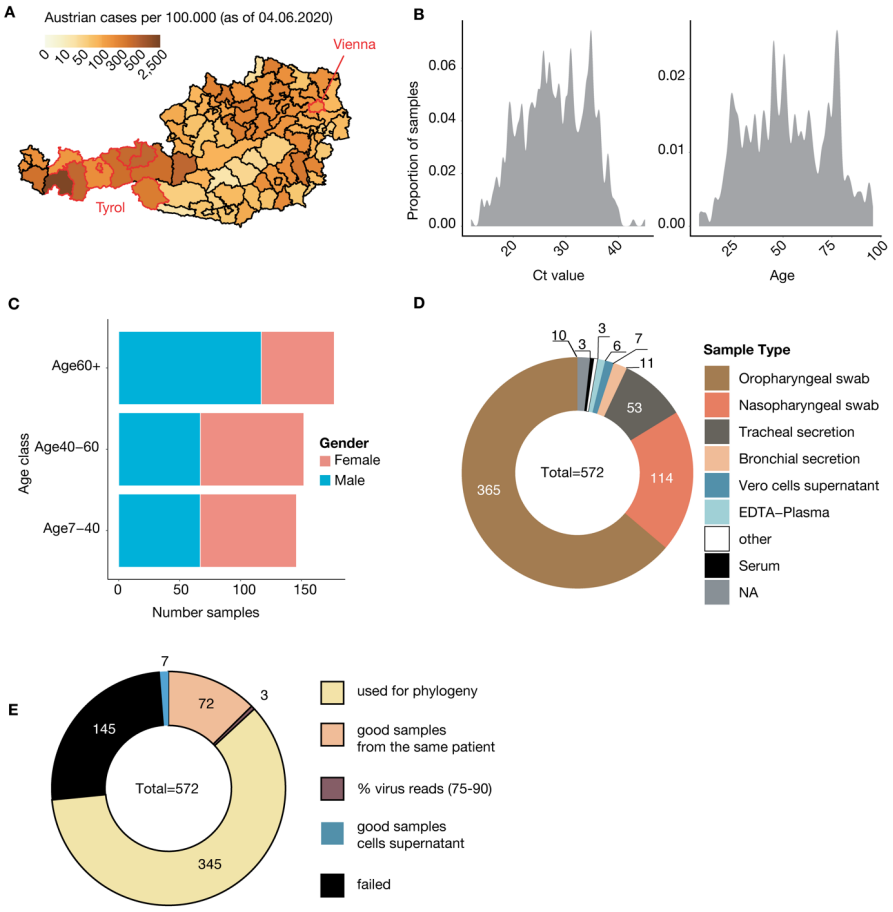
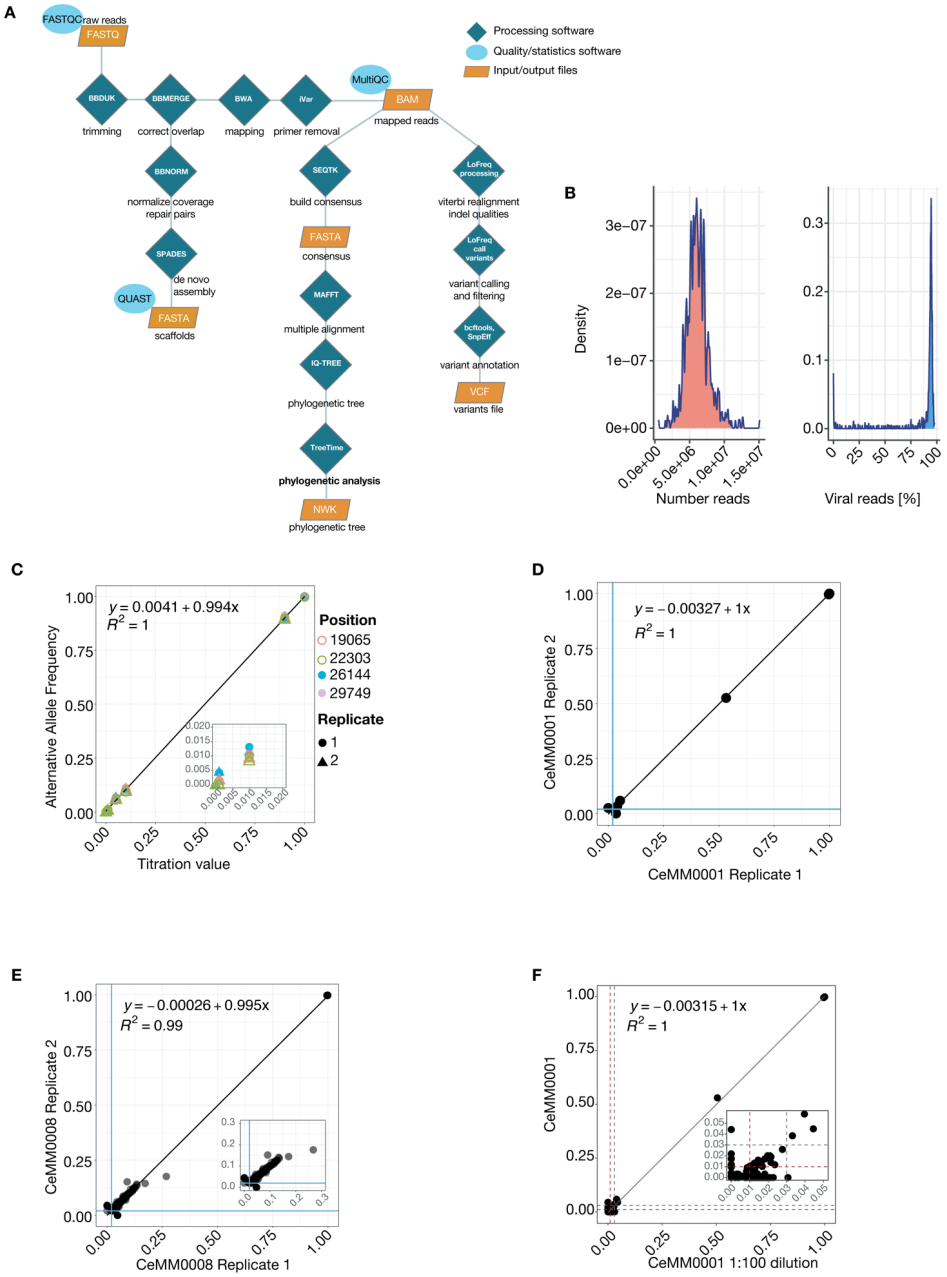


Fig. S1. **Data overview.** (**A**) Proportion of sequenced SARS-CoV-2 samples among positive cases reported across Austrian districts. (**B**) Distribution of available Ct values across the sequencesd samples (n=471 out of 572) and known age distribution of the corresponding SARS-CoV-2 cases (n=476 out of 572). (**C**) Number of female and male patient samples as a function of age class. (**D**) Number of samples for each sampling type category. **(E)** Number of samples based on sequencing quality control, patient information, and use in this study.

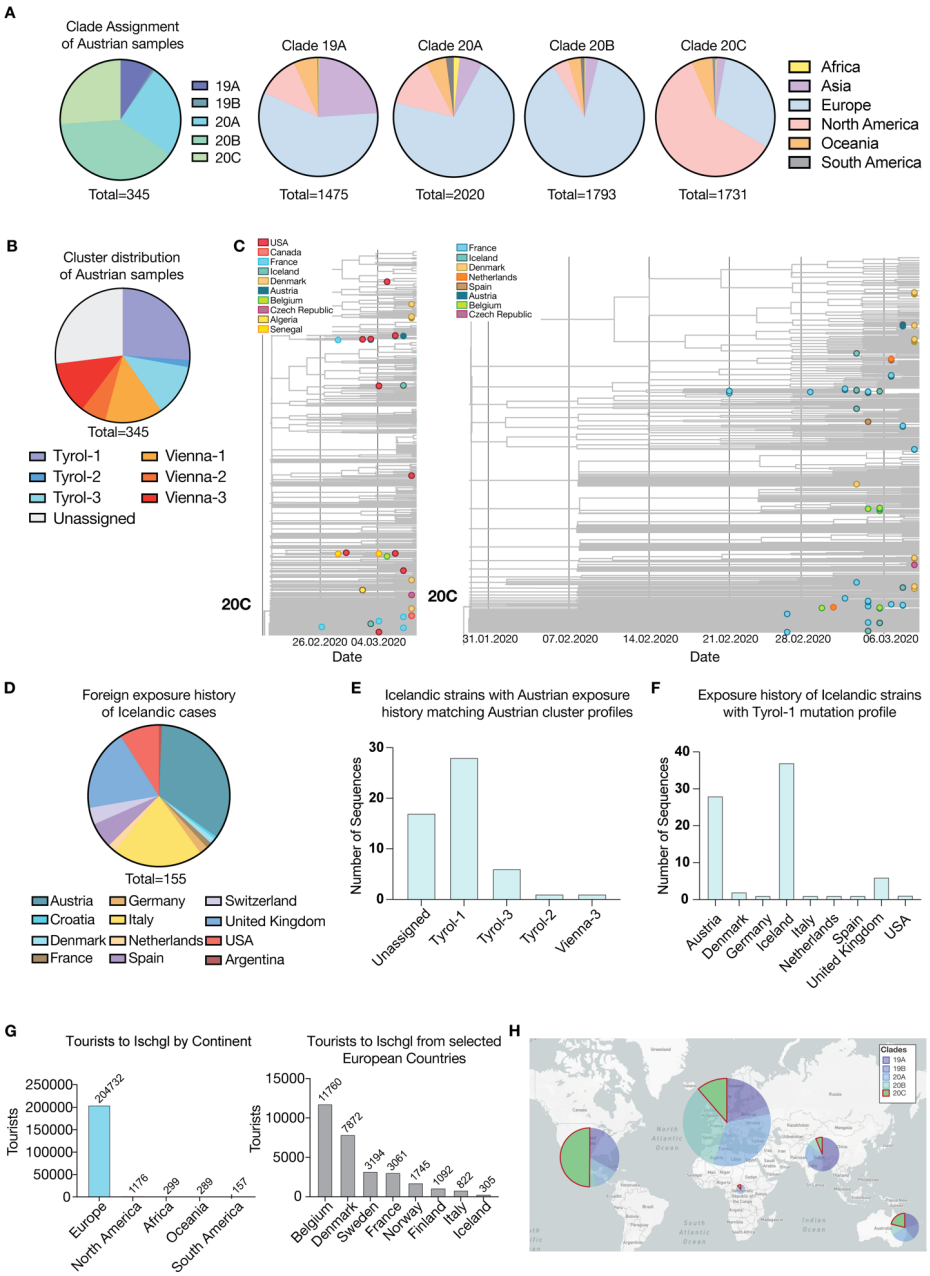
Reprinted with permission from AAAS. License number: OP-00145573



Reprinted with permission from AAAS. License number: OP-00145573

Fig. S2. Technical pipeline and controls. (**A**) Processing pipeline from raw sequencing reads to fasta genomes, phylogenetic trees and low frequency mutation calling. (**B**) Distribution of the number of reads and the percentage of viral reads for all sequenced samples. (**C**) Mixture of two synthetic viral genomes in increasing ratios (0.1%, 1%, 5%, 10%, 90% and 100%). The two technical replicates of this titration are depicted with different symbols. (**D**) Comparison of variant detection for two independent full processing (PCR amplification, library preparation, sequencing) of the same patient sample, CeMM0001. (**E**) Comparison of variant detection for two independent sequencing runs of the same patient sample CeMM0008. (**F**) Comparison of variant detection for CeMM0001 and a dilution 1:100 of the same sample.

Reprinted with permission from AAAS. License number: OP-00145573

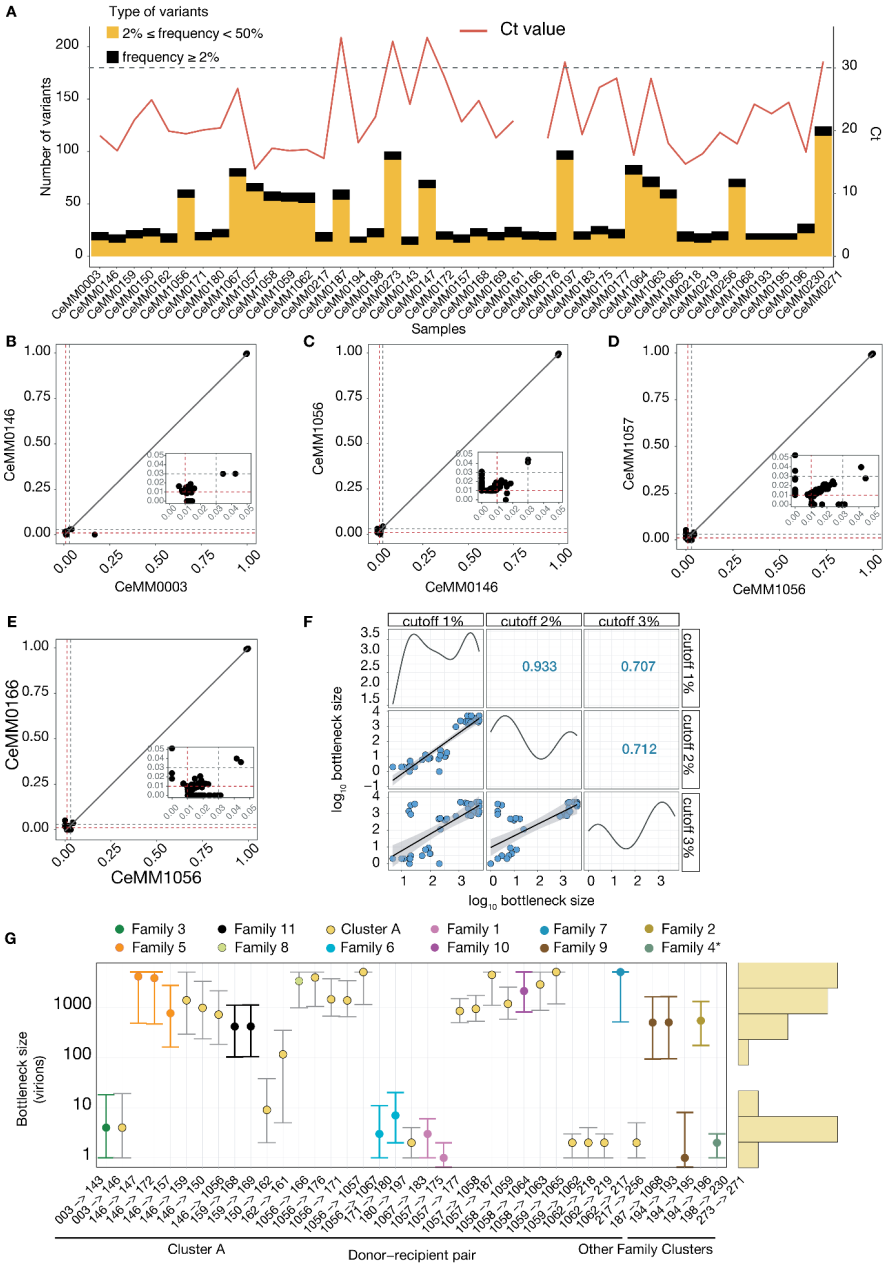


8

Reprinted with permission from AAAS. License number: OP-00145573

Fig. S3. Phylogenetic analysis of SARS-CoV-2 sequences from Austrian COVID-19 patients in global context. (**A**) Nextstrain clade assignment of Austrian samples (left) and geographic distribution of strains in clades defined by Nextstrain (right). The analysis of the geographic distribution of clades bases on information for 8,011 strains in the global phylogenetic analysis in this study. (**B**) Distribution of SARS-CoV-2 from Austrian COVID-19 sequences over the six phylogenetic clusters identified in this publication. (**C**) Clade 20C in time-resolved phylogenetic trees reconstructed from 7,666 randomly subsampled global strains and 345 Austrian strains (left) or all 7,731 European high-quality sequences dated before 31st of March (right). The shown section shows Clade 20C in the time frame from the phylodynamically inferred emergence of clade 20C until 8th of March. (**D**) Statistics of foreign exposure history of Icelandic COVID-19 cases as reported in GISAID. (**E**) Icelandic strains with Austrian exposure history matching Austrian cluster profiles. (**F**) Exposure history of all SARS-CoV-2 sequences from Icelandic COVID-19 cases available on GISAID that match the mutation profile of the phylogenetic cluster Tyrol-1. (**G**) International tourists visiting Ischgl between December 2019 and March 2020 by continent and selected European countries. (**H**) Distribution of SARS-CoV-2 samples over global clades across continents.

Reprinted with permission from AAAS. License number: OP-00145573

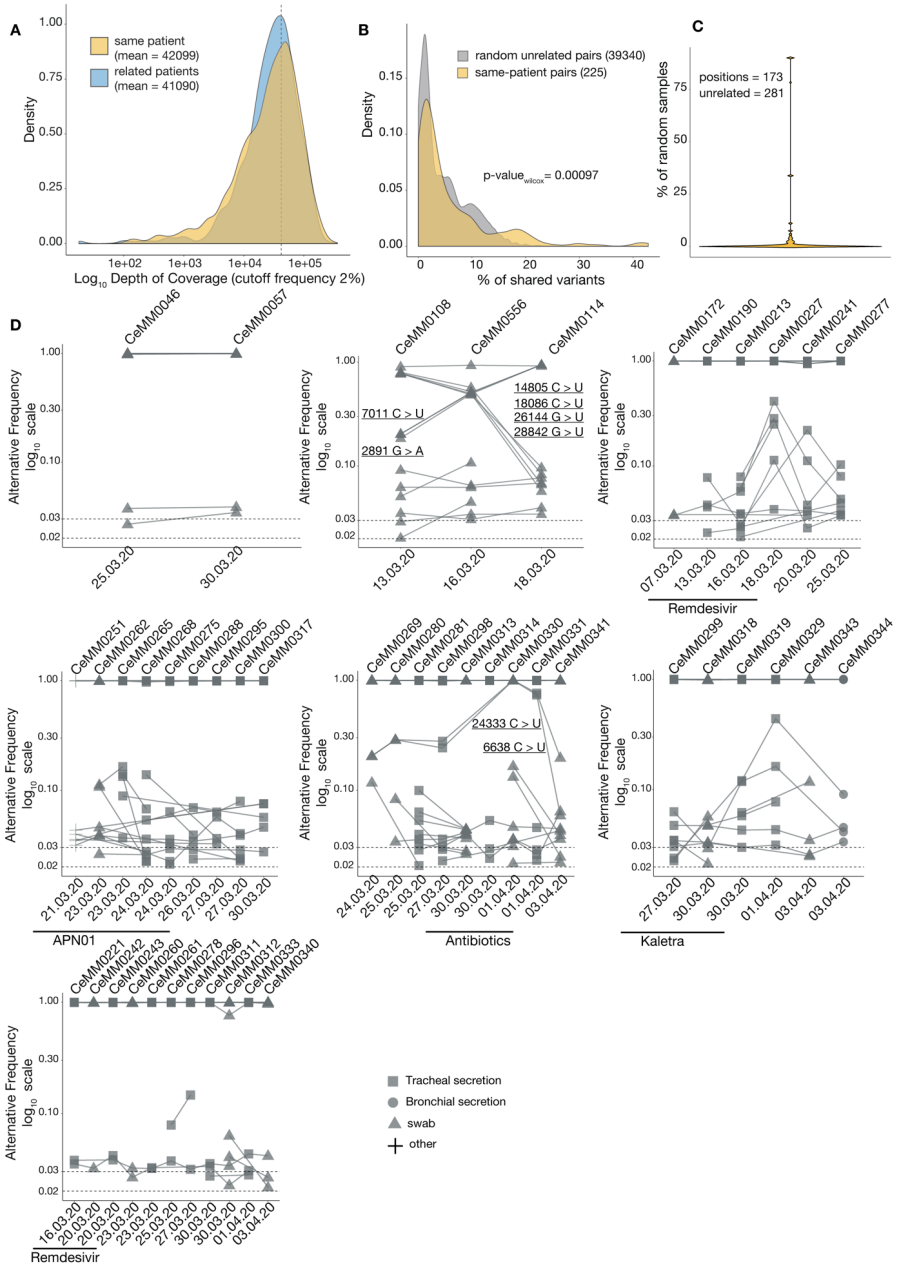


Reprinted with permission from AAAS. License number: OP-00145573

.

Fig. S4: Bottleneck size estimations. (**A**) The Ct values (red line) and number of detected variants per samples for each one of the 43 samples in the transmission chain (**Fig. 5A-B**). All variants detected (cutoff 0.02) are represented in black, with minor variants [0.02, 0.5] represented in yellow. (**B**) Scatter plot of alternative frequency variants between samples CeMM0003 and CeMM0146, leading to a low bottleneck size. (**C**) Scatter plot of alternative frequency variants between samples CeMM0146 and CeMM1056, leading to an intermediary bottleneck size. (**D**) Scatter plot of alternative frequency variants between samples CeMM1056 and CeMM1057, leading to a large bottleneck size. (**E**) Scatter plot of alternative frequency variants between samples CeMM1056 and CeMM0166, leading to a low bottleneck size with a cutoff of 0.01 and a large bottleneck size with a cutoff of 0.03 for the alternative frequency. (**F**) Comparison and correlations of bottleneck size estimations for 3 independent cutoffs: [0.01, 0.95], [0.02, 0.95], and [0.03, 0.95]. (**G**) Bottleneck size (number of virions that initiate the infection in a recipient) estimation across donor-recipient pairs based on **Fig. 5A** and ordered according to the timeline of cluster A for the respective pairs, with a cutoff of [0.03, 0.95] for alternative allele frequency. For patients with multiple samples, the earliest sample was considered for bottleneck size inference. Centered dots are maximum likelihood estimates, with 95% confidence intervals. A star (*) for family 4 indicates that the infection order was inferred as detailed in the Methods. The histogram (yellow bars) of all the bottleneck values is provided on the right side of the graph.

Reprinted with permission from AAAS. License number: OP-00145573



Reprinted with permission from AAAS. License number: OP-00145573

Fig. S5: Viral intra-host diversity in individual patients. (**A**) The depth of coverage distribution for the total number of variants (≥ 0.02) detected in each pair of samples. Two categories of pairs were considered: related patients (pairs in **Fig. 5A-B, data file S4**) (blue) and intra-patient samples (**data file S5**) (yellow). (**B**) Density plots of the percentage of shared variants ([0.02-0.5)) between intra-patient samples (225 pairs; yellow) and unrelated cases (39340 pairs for 281 samples; grey). The p-value of the Wilcoxon test is reported. (**C**) Violin plot of the percentage of unrelated samples (281) detecting each of the 173 variants shared between intra-patient samples ([0.02-0.5)). (**D**) Alternative allele frequency (y axis) of intra-patient variants across time points (x axis). The different sample types are highlighted with different symbols. Clinical treatment information is provided when available. The variants changing frequencies from fixed to minor and vice-versa are reported on the respective graphs. Additional information is provided in **data file S5**.

Supplementary data files (Microsoft Excel format):

Data file S1: Sample and sequencing information of the 572 samples and the controls.

Data file S2: Acknowledgements for SARS-CoV-2 genome sequences derived from GISAID

Data file S3: Epidemiological clusters referred to in this study.

Data file S4: Transmission chain and sample information for ClusterA/ClusterAL and family-related cases.

Data file S5: Clinical information of patients with COVID-19 relating to Figure 5 and Figure S5.

Reprinted with permission from AAAS. License number: OP-00145573

3. Discussion

The emergence of SARS-CoV-2 initiated an unprecedented global research response across scientific disciplines to tackle this emerging infectious disease (Haleem *et al*, 2020). No records existed about this virus before the first clinical reports and the genome sequence "Wuhan-Hu-1" were published. Main objectives of the biomedical research community were to investigate SARS-CoV-2 as the etiological agent of COVID-19, to develop molecular diagnostics for pathogen surveillance, to define a clinical characterization of COVID-19 with therapy guidelines to reduce morbidity and mortality, and to design public health measures based on epidemiological knowledge to contain the disease (World Health Organization, 2020d).

Their high potential for adaptation to selection pressures allows RNA viruses to quickly adapt to new hosts, change clinical and transmission characteristics, subvert immune responses, vaccines, and drug regimens, and escape detectability by molecular diagnostic tests (Villa *et al*, 2021). Several studies have shown a lower sensitivity of new variants of SARS-CoV-2 for neutralizing antibodies resulting from previous infection or vaccination (Li *et al*, 2020; Gupta, 2021; Planas *et al*, 2021; DeGrace *et al*, 2022; Carabelli *et al*, 2023). This highlights the risk of new variants for pandemic response endeavors like vaccination campaigns. Therefore, surveillance of the mutational trajectory during viral spread through the population served as a cornerstone in the pandemic research response and to address the objectives defined above (The Lancet, 2021; Burki, 2021; Carabelli *et al*, 2023).

The work presented in this thesis started during the first wave of SARS-CoV-2 in Austria in March 2020 shortly after the first European cases were identified (Rothe *et al*, 2020; Kreidl *et al*, 2020; Böhmer *et al*, 2020). Academic institutions and organizations with genome sequencing capacities around the world initiated similar pathogen surveillance programs and contributed viral genome data to the GISAID database (Gonzalez-Reiche *et al*, 2020; Deng *et al*, 2020; Gudbjartsson *et al*, 2020; Rothe *et al*, 2020; Pfefferle *et al*, 2021; Böhmer *et al*, 2020; The Lancet, 2021; Burki, 2021; DeGrace *et al*, 2022). The project presented in this thesis took advantage of the centralized Austrian healthcare system and combined national-scale epidemiological data from contact tracing with deep viral genome sequencing to reconstruct transmission chains. Our study connects mutational dynamics of the virus in individuals with international viral spread to shine light on the emergence of new virus variants. We traced the establishment of low-frequency variants to fixed mutations in transmission pairs of local infection clusters and their subsequent international spread across countries. We further

investigated the transmission properties of SARS-CoV-2 using well-curated transmission pairs and determined the genetic bottleneck size for human-to-human transmission. Finally, the comparative analysis of intrahost and interhost viral mutational dynamics presented in this study provides data for t

3.1. Genomic epidemiology is a powerful tool for pathogen surveillance

Several independent introduction events brought SARS-CoV-2 to Austria and caused distinct outbreaks in the beginning of 2020 (Kreidl *et al*, 2020). Virus genome sequencing of patient samples was available broadly and more affordable than during previous outbreaks of SARS-CoV and MERS-CoV, which was a big advantage for modern pathogen surveillance during the SARS-CoV-2 outbreak. We used deep virus genome sequencing of patient samples to determine the spread and distribution of the pathogen and, we could thereby reconstruct epidemiologically defined infection clusters. We collected 572 RNA samples from 449 unique cases from the start to the peak of the first wave of SARS-CoV-2 in Austria. This represented the most comprehensive collection of SARS-CoV-2 genome sequences from Austria (see Figure 9). The data collection contained complete transmission chains of defined infection clusters as well as superspreading events that played a pivotal role for the spread of SARS-CoV-2 across Europe and other countries (Gonzalez-Reiche *et al*, 2020; Gudbjartsson *et al*, 2020; Alm *et al*, 2020).

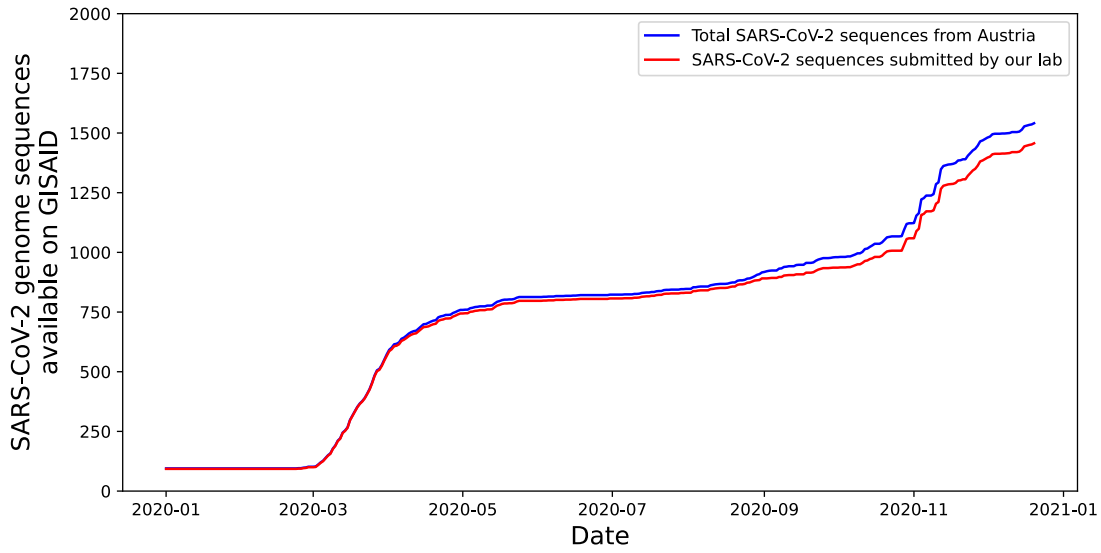


Figure 9: Cumulative number of SARS-CoV-2 genome sequences from Austrian samples available on GISAID. In 2020, the majority of complete, high-coverage SARS-CoV-2 genome sequences from Austria available on GISAID was provided by our lab. Data obtained from GISAID (Shu & McCauley, 2017).

3.1.1. Genomic epidemiology improves definition of transmission chains

Broad diagnostic testing in the population, thorough epidemiological contact tracing, and central collection of epidemiological data in Austria allowed to dissect and distinguish infection clusters (Kreidl *et al*, 2020; Leber *et al*, 2021). As part of our study, we demonstrated the practical utility of genomic epidemiology to confirm and refine results from contact tracing. We compared the six phylogenetic clusters that we identified based on unique mutational profiles to the epidemiological clusters that were defined based on contact tracing. Using phylogenetic analyses, we revealed links between epidemiological clusters in the phylogenetic Vienna-1 cluster which remained elusive to conventional contact tracing epidemiology methods. These links were confirmed by collaborators at the Agentur für Gesundheit und Ernährungssicherheit (AGES) and the Center of Virology of the Medical University of Vienna through individual interviews with the members of the suspected transmission chain connecting both clusters. This approach allowed refinement of infection clusters defined by contact tracing.

The lack of international institutional collaboration on the level of epidemiological contact tracing does usually not allow to connect infection clusters across borders. We leveraged genomic epidemiology to bridge this gap in cross-border epidemiological contact tracing and thereby embedded Austrian SARS-CoV-2 outbreaks into the global transmission network of SARS-CoV-2. This way, we showed that infection cases in Vienna and a ski resort in Tyrol were linked to international travel. We could provide evidence that cases in Iceland were linked to an outbreak in the Austrian ski resort Ischgl and that mutation profiles characteristic for this infection cluster were later detectable in clusters across Europe and North America (Popa *et al*, 2020; Gudbjartsson *et al*, 2020).

By combining data from viral genome sequencing and epidemiological contact tracing, we were able to extend the impact of data obtained by national epidemiological surveillance programs to an international level. With the approach presented in this study, we connected clusters in different countries and could create a phylodynamic model of the global spread of SARS-CoV-2 with focus on transmission chains to and from Austria. Thereby, we demonstrated the value of genomic epidemiology in addition to classic epidemiological monitoring for a more comprehensive pathogen surveillance. Moreover, we showed how results from such sequence-based approaches provide evidence for the assessment and design of public health measures. Our study joins the ranks of several similar genomic epidemiology studies in other countries, that traced the spread and distribution of SARS-CoV-2 across countries and continents through phylodynamic analyses. Despite the lack of standardized protocols and coordinated international efforts, these projects were a great

success during the pandemic because they gave insights into the global spread and mutational trajectory of SARS-CoV-2. Moreover, they allowed monitoring of the emergence of new variants that might interfere with the aims of the public health response (Burki, 2021; Baker *et al*, 2022; Carabelli *et al*, 2023). The early phase of the pandemic presented a window of opportunity to determine the mutational dynamics and transmission characteristics of SARS-CoV-2 due to relatively low infection numbers during this period, which allowed for efficient contact tracing to be implemented.

3.1.2. Genomic epidemiology allows tracing of the emergence of new mutations and virus variants with new characteristics

The S protein D614G mutation was identified through genomic epidemiology, and it allowed tracing of one of the first introduction events to Europe (Böhmer *et al*, 2020; Rothe *et al*, 2020; Carabelli *et al*, 2023). Subsequently, viruses with the D614G mutation established as dominant strains all over Europe and this mutation indicated the European origin of some North American introduction events (Gonzalez-Reiche *et al*, 2020; Fauver *et al*, 2020; Zeller *et al*, 2021). Our retrospective analysis of the mutation profiles in the Tyrol clusters gave heterogeneous results for the appearance of the D614G mutation suggesting several independent introduction events and the circulation of different strains in Tyrol at that time. However, this example illustrates the value of viral genome sequencing and genomic epidemiology beyond reconstruction of infection clusters. The D614G mutation was rapidly identified through sequencing programs and research groups could quickly conduct a functional assessment of its impact. Several studies then reported that this mutation confers increased transmissibility and viral fitness (Plante *et al*, 2021; Korber *et al*, 2020; Plante *et al*, 2021; Carabelli *et al*, 2023). The D614G mutation was the first in a series of mutations that conferred substantial selection advantage such as increased viral fitness, transmissibility, or ability for immune evasion. Virus genome sequencing sped up the identification of new virus variants so that research groups could quickly start their functional assessments. Therefore, one of the main questions when monitoring the mutational trajectory of SARS-CoV-2 was how new mutations emerge and reach fixation.

In the work presented in this thesis, we were able to trace the emergence of a new mutation from low frequency to fixation in a transmission chain connecting two epidemiological clusters in the phylogenetic cluster Vienna-1. The first case with a synonymous mutation at position 20,457 presented with a frequency of 3.6% of virions carrying that mutation. After the following transmission events, virions from cases showed a wide range of frequencies of 0%, 25%, and 100% for this mutation (see page 55).

We also found a subcluster in the phylogenetic cluster Tyrol-1 that was characterized by a nonsynonymous mutation at position 15,380. Like the transmission chain in Vienna-1, we found this mutation also at frequencies below fixation in cases from the cluster Tyrol-1. However, the emergence of this cluster centered around a bar in a ski resort with a lot of international tourist traffic during high season. It is not clear whether epidemiological contact tracing captured all cases of the transmission chain in the right order. Cases could be missing because they left the ski resort and returned home which would fit the observation that this mutation was appearing in viral genome sequencing data all over Europe at the same time. Therefore, we were not able to reconstruct the emergence of the mutation at position 15,380 as clearly as for the mutation at position 20,457 in the phylogenetic cluster Vienna-1 (see page 55).

These approaches of genomic epidemiology that use mutational profiles to reconstruct transmission chains rely on the availability of high-quality pathogen genome sequences from all over the world. At the time when this project started, there was no international organization that would define best practice protocols with guidelines for sampling, sequencing and analysis and conduct unbiased sampling across the globe, so that the mutational dynamics and global spread of SARS-CoV-2 could be sufficiently captured. However, the contributions of national viral genome sequencing initiatives like the one presented in this thesis showed that genomic epidemiology is a crucial cornerstone of the pandemic research response. Therefore, future pandemic preparedness should include the logistic and scientific infrastructure to enable broad pathogen surveillance based on genome sequencing. Analysis and collection of these data should be centralized in a global epidemiological surveillance program to overcome national borders in pathogen surveillance and make it possible to monitor the spread of pathogens across countries.

3.1.3. Variants of SARS-CoV-2

RNA viruses show the highest genetic instability in nature and follow the concept of viral quasispecies evolution (Domingo *et al*, 2012). In fact, following this concept, RNA viruses are sometimes described to not exist as single genotypes. These “clouds of genotypes” are the result of the genetic instability of RNA viruses and are postulated to increase viral fitness, for example by cooperative interaction between genotypes (Eigen, 1993; Vignuzzi *et al*, 2006; Domingo *et al*, 2012; Andino & Domingo, 2015). Therefore, it was expected that SARS-CoV-2 would show a fast mutational trajectory towards better host adaptation and a very high potential to develop variants that subvert immune responses.

The development of several vaccines started very early on in the pandemic and proceeded with unparalleled speed (Krammer, 2020; Gupta, 2021; Morens *et al*, 2022; Carabelli *et al*, 2023).cOne of the major aims was to develop vaccines that would confer broad immunity against SARS-CoV-2 and establish herd immunity. The development and mass rollout of vaccines that could induce humoral and cell-mediated protection and immune memory similar to SARS-CoV-2 infection was successful. However, later the emergence of new variants raised the concern that herd immunity through mass vaccination could not be achieved if the virus kept developing variants that evaded immune responses. Later virus variants, like “Delta” or “Omicron” with several mutations in the viral S protein, were shown to have reduced sensitivity to neutralizing antibodies induced by the different vaccines (Planas *et al*, 2021; DeGrace *et al*, 2022; Carabelli *et al*, 2023). Moreover, a study based on data generated in the project presented here showed that mutations in epitopes of SARS-CoV-2 found in Austrian cases can also escape CTL responses. This follow-up study provided the first evidence of naturally occurring nonsynonymous mutations that confer the ability to subvert epitope-specific CTL responses. Mutated epitopes showed altered antigenicity and therefore did not lead to a functional activation of epitope-specific CTLs, causing decreased proliferation and effector function of CTLs able to recognize wild-type epitopes of SARS-CoV-2 (Agerer *et al*, 2021).

The presented study and the fact that SARS-CoV-2 rapidly evolved variants that could at least partially evade neutralizing antibodies and cell-mediated immune responses again emphasizes the importance of genomic epidemiology for pathogen surveillance (Baker *et al*, 2022; Carabelli *et al*, 2023).

3.2. The effect of the transmission bottleneck on interhost mutational dynamics of SARS-CoV-2

Interhost transmission bottlenecks and intrahost evolutionary dynamics are the main drivers of viral mutagenesis. In the infection clusters that we investigated for this study, we found that SARS-CoV-2 acquires on average about two new fixed mutations per month. We used the resulting mutation profiles in the viral genomes for genomic epidemiology and found in another study from our lab that these mutations can potentially mediate viral immune escape when they are situated in epitopes (Agerer *et al*, 2021). In the following, we used the virus genome data and epidemiological data obtained in the study presented here to learn more about the transmission properties of the virus and estimate its infectivity. The number of viral particles that need to be transmitted from infector to infectee for a productive infection is an important

parameter for any infectious disease. Controlled human challenge studies with SARS-CoV-2 in young healthy volunteers observed that a median tissue culture infectious dose ($TCID_{50}$) of 10 $TCID_{50}$ is needed to induce infection in 50% of participants (Killingley *et al*, 2022). The population of virions exchanged between both individuals is only a small fraction of the total population of virions in the infected individual and, due to stochastic fluctuation, can thereby have a different composition of low-frequency variants than the total population (Domingo *et al*, 2012; Zwart & Elena, 2015; Dolan *et al*, 2018). This genetic bottleneck correlates inversely with the mutation rate of the virus and can be used to estimate the number of virus particles that need to be transmitted to successfully produce progeny in the infectee (Zwart & Elena, 2015).

We used deep viral genome sequencing to identify low-frequency mutations in samples from SARS-CoV-2 cases and an epidemiologically well-curated set of 39 infector-infectee pairs. Initially, we determined a bottleneck size of around 10^3 viral particles which was in line with other studies that were in preprint at that time (Prentiss *et al*, 2022). We interpreted this bottleneck size larger than one determined for influenza A virus in a previous study to fit well to the lower mutation rates of SARS-CoV-2 compared to influenza A (Sobel Leonard *et al*, 2017). Hence, it was concluded that superspreader events provided conducive conditions for the transmission of a substantial number of virions among individuals, primarily owing to close proximity in shared or enclosed spaces, or through a high frequency of interpersonal interactions. The statistical uncertainty of our results indicated that in some cases the transmission of a much smaller number of virions led to successful infection.

Following our publication, Martin and Koelle reanalyzed the dataset published as part of this thesis and yielded a much lower bottleneck size estimate of one to three virions (Martin & Koelle, 2021). Following their response to our study, we revisited our datasets as well in order to understand the discrepancy in results (Nicholson *et al*, 2021). Thereby, we identified critical factors that had a tremendous impact on the resulting bottleneck size estimates such as the low-frequency variant calling and the robust and reproducible identification of low-frequency variants of biological origin. One major difference in the approach of Martin and Koelle was the choice of a higher minor allele frequency (MAF) cutoff of 6% under which low-frequency variants would be discarded and not considered for the analysis (Martin & Koelle, 2021). With this approach, Martin and Koelle could only use 13 out of the 39 transmission pairs for the bottleneck size estimation because the other pairs did not have any intrahost single-nucleotide variants (iSNVs) with more than 6% frequency (Martin & Koelle, 2021).The remaining pairs yielded a genetic transmission bottleneck size of 1.21 virions on average with less than three

virions starting a new infection in over 99% of cases and, thus, was three orders of magnitude below our initial results.

The results of our reanalysis were closer to what was reported by Martin and Koelle and other studies (Lythgoe *et al*, 2021; Martin & Koelle, 2021). The main difference of our revised approach to the 6% MAF cutoff of Martin and Koelle was the choice of a MAF cutoff and removal of technical artefacts. We applied a 3% MAF cutoff to keep iSNVs below 6% for the analysis and thereby retained a larger number of transmission pairs than Martin and Koelle. Additionally, we masked iSNVs that were found to be "highly shared" among many SARS-CoV-2 samples and possibly biological or technical artefacts. We concluded that the best way to exclude technical or biological artefacts and retain high-confidence iSNVs at that time was the manual curation of low-frequency variants and consideration of other knowledge about the low-frequency variants in question. A more general guideline for low-frequency variant filtering besides this prevalence-guided review of iSNVs was not possible at that time and further research is necessary to refine the identification and selection of high-confidence iSNVs. In our reanalysis, the removal of low-confidence low-frequency variants and adjustment of the low-frequency cutoff yielded a bottleneck size estimate of 1 virion in 27 out of 29 transmission pairs and 8 or 58 virions for the remaining pairs, respectively. In conclusion, we were able to refine our approach to bottleneck size estimation with retaining a high number of infector-infectee pairs and manual curation of low-frequency variants which yielded bottleneck size estimates closer to what was obtained by Martin and Koelle as well as Lythgoe *et al*. (Lythgoe *et al*, 2021; Martin & Koelle, 2021)

3.3. Intrahost mutational dynamics of SARS-CoV-2

Intrahost mutational dynamics are another major driver of viral mutagenesis besides interhost transmission bottlenecks. Due to prolonged exposure to the purifying selection pressure from the antigen-specific immune response, intrahost mutational dynamics could play an important role for the development of variants that escape the immune response (Agerer *et al*, 2021; Carabelli *et al*, 2023). Different studies investigated viral evolution of SARS-CoV-2 on the level of low-frequency variants (Jary *et al*, 2020; Kemp *et al*, 2021; Lythgoe *et al*, 2021). One study showed an increased within-host mutational diversity in the early phase of the infection with high viral titers and concluded that intrahost mutational dynamics correlate with viral load (Lythgoe *et al*, 2021).

Several clinical studies reported cases of immunosuppressed patients who were not able to clear the virus within the usual period of two to three weeks, but instead remained chronically infected for extended periods of time (Agarwal *et al*, 2020; Jary *et al*, 2020; Kemp *et al*, 2021; Clark *et al*, 2021; Sonnleitner *et al*, 2022). Some of these patients were immunocompromised due to an ongoing cancer immunotherapy or due to chronic viral infection like HIV (Jary *et al*, 2020; Kemp *et al*, 2021; Maan *et al*, 2022; Sonnleitner *et al*, 2022). Due to the lacking ability to mount immune responses, these patients were usually treated with monoclonal antibodies targeting for example the S protein of SARS-CoV-2 (Kemp *et al*, 2021; Sonnleitner *et al*, 2022). The viral population in these patients was constantly exposed to chemotherapeutic drugs that could increase viral mutagenesis by targeting mechanisms like DNA repair, RNA synthesis, proofreading mechanisms, nucleotide metabolism and others. Investigation of the mutational diversity in these patients led to growing concerns that increased viral mutagenesis without immunological suppression provides favorable conditions for the emergence of new VOCs (Kemp *et al*, 2021). In fact, clinical reports showed that immunocompromised patients treated with monoclonal antibodies against SARS-CoV-2 or convalescent serum showed recurring deletions and mutations that decreased the neutralizing effect of the antibodies (Kemp *et al*, 2021; Sonnleitner *et al*, 2022).

Our collection of RNA samples from SARS-CoV-2-positive cases contained 31 patients with at least two samples from longitudinal samplings. These cases were hospitalized patients that presented with different co-morbidities and received different treatment regimens. We were able to identify a large pool of high-confidence low-frequency variants at a MAF cutoff over 2% that could be repeatedly detected in samples from the same patient. We observed that mutational profiles of low-frequency variants were more similar between samples from the same patient compared to unrelated samples from different patients. The collection of patients investigated and presented in our study shows the emergence and presence of low-frequency variants during infection in line with other studies. However, we could not link the occurrence of specific mutations to certain drug treatments of patients.

We also observed that the type of sampling strategy (nasal swabs, serum, etc.) affected the detectability of low-frequency variants reflecting that different sampling types reflect the composition of virus populations in different body compartments. Following the concept of viral quasispecies evolution, future studies could follow the occurrence of low-frequency variants of SARS-CoV-2 in different body compartments. Comparative analyses of the mutational patterns of the "cloud of variants" in different body sites, such as the small intestine, upper and lower respiratory tracts, may reveal underlying mechanisms of viral adaptation and pathogenesis. Additionally, these analyses may offer insights into other emerging areas of

research, such as wastewater-based pathogen surveillance. Furthermore, longitudinal samples from immunocompetent but also immunocompromised patients can be valuable to study the emergence and accumulation of T cell and antibody escape mutations. Currently, research efforts are ongoing to determine the impact of chemotherapeutic and antiviral drugs, which are thought to potentially enhance mutagenesis, on the intrahost mutational diversity of SARS-CoV-2.

3.4. Science communication is a decisive component of an effective public health response

The COVID-19 pandemic had a tremendous impact on all aspects of society and economy. Success and effectiveness of countermeasures for the containment of virus spread relied strongly on public acceptance and compliance (Nan *et al*, 2022). Therefore, science communication was an important factor contributing to success of public health measures and modalities of proper communication of scientific results to the public were extensively discussed in the scientific community (Antiochou, 2021; de las Heras-Pedrosa *et al*, 2022; Nan *et al*, 2022). For more than two years, society revolved around the COVID-19 pandemic and, thus, it became the focus of misinformation campaigns targeting several aspects of the public health response like vaccination programs, social distancing, and mask mandates (Antiochou, 2021; Nan *et al*, 2022). During the pandemic, many scientists engaged in science communication via social media or the press to build trust in the pandemic response and public health countermeasures and to satisfy the general need for information about the progress of the pandemic.

We collaborated with the aforementioned efforts and utilized the Nextstrain package to conduct phylodynamic analysis and visualize the spread of the pathogen. Our findings, along with comprehensive explanations, were published on our project website (https://www.sarscov2-austria.org/cemm/de/nextstrain-austria_at/), which served as a medium to disseminate our progress in this project, as presented in this thesis, and scientific results from our genomic epidemiology analyses to the scientific community and the general public (Hadfield *et al*, 2018). A variety of studies addressed the difficulties of science communication and contributed to defining a set of best practices (de las Heras-Pedrosa *et al*, 2022; Antiochou, 2021; Nan *et al*, 2022). Overall, these studies acknowledge the positive impact of science communication to counter misinformation and science skepticism with transparency and that it improves the general scientific literacy in the population (Antiochou, 2021; de las Heras-Pedrosa *et al*, 2022; Nan *et al*, 2022; Ward & Rawle, 2022). Therefore, future pandemic preparedness programs should consider teaching and enabling scientists to

communicate their research and engage with the public, as this was proven to be a powerful tool to support public health measures (Antiochou, 2021; de las Heras-Pedrosa *et al*, 2022; Nan *et al*, 2022; Ward & Rawle, 2022).

3.5. Conclusion and outlook

The work presented in this thesis demonstrates the power of genomic epidemiology for pathogen surveillance. It comprises a thorough analysis of fixed mutations and low-frequency variants in samples from Austrian infection clusters and superspreader events with importance for the European spread of SARS-CoV-2 during the first infection wave in early 2020. We reconstructed transmission chains in infection clusters and found connections between previously unrelated clusters. These well-curated infection clusters served for tracing new mutations from emergence to fixation by combining data from virus sequencing and epidemiological contact tracing. Next, we used confirmed infector-infectee pairs to estimate the transmission bottleneck size and refined selection of high-confidence low-frequency variants and MAF cutoffs to obtain better results. Finally, we presented low-frequency mutational dynamics of SARS-CoV-2 in longitudinal measurements of patients.

The scientific and logistic infrastructure for genomic epidemiology projects like this was still underdeveloped at the beginning of the pandemic. Virus genome sequencing projects were initiated around the world, but international standards for unbiased sample acquisition, sequencing, analysis, and centralized collection of the results were lacking. Studies like the one presented in this thesis served as pioneers and proved the value of population-wide genomic epidemiology for pathogen surveillance, especially for identification and monitoring of new virus variant with advanced characteristics like immune escape mechanisms. However, identification and robust quantification of emerging virus variants in the population remains a challenging task if this affords collecting analyzing single samples from infected individuals. Such an approach entails encountering different obstacles. For example, it requires easily accessible diagnostic programs for the broader population to avoid bias in pathogen surveillance, which may occur if only hospital-based samples are collected. However, such programs may incur considerable financial costs. Therefore, several studies presented methods for pathogen surveillance via pathogen genome sequencing in wastewater as a viable tool to obtain a population-wide overview of circulating pathogens. This also allowed the unbiased identification and quantification of emerging virus variants of SARS-CoV-2 (Rothman *et al*, 2021; Amman *et al*, 2022; Wilhelm *et al*, 2022).

This thesis started with an overview of emerging infectious diseases throughout human history with the aim to provide an understanding that the emergence and re-emergence of infectious diseases is a stochastic event influenced by a variety of factors (Morens & Fauci, 2020; van Doorn, 2021; Baker *et al*, 2022). Despite the development of a plethora of medical treatment options, drugs and vaccinations, infectious diseases are still a leading cause of morbidity and mortality in the world (Fauci & Morens, 2012). In 2007, approximately 25.5% of global deaths (in absolute numbers about 15 million deaths) were attributed to infectious diseases – roughly a third of these (4.3 million deaths) were due to respiratory infections (Fauci & Morens, 2012; Morens *et al*, 2008). Moreover, with continued research interest, more diseases can be reclassified as being caused by infectious diseases - for example cervical cancer resulting from human papillomavirus (HPV) infection (Morens *et al*, 2008). Therefore, infectious diseases remain a recurrent challenge for humanity due to their adaptability which leads frequently to the emergence of new pathogens or re-emergence of variants of known pathogens with epidemic or pandemic potential. Since its emergence more than 3 years ago, COVID-19 led to the death of approximately 6.8 million people worldwide (Dong *et al*, 2020: COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University). Despite all efforts with social distancing and quarantine measures, mask mandates and mass deployment of vaccines, SARS-CoV-2 could not be eradicated but is about to become endemic in the human population.

Respiratory infections like influenza were considerably reduced since 2020 due to public health countermeasures to COVID-19 like mask mandates and social distancing (Paget *et al*, 2022). These countermeasures did not just affect the spread of SARS-CoV-2 but it is now discussed whether this led to the extinction of some virus strains like Influenza B/Yamagata (Paget *et al*, 2022; Baker *et al*, 2022). On the other hand, some influenza strains resurged and are currently under investigation regarding their pathogenicity and global risk potential (Sominina *et al*, 2022; Zhang *et al*, 2022; Wille & Barr, 2022). Among these are avian influenza strains like H5N6 with 33 cases between 2021 and 2022 that showed an alarming case fatality rate of 33% (Wille & Barr, 2022; Zhang *et al*, 2022). Other H5 influenza viruses like H5N1 and H5N8 were also found, however with lower case numbers. These highly pathogenic avian influenza viruses are currently under investigation and caused the concern of resurgence of pathogenic influenza strains with pandemic potential (Wille & Barr, 2022).

To date, only one virus was successfully eradicated – the global vaccination program against smallpox led to its extinction in the 1980s (Gessain *et al*, 2022; Kmiec & Kirchhoff, 2022). However, the vaccination campaign was stopped and today 70% of the world population are not immunized. On May 6th, 2022 the first case of a new Mpox outbreak outside of the African

continent was confirmed in the United Kingdom. Mpox is a double-stranded DNA viruses from the genus of orthopoxviruses and thereby a close relative to smallpox. The smallpox vaccination provides protection against Mpox virus to some degree (Gessain *et al*, 2022; Kmiec & Kirchhoff, 2022). Since its re-emergence in Europe in May 2022, Mpox spread across the globe primarily transmitted via skin contact with more than 70,000 cases by October 2022. This sudden increase in case numbers led to the initiation of a coordinated international public health intervention which led to mass vaccine deployment to risk groups (Gessain *et al*, 2022).

These were two examples of infectious diseases that re-emerged within the last two years during the ongoing pandemic of SARS-CoV-2 and therefore serve as contemporary examples that new pathogens can always emerge and re-emerge. These examples illustrate the importance of modern pathogen surveillance tools for monitoring and assessing potential threats from newly emerging or re-emerging pathogens.

The swift identification of SARS-CoV-2 as the cause of COVID-19 serves as a prime example of rapid action during the emergence of a new pathogen with epidemic or pandemic potential. Early sequencing and access to the full virus genome facilitated the quick development of diagnostic tests, forming the basis for the surveillance of disease spreading throughout the pandemic. Additionally, national and international sequencing initiatives allowed to conduct genomic epidemiology to extend the reach of traditional epidemiological contact tracing across borders and continents. These sequencing programs were later also pivotal in pathogen surveillance, detecting emergent variants with potentially altered characteristics like changed morbidity, transmissibility or the ability to evade acquired immunity from previous infection or vaccination. This study highlights the benefits of virus genome sequencing in individual and population-wide samples, emphasizing the importance of a robust scientific and logistical infrastructure. Therefore, contemporary pathogen surveillance tools, including virus genome sequencing from infected individuals and wastewater monitoring, should be expanded to enhance pandemic preparedness and facilitate the timely detection of potential hazards for the population.

4. Materials and Methods

The materials and methods used to obtain and analyze data presented in this thesis are described in the “Materials and Methods” section of the published research article “Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2“. See also pages 57 – 59 and pages 63 - 65 in section 2.2 of this thesis.

References

Agarwal V, Venkatakrishnan AJ, Puranik A, Kirkup C, Lopez-Marquez A, Challener DW, Theel ES, O'Horo JC, Binnicker MJ, Kremers WK, *et al* (2020) Long-term SARS-CoV-2 RNA shedding and its temporal association to IgG seropositivity. *Cell Death Discov* 6: 138

Agerer B, Koblischke M, Gudipati V, Montaño-Gutierrez LF, Smyth M, Popa A, Genger J-W, Endler L, Florian DM, Mühlgrabner V, *et al* (2021) SARS-CoV-2 mutations in MHC-I-restricted epitopes evade CD8 $^{+}$ T cell responses. *Sci Immunol* 6: eabg6461

Akira S, Uematsu S & Takeuchi O (2006) Pathogen Recognition and Innate Immunity. *Cell* 124: 783–801

Alcami A & Koszinowski UH (2000) Viral mechanisms of immune evasion. *Trends in Microbiology* 8: 410–418

Alfaraj SH, Al-Tawfiq JA, Assiri AY, Alzahrani NA, Alanazi AA & Memish ZA (2019) Clinical predictors of mortality of Middle East Respiratory Syndrome Coronavirus (MERS-CoV) infection: A cohort study. *Travel Medicine and Infectious Disease* 29: 48–50

Alm E, Broberg EK, Connor T, Hodcroft EB, Komissarov AB, Maurer-Stroh S, Melidou A, Neher RA, O'Toole Á, Pereyaslov D, *et al* (2020) Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. *Eurosurveillance* 25

Almeida JD & Tyrrell DAJ (1967) The Morphology of Three Previously Uncharacterized Human Respiratory Viruses that Grow in Organ Culture. *Journal of General Virology* 1: 175–178

Al-Tawfiq JA, Momattin H, Dib J & Memish ZA (2014) Ribavirin and interferon therapy in patients infected with the Middle East respiratory syndrome coronavirus: an observational study. *International Journal of Infectious Diseases* 20: 42–46

Amman F, Markt R, Endler L, Hupfauf S, Agerer B, Schedl A, Richter L, Zechmeister M, Bicher M, Heiler G, *et al* (2022) Viral variant-resolved wastewater surveillance of SARS-CoV-2 at national scale. *Nat Biotechnol* 40: 1814–1822

Anderson RM, Fraser C, Ghani AC, Donnelly CA, Riley S, Ferguson NM, Leung GM, Lam TH & Hedley AJ (2004) Epidemiology, transmission dynamics and control of SARS: the 2002–2003 epidemic. *Phil Trans R Soc Lond B* 359: 1091–1105

Andino R & Domingo E (2015) Viral quasispecies. *Virology* 479–480: 46–51

von Andrian UH & Mempel TR (2003) Homing and cellular traffic in lymph nodes. *Nat Rev Immunol* 3: 867–878

Antiochou K (2021) Science communication: challenges and dilemmas in the age of COVID-19. *HPLS* 43: 87

Araujo SBL, Braga MP, Brooks DR, Agosta SJ, Hoberg EP, von Hartenthal FW & Boeger WA (2015) Understanding Host-Switching by Ecological Fitting. *PLoS ONE* 10: e0139225

Arbel R, Hammerman A, Sergienko R, Friger M, Peretz A, Netzer D & Yaron S (2021) BNT162b2 Vaccine Booster and Mortality Due to Covid-19. *N Engl J Med* 385: 2413–2420

Assiri A, Al-Tawfiq JA, Al-Rabeeah AA, Al-Rabiah FA, Al-Hajjar S, Al-Barrak A, Flemban H, Al-Nassir WN, Balkhy HH, Al-Hakeem RF, *et al* (2013) Epidemiological, demographic, and clinical characteristics of 47 cases of Middle East respiratory syndrome coronavirus disease from Saudi Arabia: a descriptive study. *The Lancet Infectious Diseases* 13: 752–761

Azhar EI, El-Kafrawy SA, Farraj SA, Hassan AM, Al-Saeed MS, Hashem AM & Madani TA (2014) Evidence for Camel-to-Human Transmission of MERS Coronavirus. *N Engl J Med* 370: 2499–2505

Baker RE, Mahmud AS, Miller IF, Rajeev M, Rasambainarivo F, Rice BL, Takahashi S, Tatem AJ, Wagner CE, Wang L-F, *et al* (2022) Infectious disease in an era of global change. *Nat Rev Microbiol* 20: 193–205

Barbieri R, Signoli M, Chevé D, Costedoat C, Tzortzis S, Aboudharam G, Raoult D & Drancourt M (2020) Yersinia pestis: the Natural History of Plague. *Clin Microbiol Rev* 34: e00044-19

Bar-On YM, Goldberg Y, Mandel M, Bodenheimer O, Freedman L, Alroy-Preis S, Ash N, Huppert A & Milo R (2021) Protection against Covid-19 by BNT162b2 Booster across Age Groups. *N Engl J Med* 385: 2421–2430

Barré-Sinoussi F, Ross AL & Delfraissy J-F (2013) Past, present and future: 30 years of HIV research. *Nat Rev Microbiol* 11: 877–883

Bastard P, Rosen LB, Zhang Q, Michailidis E, Hoffmann H-H, Zhang Y, Dorgham K, Philippot Q, Rosain J, Béziat V, *et al* (2020) Autoantibodies against type I IFNs in patients with life-threatening COVID-19. *Science* 370: eabd4585

Beach JR & Schalm OW (1936) A Filterable Virus, Distinct from that of Laryngotracheitis, the Cause of a Respiratory Disease of Chicks. *Poultry Science* 15: 199–206

Bernard Stoecklin S, Rolland P, Silue Y, Mailles A, Campese C, Simondon A, Mechain M, Meurice L, Nguyen M, Bassi C, *et al* (2020) First cases of coronavirus disease 2019 (COVID-19) in France: surveillance, investigations and control measures, January 2020. *Eurosurveillance* 25

Binn LN, Lazar EC, Keenan KP, Huxsoll DL, Marchwicki RH & Strano AJ (1974) Recovery and characterization of a coronavirus from military dogs with diarrhea. *Proc Annu Meet U S Anim Health Assoc*: 359–366

Björkström NK, Strunz B & Ljunggren H-G (2022) Natural killer cells in antiviral immunity. *Nat Rev Immunol* 22: 112–123

Bluhm A, Christandl M, Gesmundo F, Ravn Klausen F, Mančinska L, Steffan V, Stilck França D & Werner AH (2020) SARS-CoV-2 transmission routes from genetic data: A Danish case study. *PLoS ONE* 15: e0241405

Boechat JL, Chora I, Morais A & Delgado L (2021) The immune response to SARS-CoV-2 and COVID-19 immunopathology – Current perspectives. *Pulmonology* 27: 423–437

Böhmer MM, Buchholz U, Corman VM, Hoch M, Katz K, Marosevic DV, Böhm S, Woudenberg T, Ackermann N, Konrad R, *et al* (2020) Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: a case series. *The Lancet Infectious Diseases* 20: 920–928

Boni MF, Lemey P, Jiang X, Lam TT-Y, Perry BW, Castoe TA, Rambaut A & Robertson DL (2020) Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat Microbiol* 5: 1408–1417

Bowen DG & Walker CM (2005) Mutational escape from CD8+ T cell immunity. *Journal of Experimental Medicine* 201: 1709–1714

Bridger JC, Caul EO & Egglestone SI (1978) Replication of an enteric bovine coronavirus in intestinal organ cultures. *Archives of Virology* 57: 43–51

Brown M (2020) Immune correlates of SARS-CoV-2 protection. *Nat Rev Immunol* 20: 593–593

Burki T (2021) Understanding variants of SARS-CoV-2. *The Lancet* 397: 462

Carabelli AM, Peacock TP, Thorne LG, Harvey WT, Hughes J, COVID-19 Genomics UK Consortium, de Silva TI, Peacock SJ, Barclay WS, de Silva TI, *et al* (2023) SARS-CoV-2 variant biology: immune escape, transmission and fitness. *Nat Rev Microbiol*

Chan JF-W, Yuan S, Kok K-H, To KK-W, Chu H, Yang J, Xing F, Liu J, Yip CC-Y, Poon RW-S, *et al* (2020) A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet* 395: 514–523

Chazal N (2021) Coronavirus, the King Who Wanted More Than a Crown: From Common to the Highly Pathogenic SARS-CoV-2, Is the Key in the Accessory Genes? *Front Microbiol* 12: 682603

Cheever FS, Daniels JB, Pappenheimer AM & Bailey OT (1949) A MURINE VIRUS (JHM) CAUSING DISSEMINATED ENCEPHALOMYELITIS WITH EXTENSIVE DESTRUCTION OF MYELIN. *Journal of Experimental Medicine* 90: 181–194

Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y, *et al* (2020) Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet* 395: 507–513

Choi WJ, Lee K-N, Kang E-J & Lee H (2016) Middle East Respiratory Syndrome-Coronavirus Infection: A Case Report of Serial Computed Tomographic Findings in a Young Male Patient. *Korean J Radiol* 17: 166

Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ, Chu DK, Akl EA, El-harakeh A, Bognanni A, *et al* (2020) Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *The Lancet* 395: 1973–1987

Chu DKW, Hui KPY, Perera RAPM, Miguel E, Niemeyer D, Zhao J, Channappanavar R, Dudas G, Oladipo JO, Traoré A, *et al* (2018) MERS coronaviruses from camels in Africa exhibit region-dependent genetic diversity. *Proc Natl Acad Sci USA* 115: 3144–3149

Clark SA, Clark LE, Pan J, Coscia A, McKay LGA, Shankar S, Johnson RI, Brusic V, Choudhary MC, Regan J, *et al* (2021) SARS-CoV-2 evolution in an immunocompromised host reveals shared neutralization escape mechanisms. *Cell* 184: 2605-2617.e18

Combe M & Sanjuán R (2014) Variation in RNA Virus Mutation Rates across Host Cells. *PLoS Pathog* 10: e1003855

Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, Haagmans BL, Lauber C, Leontovich AM, *et al* (2020) The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* 5: 536–544

Cui J, Li F & Shi Z-L (2019) Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol* 17: 181–192

Dan JM, Mateus J, Kato Y, Hastie KM, Yu ED, Faliti CE, Grifoni A, Ramirez SI, Haupt S, Frazier A, *et al* (2021) Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science* 371: eabf4063

Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, Pearson CAB, Russell TW, Tully DC, Washburne AD, *et al* (2021) Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* 372: eabg3055

De Cock KM, Jaffe HW & Curran JW (2011) Reflections on 30 Years of AIDS. *Emerg Infect Dis* 17: 1044–1048

DeGrace MM, Ghedin E, Frieman MB, Krammer F, Grifoni A, Alisoltani A, Alter G, Amara RR, Baric RS, Barouch DH, *et al* (2022) Defining the risk of SARS-CoV-2 variants on immune protection. *Nature* 605: 640–652

Deng K, Pertea M, Rongvaux A, Wang L, Durand CM, Ghiaur G, Lai J, McHugh HL, Hao H, Zhang H, *et al* (2015) Broad CTL response is required to clear latent HIV-1 due to dominance of escape mutations. *Nature* 517: 381–385

Deng S-Q & Peng H-J (2020) Characteristics of and Public Health Responses to the Coronavirus Disease 2019 Outbreak in China. *JCM* 9: 575

Deng X, Gu W, Federman S, du Plessis L, Pybus OG, Faria NR, Wang C, Yu G, Bushnell B, Pan C-Y, *et al* (2020) Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California. *Science* 369: 582–587

Diekmann O, Othmer HG, Planqué R & Bootsma MCJ (2021) The discrete-time Kermack–McKendrick model: A versatile and computationally attractive framework for modeling epidemics. *Proc Natl Acad Sci USA* 118: e2106332118

Ding Y, He L, Zhang Q, Huang Z, Che X, Hou J, Wang H, Shen H, Qiu L, Li Z, *et al* (2004) Organ distribution of severe acute respiratory syndrome(SARS) associated coronavirus(SARS-CoV) in SARS patients: implications for pathogenesis and virus transmission pathways. *J Pathol* 203: 622–630

Dobson AP & Carper ER (1996) Infectious Diseases and Human Population History. *BioScience* 46: 115–126

Dolan PT, Whitfield ZJ & Andino R (2018) Mechanisms and Concepts in RNA Virus Population Dynamics and Evolution. *Annu Rev Virol* 5: 69–92

Domingo E & Holland JJ (1997) RNA VIRUS MUTATIONS AND FITNESS FOR SURVIVAL. *Annu Rev Microbiol* 51: 151–178

Domingo E, Sheldon J & Perales C (2012) Viral Quasispecies Evolution. *Microbiol Mol Biol Rev* 76: 159–216

Dong E, Du H & Gardner L (2020) An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases* 20: 533–534

van Doorn HR (2021) The epidemiology of emerging infectious diseases and pandemics. *Medicine* 49: 659–662

Doyle LP & Hutchings LM (1946) A transmissible gastroenteritis in pigs. *J Am Vet Med Assoc* 108: 257–259

Drosten C, Günther S, Preiser W, van der Werf S, Brodt H-R, Becker S, Rabenau H, Panning M, Kolesnikova L, Fouchier RAM, *et al* (2003) Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *N Engl J Med* 348: 1967–1976

Earle KA, Ambrosino DM, Fiore-Gartland A, Goldblatt D, Gilbert PB, Siber GR, Dull P & Plotkin SA (2021) Evidence for antibody as a protective correlate for COVID-19 vaccines. *Vaccine* 39: 4423–4428

Eckerle LD, Lu X, Sperry SM, Choi L & Denison MR (2007) High Fidelity of Murine Hepatitis Virus Replication Is Decreased in nsp14 Exoribonuclease Mutants. *J Virol* 81: 12135–12144

Eigen M (1993) Viral Quasispecies. *Sci Am* 269: 42–49

Eigen M & Schuster P (1977) A principle of natural self-organization: Part A: Emergence of the hypercycle. *Naturwissenschaften* 64: 541–565

Escalera A, Gonzalez-Reiche AS, Aslam S, Mena I, Laporte M, Pearl RL, Fossati A, Rathnasinghe R, Alshammary H, van de Guchte A, *et al* (2022) Mutations in SARS-CoV-2 variants of concern link to increased spike cleavage and virus transmission. *Cell Host & Microbe* 30: 373-387.e7

European Centre for Disease Prevention and Control (2020) Contact tracing: public health management of persons, including healthcare workers, having had contact with COVID-19 cases in the European Union.

Fauci AS & Morens DM (2012) The Perpetual Challenge of Infectious Diseases. *N Engl J Med* 366: 454–461

Fauver JR, Petrone ME, Hodcroft EB, Shioda K, Ehrlich HY, Watts AG, Vogels CBF, Brito AF, Alpert T, Muyombwe A, *et al* (2020) Coast-to-Coast Spread of SARS-CoV-2 during the Early Epidemic in the United States. *Cell* 181: 990-996.e5

Feng S, Phillips DJ, White T, Sayal H, Aley PK, Bibi S, Dold C, Fuskova M, Gilbert SC, Hirsch I, *et al* (2021) Correlates of protection against symptomatic and asymptomatic SARS-CoV-2 infection. *Nat Med* 27: 2032–2040

Frérot M, Lefebvre A, Aho S, Callier P, Astruc K & Aho Glélé LS (2018) What is epidemiology? Changing definitions of epidemiology 1978-2017. *PLoS ONE* 13: e0208442

Funk T, Pharris A, Spiteri G, Bundle N, Melidou A, Carr M, Gonzalez G, Garcia-Leon A, Crispie F, O'Connor L, *et al* (2021) Characteristics of SARS-CoV-2 variants of concern B.1.1.7, B.1.351 or P.1: data from seven EU/EEA countries, weeks 38/2020 to 10/2021. *Eurosurveillance* 26

German Advisory Committee Blood (Arbeitskreis Blut) S 'Assessment of PT by B (2016) Human Immunodeficiency Virus (HIV). *Transfus Med Hemother* 43: 203–222

Gessain A, Nakoune E & Yazdanpanah Y (2022) Monkeypox. *N Engl J Med* 387: 1783–1793

Ghosh S, Dellibovi-Ragheb TA, Kerviel A, Pak E, Qiu Q, Fisher M, Takvorian PM, Bleck C, Hsu VW, Fehr AR, *et al* (2020) β-Coronaviruses Use Lysosomes for Egress Instead of the Biosynthetic Secretory Pathway. *Cell* 183: 1520-1535.e14

Gledhill AW & Andrewes CH (1951) A hepatitis virus of mice. *Br J Exp Pathol* 32: 559–568

Global Initiative on Sharing All Influenza Data (GISAID) (2021) Clade and lineage nomenclature, March 2, 2021 - Clade and lineage nomenclature aids in genomic epidemiology studies of active hCoV-19 viruses.

Gonzalez-Reiche AS, Hernandez MM, Sullivan MJ, Ciferri B, Alshammary H, Obla A, Fabre S, Kleiner G, Polanco J, Khan Z, *et al* (2020) Introductions and early spread of SARS-CoV-2 in the New York City area. *Science* 369: 297–301

Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, Luo SW, Li PH, Zhang LJ, Guan YJ, *et al* (2003) Isolation and Characterization of Viruses Related to the SARS Coronavirus from Animals in Southern China. *Science* 302: 276–278

Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL, Saemundsdottir J, Sigurdsson A, Sulem P, Agustsdottir AB, *et al* (2020) Spread of SARS-CoV-2 in the Icelandic Population. *N Engl J Med* 382: 2302–2315

Gupta RK (2021) Will SARS-CoV-2 variants of concern affect the promise of vaccines? *Nat Rev Immunol* 21: 340–341

Haagmans BL, Al Dhahiry SHS, Reusken CBEM, Raj VS, Galiano M, Myers R, Godeke G-J, Jonges M, Farag E, Diab A, *et al* (2014) Middle East respiratory syndrome coronavirus in dromedary camels: an outbreak investigation. *The Lancet Infectious Diseases* 14: 140–145

Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T & Neher RA (2018) Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 34: 4121–4123

Haleem A, Javaid M, Vaishya R & Deshmukh SG (2020) Areas of academic research with the impact of COVID-19. *The American Journal of Emergency Medicine* 38: 1524–1526

Hamming I, Timens W, Bulthuis M, Lely A, Navis G & van Goor H (2004) Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J Pathol* 203: 631–637

Hamre D & Procknow JJ (1966) A New Virus Isolated from the Human Respiratory Tract. *Experimental Biology and Medicine* 121: 190–193

Han AX, Parker E, Scholer F, Maurer-Stroh S & Russell CA (2019) Phylogenetic Clustering by Linear Integer Programming (PhyCLIP). *Molecular Biology and Evolution* 36: 1580–1595

Harbeck M, Seifert L, Hänsch S, Wagner DM, Birdsell D, Parise KL, Wiechmann I, Grupe G, Thomas A, Keim P, *et al* (2013) Yersinia pestis DNA from Skeletal Remains from the 6th Century AD Reveals Insights into Justinianic Plague. *PLoS Pathog* 9: e1003349

Hayn M, Hirschenberger M, Koepke L, Nchioua R, Straub JH, Klute S, Hunszinger V, Zech F, Prelli Bozzo C, Aftab W, *et al* (2021) Systematic functional analysis of SARS-CoV-2 proteins uncovers viral innate immune antagonists and remaining vulnerabilities. *Cell Reports* 35: 109126

Henderson LA & Yeung RSM (2021) MIS-C: early lessons from immune profiling. *Nat Rev Rheumatol* 17: 75–76

de las Heras-Pedrosa C, Jambrino-Maldonado C, Rando-Cueto D & Iglesias-Sánchez PP (2022) COVID-19 Study on Scientific Articles in Health Communication: A Science Mapping Analysis in Web of Science. *IJERPH* 19: 1705

Hijawi B, Abdallat M, Sayaydeh A, Alqasrawi S, Haddadin A, Jaarour N, Alsheikh S & Alsanouri T (2013) Novel coronavirus infections in Jordan, April 2012: epidemiological findings from a retrospective investigation. *East Mediterr Health J* 19 Suppl 1: S12-18

Hilligan KL & Ronchese F (2020) Antigen presentation by dendritic cells and their instruction of CD4+ T helper cell responses. *Cell Mol Immunol* 17: 587–599

Hodcroft EB, Hadfield J, Neher RA & Bedford T (2021) Year-letter Genetic Clade Naming for SARS-CoV-2 on Nextstrain.org.

van der Hoek L (2007) Human coronaviruses: what do they cause? *Antivir Ther* 12: 651–658

van der Hoek L, Pyrc K, Jebbink MF, Vermeulen-Oost W, Berkhout RJM, Wolthers KC, Wertheim-van Dillen PME, Kaandorp J, Spaargaren J & Berkhout B (2004) Identification of a new human coronavirus. *Nat Med* 10: 368–373

Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu N-H, Nitsche A, *et al* (2020) SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 181: 271-280.e8

Holland J, Spindler K, Horodyski F, Grabau E, Nichol S & VandePol S (1982) Rapid Evolution of RNA Genomes. *Science* 215: 1577–1585

Holmes EC, Goldstein SA, Rasmussen AL, Robertson DL, Crits-Christoph A, Wertheim JO, Anthony SJ, Barclay WS, Boni MF, Doherty PC, *et al* (2021) The origins of SARS-CoV-2: A critical review. *Cell* 184: 4848–4856

Hsu L-Y, Lee C-C, Green JA, Ang B, Paton NI, Lee L, Villacian JS, Lim P-L, Earnest A & Leo Y-S (2003) Severe Acute Respiratory Syndrome (SARS) in Singapore: Clinical Features of Index Patient and Initial Contacts. *Emerg Infect Dis* 9: 713–717

Hu B, Guo H, Zhou P & Shi Z-L (2021) Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol* 19: 141–154

Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, *et al* (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* 395: 497–506

Humphries F, Shmuel-Galia L, Jiang Z, Wilson R, Landis P, Ng S-L, Parsi KM, Maehr R, Cruz J, Morales A, *et al* (2021) A diamidobenzimidazole STING agonist protects against SARS-CoV-2 infection. *Sci Immunol* 6: eabi9002

International Committee on Taxonomy of Viruses (ICTV) (2022) The International Code of Virus Classification and Nomenclature (ICVCN).

Ismail MM, Tang Y & Saif YM (2003) Pathogenicity of Turkey Coronavirus in Turkeys and Chickens. *Avian Diseases* 47: 515–522

Jary A, Leducq V, Malet I, Marot S, Klement-Frutos E, Teyssou E, Soulié C, Abdi B, Wirden M, Pourcher V, *et al* (2020) Evolution of viral quasispecies during SARS-CoV-2 infection. *Clinical Microbiology and Infection* 26: 1560.e1-1560.e4

Joffre OP, Segura E, Savina A & Amigorena S (2012) Cross-presentation by dendritic cells. *Nat Rev Immunol* 12: 557–569

Johnson WE (2019) Origins and evolutionary consequences of ancient endogenous retroviruses. *Nat Rev Microbiol* 17: 355–370

Kaech SM & Cui W (2012) Transcriptional control of effector and memory CD8+ T cell differentiation. *Nature Reviews Immunology* 12: 749–761

Kan B, Wang M, Jing H, Xu H, Jiang X, Yan M, Liang W, Zheng H, Wan K, Liu Q, *et al* (2005) Molecular Evolution Analysis and Geographic Investigation of Severe Acute Respiratory Syndrome Coronavirus-Like Virus in Palm Civets at an Animal Market and on Farms. *J Virol* 79: 11892–11900

Kautz T & Forrester N (2018) RNA Virus Fidelity Mutants: A Useful Tool for Evolutionary Biology or a Complex Challenge? *Viruses* 10: 600

Keeton R, Tincho MB, Ngomti A, Baguma R, Benede N, Suzuki A, Khan K, Cele S, Bernstein M, Karim F, *et al* (2022) T cell responses to SARS-CoV-2 spike cross-recognize Omicron. *Nature* 603: 488–492

Kemp SA, Collier DA, Datir RP, Ferreira IATM, Gayed S, Jahun A, Hosmillo M, Rees-Spear C, Mlcochova P, Lumb IU, *et al* (2021) SARS-CoV-2 evolution during treatment of chronic infection. *Nature* 592: 277–282

Kent SJ, Khoury DS, Reynaldi A, Juno JA, Wheatley AK, Stadler E, John Wherry E, Triccas J, Sasson SC, Cromer D, *et al* (2022) Disentangling the relative importance of T cell responses in COVID-19: leading actors or supporting cast? *Nat Rev Immunol* 22: 387–397

Khoury DS, Cromer D, Reynaldi A, Schlub TE, Wheatley AK, Juno JA, Subbarao K, Kent SJ, Triccas JA & Davenport MP (2021) Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. *Nat Med* 27: 1205–1211

Killingley B, Mann AJ, Kalinova M, Boyers A, Goonawardane N, Zhou J, Lindsell K, Hare SS, Brown J, Frise R, *et al* (2022) Safety, tolerability and viral kinetics during SARS-CoV-2 human challenge in young adults. *Nat Med* 28: 1031–1041

Kim D, Lee J-Y, Yang J-S, Kim JW, Kim VN & Chang H (2020) The Architecture of SARS-CoV-2 Transcriptome. *Cell* 181: 914-921.e10

Kim Y-M & Shin E-C (2021) Type I and III interferon responses in SARS-CoV-2 infection. *Exp Mol Med* 53: 750–760

Kindred B & Shreffler DC (1972) H-2 dependence of co-operation between T and B cells in vivo. *J Immunol* 109: 940–943

Kmiec D & Kirchhoff F (2022) Monkeypox: A New Threat? *IJMS* 23: 7866

Knoops K, Kikkert M, Worm SHE van den, Zevenhoven-Dobbe JC, van der Meer Y, Koster AJ, Mommaas AM & Snijder EJ (2008) SARS-Coronavirus Replication Is Supported by a Reticulovesicular Network of Modified Endoplasmic Reticulum. *PLoS Biol* 6: e226

Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi EE, Bhattacharya T, Foley B, *et al* (2020) Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* 182: 812-827.e19

Korea Centers for Disease Control & Prevention (2015) Middle East Respiratory Syndrome Coronavirus Outbreak in the Republic of Korea, 2015. *Osong Public Health and Research Perspectives* 6: 269–278

Krammer F (2020) SARS-CoV-2 vaccines in development. *Nature* 586: 516–527

Kreidl P, Schmid D, Maritschnik S, Richter L, Borena W, Genger J-W, Popa A, Penz T, Bock C, Bergthaler A, *et al* (2020) Emergence of coronavirus disease 2019 (COVID-19) in Austria. *Wien Klin Wochenschr* 132: 645–652

Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, Tong S, Urbani C, Comer JA, Lim W, *et al* (2003) A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *N Engl J Med* 348: 1953–1966

Lazear HM, Schoggins JW & Diamond MS (2019) Shared and Distinct Functions of Type I and Type III Interferons. *Immunity* 50: 907–923

Leber W, Lammel O, Redlberger-Fritz M, Mustafa-Korninger ME, Glehr RC, Camp J, Agerer B, Lercher A, Popa A, Genger J-W, *et al* (2021) Rapid, early and accurate SARS-CoV-2 detection using RT-qPCR in primary care: a prospective cohort study (REAP-1). *BMJ Open* 11: e045225

Lee N, Hui D, Wu A, Chan P, Cameron P, Joynt GM, Ahuja A, Yung MY, Leung CB, To KF, *et al* (2003) A Major Outbreak of Severe Acute Respiratory Syndrome in Hong Kong. *N Engl J Med* 348: 1986–1994

Letko M, Marzi A & Munster V (2020) Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol* 5: 562–569

Leung GM, Hedley AJ, Ho L-M, Chau P, Wong IOL, Thach TQ, Ghani AC, Donnelly CA, Fraser C, Riley S, *et al* (2004) The Epidemiology of Severe Acute Respiratory Syndrome in

the 2003 Hong Kong Epidemic: An Analysis of All 1755 Patients. *Ann Intern Med* 141: 662

Li M, Ferretti M, Ying B, Descamps H, Lee E, Dittmar M, Lee JS, Whig K, Kamalia B, Dohnalová L, *et al* (2021) Pharmacological activation of STING blocks SARS-CoV-2 infection. *Sci Immunol* 6: eabi9007

Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, Zhao C, Zhang Q, Liu H, Nie L, *et al* (2020) The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell* 182: 1284-1294.e9

Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, Somasundaran M, Sullivan JL, Luzuriaga K, Greenough TC, *et al* (2003) Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 426: 450–454

Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, Wang H, Crameri G, Hu Z, Zhang H, *et al* (2005) Bats Are Natural Reservoirs of SARS-Like Coronaviruses. *Science* 310: 676–679

Lim PL, Kurup A, Gopalakrishna G, Chan KP, Wong CW, Ng LC, Se-Thoe SY, Oon L, Bai X, Stanton LW, *et al* (2004) Laboratory-Acquired Severe Acute Respiratory Syndrome. *N Engl J Med* 350: 1740–1745

Littman RJ (2009) The Plague of Athens: Epidemiology and Paleopathology: THE PLAGUE OF ATHENS. *Mt Sinai J Med* 76: 456–467

Liu W, Peng L, Liu H & Hua S (2015) Pulmonary Function and Clinical Manifestations of Patients Infected with Mild Influenza A Virus Subtype H1N1: A One-Year Follow-Up. *PLoS ONE* 10: e0133698

Liu Y, Liu J, Plante KS, Plante JA, Xie X, Zhang X, Ku Z, An Z, Scharton D, Schindewolf C, *et al* (2022) The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. *Nature* 602: 294–299

Lloyd SB, Kent SJ & Winnall WR (2014) The High Cost of Fidelity. *AIDS Research and Human Retroviruses* 30: 8–16

Lokugamage KG, Hage A, de Vries M, Valero-Jimenez AM, Schindewolf C, Dittmann M, Rajsbaum R & Menachery VD (2020) Type I Interferon Susceptibility Distinguishes SARS-CoV-2 from SARS-CoV. *J Virol* 94: e01410-20

Long Q-X, Tang X-J, Shi Q-L, Li Q, Deng H-J, Yuan J, Hu J-L, Xu W, Zhang Y, Lv F-J, *et al* (2020) Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat Med* 26: 1200–1204

Lopez Bernal J, Andrews N, Gower C, Gallagher E, Simmons R, Thelwall S, Stowe J, Tessier E, Groves N, Dabrera G, *et al* (2021) Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant. *N Engl J Med* 385: 585–594

Lou Z & Rao Z (2022) The Life of SARS-CoV-2 Inside Cells: Replication–Transcription Complex Assembly and Function. *Annu Rev Biochem* 91: 381–401

Lythgoe KA, Hall M, Ferretti L, de Cesare M, MacIntyre-Cockett G, Trebes A, Andersson M, Otecko N, Wise EL, Moore N, *et al* (2021) SARS-CoV-2 within-host diversity and transmission. *Science* 372: eabg0821

Lytras S, Hughes J, Martin D, Swanepoel P, de Klerk A, Lourens R, Kosakovsky Pond SL, Xia W, Jiang X & Robertson DL (2022) Exploring the Natural Origins of SARS-CoV-2 in the Light of Recombination. *Genome Biology and Evolution* 14: evac018

Maan I, Paraskevopoulou SM, Cwynarski K, Shrestha M, Waters L, Miller R & Ahmed N (2022) Prolonged SARS-CoV-2 shedding in a person living with advanced HIV and diffuse large B-cell lymphoma: a case report. *Infectious Diseases* 54: 529–533

Madden EA & Diamond MS (2022) Host cell-intrinsic innate immune recognition of SARS-CoV-2. *Current Opinion in Virology* 52: 30–38

Maris NA, Dessing MC, de Vos AF, Bresser P, van der Zee JS, Jansen HM, Spek CA & van der Poll T (2006) Toll-like receptor mRNA levels in alveolar macrophages after inhalation of endotoxin. *European Respiratory Journal* 28: 622–626

Marra MA, Jones SJM, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YSN, Khattra J, Asano JK, Barber SA, Chan SY, *et al* (2003) The Genome Sequence of the SARS-Associated Coronavirus. *Science* 300: 1399–1404

Martin MA & Koelle K (2021) Comment on “Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2”. *Sci Transl Med* 13: eabh1803

Matsuyama S, Nagata N, Shirato K, Kawase M, Takeda M & Taguchi F (2010) Efficient Activation of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein by the Transmembrane Protease TMPRSS2. *J Virol* 84: 12658–12664

McMahan K, Yu J, Mercado NB, Loos C, Tostanoski LH, Chandrashekar A, Liu J, Peter L, Atyeo C, Zhu A, *et al* (2021) Correlates of protection against SARS-CoV-2 in rhesus macaques. *Nature* 590: 630–634

McNab F, Mayer-Barber K, Sher A, Wack A & O’Garra A (2015) Type I interferons in infectious disease. *Nat Rev Immunol* 15: 87–103

Merad M, Blish CA, Sallusto F & Iwasaki A (2022) The immunology and immunopathology of COVID-19. *Science* 375: 1122–1127

Mesev EV, LeDesma RA & Ploss A (2019) Decoding type I and III interferon signalling during viral infection. *Nat Microbiol* 4: 914–924

Moreno-Eutimio MA, López-Macías C & Pastelin-Palacios R (2020) Bioinformatic analysis and identification of single-stranded RNA sequences recognized by TLR7/8 in the SARS-CoV-2, SARS-CoV, and MERS-CoV genomes. *Microbes and Infection* 22: 226–229

Morens DM & Fauci AS (2020) Emerging Pandemic Diseases: How We Got to COVID-19. *Cell* 182: 1077–1092

Morens DM, Folkers GK & Fauci AS (2004) The challenge of emerging and re-emerging infectious diseases. *Nature* 430: 242–249

Morens DM, Folkers GK & Fauci AS (2008) Emerging infections: a perpetual challenge. *The Lancet Infectious Diseases* 8: 710–719

Morens DM, Taubenberger JK & Fauci AS (2022) Universal Coronavirus Vaccines — An Urgent Need. *N Engl J Med* 386: 297–299

Moriyama M, Hugentobler WJ & Iwasaki A (2020) Seasonality of Respiratory Viral Infections. *Annu Rev Virol* 7: 83–101

Müller MA, Corman VM, Jores J, Meyer B, Younan M, Liljander A, Bosch B-J, Lattwein E, Hilali M, Musa BE, *et al* (2014) MERS Coronavirus Neutralizing Antibodies in Camels, Eastern Africa, 1983–1997. *Emerg Infect Dis* 20

Nalbandian A, Sehgal K, Gupta A, Madhavan MV, McGroder C, Stevens JS, Cook JR, Nordvig AS, Shalev D, Sehrawat TS, *et al* (2021) Post-acute COVID-19 syndrome. *Nat Med* 27: 601–615

Nan X, Iles IA, Yang B & Ma Z (2022) Public Health Messaging during the COVID-19 Pandemic and Beyond: Lessons from Communication Science. *Health Communication* 37: 1–19

Nicholson MD, Endler L, Popa A, Genger J-W, Bock C, Michor F & Bergthaler A (2021) Response to comment on “Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2”. *Sci Transl Med* 13: eabj3222

O’Toole Á, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, Colquhoun R, Ruis C, Abu-Dahab K, Taylor B, *et al* (2021) Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evolution* 7: veab064

Paget J, Caini S, Del Riccio M, van Waarden W & Meijer A (2022) Has influenza B/Yamagata become extinct and what implications might this have for quadrivalent influenza vaccines? *Eurosurveillance* 27

Parrish CR, Holmes EC, Morens DM, Park E-C, Burke DS, Calisher CH, Laughlin CA, Saif LJ & Daszak P (2008) Cross-Species Virus Transmission and the Emergence of New Epidemic Diseases. *Microbiol Mol Biol Rev* 72: 457–470

Parry J (2004) Breaches of safety regulations are probable cause of recent SARS outbreak, WHO says. *BMJ* 328: 1222.3

Peacock TP, Goldhill DH, Zhou J, Baillon L, Frise R, Swann OC, Kugathasan R, Penn R, Brown JC, Sanchez-David RY, *et al* (2021) The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. *Nat Microbiol* 6: 899–909

Pedersen NC, Evermann JF, McKeirnan AJ & Ott RL (1984) Pathogenicity studies of feline coronavirus isolates 79-1146 and 79-1683. *Am J Vet Res* 45: 2580–2585

Peiris J, Chu C, Cheng V, Chan K, Hung I, Poon L, Law K, Tang B, Hon T, Chan C, *et al* (2003a) Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *The Lancet* 361: 1767–1772

Peiris JSM, Lai ST, Poon LLM, Guan Y, Yam LYC, Lim W, Nicholls J, Yee WKS, Yan WW, Cheung MT, *et al* (2003b) Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* 361: 1319–1325

Perera RA, Wang P, Gomaa MR, El-Shesheny R, Kandeil A, Bagato O, Siu LY, Shehata MM, Kayed AS, Moatasim Y, *et al* (2013) Seroepidemiology for MERS coronavirus using microneutralisation and pseudoparticle virus neutralisation assays reveal a high prevalence of antibody in dromedary camels in Egypt, June 2013. *Eurosurveillance* 18

Perlman S & Netland J (2009) Coronaviruses post-SARS: update on replication and pathogenesis. *Nat Rev Microbiol* 7: 439–450

Petersen E, Koopmans M, Go U, Hamer DH, Petrosillo N, Castelli F, Storgaard M, Al Khalili S & Simonsen L (2020) Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *The Lancet Infectious Diseases* 20: e238–e244

Petrova VN & Russell CA (2018) The evolution of seasonal influenza viruses. *Nat Rev Microbiol* 16: 47–60

Pfefferle S, Günther T, Kobbe R, Czech-Sioli M, Nörz D, Santer R, Oh J, Kluge S, Oestereich L, Peldschus K, *et al* (2021) SARS Coronavirus-2 variant tracing within the first Coronavirus Disease 19 clusters in northern Germany. *Clinical Microbiology and Infection* 27: 130.e5-130.e8

Piccoli L, Park Y-J, Tortorici MA, Czudnochowski N, Walls AC, Beltramello M, Silacci-Fregni C, Pinto D, Rosen LE, Bowen JE, *et al* (2020) Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology. *Cell* 183: 1024-1042.e21

Pircher H, Moskophidis D, Rohrer U, Bürki K, Hengartner H & Zinkernagel RM (1990) Viral escape by selection of cytotoxic T cell-resistant virus variants in vivo. *Nature* 346: 629–633

Planas D, Veyer D, Baidaliuk A, Staropoli I, Guivel-Benhassine F, Rajah MM, Planchais C, Porrot F, Robillard N, Puech J, *et al* (2021) Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* 596: 276–280

Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, Zhang X, Muruato AE, Zou J, Fontes-Garfias CR, *et al* (2021) Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* 592: 116–121

Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, Perez JL, Pérez Marc G, Moreira ED, Zerbini C, *et al* (2020) Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N Engl J Med* 383: 2603–2615

Popa A, Genger J-W, Nicholson MD, Penz T, Schmid D, Aberle SW, Agerer B, Lercher A, Endler L, Colaço H, *et al* (2020) Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. *Sci Transl Med* 12: eabe2555

Porter R (1998) The greatest benefit to mankind: a medical history of humanity 1st American ed. New York: W. W. Norton

Prentiss M, Chu A & Berggren KK (2022) Finding the infectious dose for COVID-19 by applying an airborne-transmission model to superspreader events. *PLoS ONE* 17: e0265816

Qi H, Liu B, Wang X & Zhang L (2022) The humoral response and antibodies against SARS-CoV-2 infection. *Nat Immunol* 23: 1008–1020

Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, du Plessis L & Pybus OG (2020) A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* 5: 1403–1407

Ranjan R & Biswal JK (2022) Monkeypox: Re-Emerging Zoonotic Threat. *Zoonotic Diseases* 2: 234–246

Roper RL & Rehm KE (2009) SARS vaccines: where are we? *Expert Review of Vaccines* 8: 887–898

Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, Zimmer T, Thiel V, Janke C, Guggemos W, *et al* (2020) Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. *N Engl J Med* 382: 970–971

Rothman JA, Loveless TB, Kapcia J, Adams ED, Steele JA, Zimmer-Faust AG, Langlois K, Wanless D, Griffith M, Mao L, *et al* (2021) RNA Viromics of Southern California Wastewater and Detection of SARS-CoV-2 Single-Nucleotide Variants. *Appl Environ Microbiol* 87: e01448-21

Rothman KJ, Greenland S & Lash TL (2008) Modern epidemiology 3rd ed., thoroughly rev. and updated. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins

Sawicki SG & Sawicki DL (1995) Coronaviruses use Discontinuous Extension for Synthesis of Subgenome-Length Negative Strands. In *Corona- and Related Viruses*, Talbot PJ & Levy GA (eds) pp 499–506. Boston, MA: Springer US

Sawicki SG, Sawicki DL & Siddell SG (2007) A Contemporary View of Coronavirus Transcription. *Journal of Virology* 81: 20–29

Schalk AF & Hawn MC (1931) An Apparently New Respiratory Disease of Baby Chicks. *Journal of the American Veterinary Medical Association* 78: 413–423

Senio K (2003) Recent Singapore SARS case a laboratory accident. *The Lancet Infectious Diseases* 3: 679

Sette A & Crotty S (2021) Adaptive immunity to SARS-CoV-2 and COVID-19. *Cell* 184: 861–880

Shi G & Suzuki T (2018) Molecular Basis of Encapsidation of Hepatitis C Virus Genome. *Front Microbiol* 9: 396

Shu Y & McCauley J (2017) GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance* 22

Simmons G, Gosalia DN, Rennekamp AJ, Reeves JD, Diamond SL & Bates P (2005) Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. *Proc Natl Acad Sci USA* 102: 11876–11881

Slifka MK & Whitton JL (2000) Antigen-Specific Regulation of T Cell–Mediated Cytokine Production. *Immunity* 12: 451–457

Snijder EJ, Decroly E & Ziebuhr J (2016) The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. In *Advances in Virus Research* pp 59–126. Elsevier

Sobel Leonard A, Weissman DB, Greenbaum B, Ghedin E & Koelle K (2017) Transmission Bottleneck Size Estimation from Pathogen Deep-Sequencing Data, with an Application to Human Influenza A Virus. *J Virol* 91: e00171-17

Sola I, Almazán F, Zúñiga S & Enjuanes L (2015) Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annu Rev Virol* 2: 265–288

Sominina A, Danilenko D, Komissarov A, Karpova L, Pisareva M, Fadeev A, Konovalova N, Eropkin M, Stolyarov K, Shtro A, *et al* (2022) Resurgence of Influenza Circulation in the Russian Federation during the Delta and Omicron COVID-19 Era. *Viruses* 14: 1909

Sonnleitner ST, Prelog M, Sonnleitner S, Hinterbichler E, Halbfurter H, Kopecky DBC, Almanzar G, Koblmüller S, Sturmbauer C, Feist L, *et al* (2022) Cumulative SARS-CoV-2 mutations and corresponding changes in immunity in an immunocompromised patient indicate viral evolution within the host. *Nat Commun* 13: 2560

Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, Liu W, Bi Y & Gao GF (2016) Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends in Microbiology* 24: 490–502

Sudre CH, Murray B, Varsavsky T, Graham MS, Penfold RS, Bowyer RC, Pujol JC, Klaser K, Antonelli M, Canas LS, *et al* (2021) Attributes and predictors of long COVID. *Nat Med* 27: 626–631

Suthar MS, Zimmerman MG, Kauffman RC, Mantus G, Linderman SL, Hudson WH, Vanderheiden A, Nyhoff L, Davis CW, Adekunle O, *et al* (2020) Rapid Generation of Neutralizing Antibody Responses in COVID-19 Patients. *Cell Reports Medicine* 1: 100040

Taubenberger JK, Kash JC & Morens DM (2019) The 1918 influenza pandemic: 100 years of questions answered and unanswered. *Sci Transl Med* 11: eaau5485

Tay MZ, Poh CM, Rénia L, MacAry PA & Ng LFP (2020) The trinity of COVID-19: immunity, inflammation and intervention. *Nat Rev Immunol* 20: 363–374

The COVID STEROID 2 Trial Group, Russell L, Uhre KR, Lindgaard ALS, Degn JF, Wetterslev M, Sivapalan P, Anthon CT, Mikkelsen VS, la Porta LC, *et al* (2021) Effect of 12 mg vs 6 mg of Dexamethasone on the Number of Days Alive Without Life Support in Adults With COVID-19 and Severe Hypoxemia: The COVID STEROID 2 Randomized Trial. *JAMA* 326: 1807

The Lancet (2021) Genomic sequencing in pandemics. *The Lancet* 397: 445

The WHO Rapid Evidence Appraisal for COVID-19 Therapies (REACT) Working Group, Sterne JAC, Murthy S, Diaz JV, Slutsky AS, Villar J, Angus DC, Annane D, Azevedo LCP, Berwanger O, *et al* (2020) Association Between Administration of Systemic Corticosteroids and Mortality Among Critically Ill Patients With COVID-19: A Meta-analysis. *JAMA* 324: 1330

Thoms M, Buschauer R, Ameismeier M, Koepke L, Denk T, Hirschenberger M, Kratzat H, Hayn M, Mackens-Kiani T, Cheng J, *et al* (2020) Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science* 369: 1249–1255

Thomson EC, Rosen LE, Shepherd JG, Spreafico R, da Silva Filipe A, Wojcechowskyj JA, Davis C, Piccoli L, Pascall DJ, Dillen J, *et al* (2021) Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell* 184: 1171-1187.e20

Thorne LG, Bouhaddou M, Reuschl A-K, Zuliani-Alvarez L, Polacco B, Pelin A, Batra J, Whelan MVX, Hosmillo M, Fossati A, *et al* (2022) Evolution of enhanced innate immune evasion by SARS-CoV-2. *Nature* 602: 487–495

Tian L, Zhou W, Wu X, Hu Z, Qiu L, Zhang H, Chen X, Zhang S & Lu Z (2022) CTLs: Killers of intracellular bacteria. *Front Cell Infect Microbiol* 12: 967679

Toor SM, Saleh R, Sasidharan Nair V, Taha RZ & Elkord E (2021) T-cell responses and therapies against SARS-CoV-2 infection. *Immunology* 162: 30–43

Tortorici MA & Veesler D (2019) Structural insights into coronavirus entry. In *Advances in Virus Research* pp 93–116. Elsevier

Tuite AR, Ng V, Rees E & Fisman D (2020) Estimation of COVID-19 outbreak size in Italy. *The Lancet Infectious Diseases* 20: 537

Tyrrell DAJ & Bynoe ML (1965) Cultivation of a Novel Type of Common-cold Virus in Organ Cultures. *BMJ* 1: 1467–1470

Vabret N, Britton GJ, Gruber C, Hegde S, Kim J, Kuksin M, Levantovsky R, Malle L, Moreira A, Park MD, *et al* (2020) Immunology of COVID-19: Current State of the Science. *Immunity* 52: 910–941

Vazquez C, Swanson SE, Negatu SG, Dittmar M, Miller J, Ramage HR, Cherry S & Jurado KA (2021) SARS-CoV-2 viral proteins NSP1 and NSP13 inhibit interferon activation through distinct mechanisms. *PLoS ONE* 16: e0253089

Viana R, Moyo S, Amoako DG, Tegally H, Scheepers C, Althaus CL, Anyaneji UJ, Bester PA, Boni MF, Chand M, *et al* (2022) Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature* 603: 679–686

Vignuzzi M, Stone JK, Arnold JJ, Cameron CE & Andino R (2006) Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 439: 344–348

Vihta KD, Pouwels KB, Peto1 TE, Pritchard E, House T, Studley R, Rourke E, Cook D, Diamond I, Cook D, *et al* (2022) Omicron-associated changes in SARS-CoV-2 symptoms in the United Kingdom. *Clinical Infectious Diseases*: ciac613

Villa TG, Abril AG, Sánchez S, de Miguel T & Sánchez-Pérez A (2021) Animal and human RNA viruses: genetic variability and ability to overcome vaccines. *Arch Microbiol* 203: 443–464

Virgin HW, Wherry EJ & Ahmed R (2009) Redefining Chronic Viral Infection. *Cell* 138: 30–50

V'kovski P, Kratzel A, Steiner S, Stalder H & Thiel V (2021) Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol* 19: 155–170

Voskoboinik I, Whisstock JC & Trapani JA (2015) Perforin and granzymes: function, dysfunction and human pathology. *Nat Rev Immunol* 15: 388–400

Voysey M, Clemens SAC, Madhi SA, Weckx LY, Folegatti PM, Aley PK, Angus B, Baillie VL, Barnabas SL, Bhorat QE, *et al* (2021) Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *The Lancet* 397: 99–111

Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT & Veesler D (2020) Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 181: 281-292.e6

Wang Y, Zhang D, Du G, Du R, Zhao J, Jin Y, Fu S, Gao L, Cheng Z, Lu Q, *et al* (2020) Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. *The Lancet* 395: 1569–1578

Ward M & Rawle F (2022) Teaching the process of science through COVID -19 pandemic themes. *Biochem Molecular Bio Educ* 50: 307–309

Watanabe Y, Allen JD, Wrapp D, McLellan JS & Crispin M (2020) Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* 369: 330–333

Whitby M (2008) :Plague and the End of Antiquity: The Pandemic of 541–750. *AM HIST REV* 113: 1211–1212

Wilhelm A, Schoth J, Meinert-Berning C, Agrawal S, Bastian D, Orschler L, Ciesek S, Teichgräber B, Wintgens T, Lackner S, *et al* (2022) Wastewater surveillance allows early detection of SARS-CoV-2 omicron in North Rhine-Westphalia, Germany. *Science of The Total Environment* 846: 157375

Wilkinson E, Giovanetti M, Tegally H, San JE, Lessells R, Cuadros D, Martin DP, Rasmussen DA, Zekri A-RN, Sangare AK, *et al* (2021) A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa. *Science* 374: 423–431

Wille M & Barr IG (2022) Resurgence of avian influenza virus. *Science* 376: 459–460

de Wit E, van Doremalen N, Falzarano D & Munster VJ (2016) SARS and MERS: recent insights into emerging coronaviruses. *Nat Rev Microbiol* 14: 523–534

Wolter N, Jassat W, Walaza S, Welch R, Moultrie H, Groome M, Amoako DG, Everatt J, Bhiman JN, Scheepers C, *et al* (2022) Early assessment of the clinical severity of the SARS-CoV-2 omicron variant in South Africa: a data linkage study. *The Lancet* 399: 437–446

Woo PCY, Lau SKP, Chu C, Chan K, Tsoi H, Huang Y, Wong BHL, Poon RWS, Cai JJ, Luk W, *et al* (2005) Characterization and Complete Genome Sequence of a Novel Coronavirus, Coronavirus HKU1, from Patients with Pneumonia. *J Virol* 79: 884–895

Woolhouse J, Adair K & Brierley L (2013) RNA Viruses: A Case Study of the Biology of Emerging Infectious Diseases. *Microbiol Spectr* 1: 1.1.03

World Health Organization (2020a) Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19): https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf (accessed 10/08/2022)

World Health Organization (2020b) Novel Coronavirus (2019-nCoV): Situation Report, 11 World Health Organization

World Health Organization (2020c) Novel Coronavirus(2019-nCoV): Situation Report, 22 World Health Organization

World Health Organization (2020d) Coronavirus disease 2019 (COVID-19): Situation Report, 51 World Health Organization

World Health Organization (2022) Tracking SARS-CoV-2 variants: https://www.who.int/activities/tracking-SARS-CoV-2-variants (accessed: 22/08/2022).

Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, Hu Y, Tao Z-W, Tian J-H, Pei Y-Y, *et al* (2020) A new coronavirus associated with human respiratory disease in China. *Nature* 579: 265–269

Xia H, Cao Z, Xie X, Zhang X, Chen JY-C, Wang H, Menachery VD, Rajsbaum R & Shi P-Y (2020) Evasion of Type I Interferon by SARS-CoV-2. *Cell Reports* 33: 108234

Xu R-H, He J-F, Evans MR, Peng G-W, Field HE, Yu D-W, Lee C-K, Luo H-M, Lin W-S, Lin P, *et al* (2004) Epidemiologic Clues to SARS Origin in China. *Emerg Infect Dis* 10: 1030–1037

Yamada T, Sato S, Sotoyama Y, Orba Y, Sawa H, Yamauchi H, Sasaki M & Takaoka A (2021) RIG-I triggers a signaling-abortive anti-SARS-CoV-2 defense in human lung cells. *Nat Immunol* 22: 820–828

Yount B, Roberts RS, Sims AC, Deming D, Frieman MB, Sparks J, Denison MR, Davis N & Baric RS (2005) Severe Acute Respiratory Syndrome Coronavirus Group-Specific Open Reading Frames Encode Nonessential Functions for Replication in Cell Cultures and Mice. *J Virol* 79: 14909–14922

Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus ADME & Fouchier RAM (2012) Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *N Engl J Med* 367: 1814–1820

Zeller M, Gangavarapu K, Anderson C, Smither AR, Vanchiere JA, Rose R, Snyder DJ, Dudas G, Watts A, Matteson NL, *et al* (2021) Emergence of an early SARS-CoV-2 epidemic in the United States. *Cell* 184: 4939-4952.e15

Zhang H, Zhou G, Zhi L, Yang H, Zhai Y, Dong X, Zhang X, Gao X, Zhu Y & He F (2005) Association between Mannose-Binding Lectin Gene Polymorphisms and Susceptibility to Severe Acute Respiratory Syndrome Coronavirus Infection. *The Journal of Infectious Diseases* 192: 1355–1361

Zhang J, Ye H, Liu Y, Liao M & Qi W (2022) Resurgence of H5N6 avian influenza virus in 2021 poses new threat to public health. *The Lancet Microbe* 3: e558

Zheng M, Karki R, Williams EP, Yang D, Fitzpatrick E, Vogel P, Jonsson CB & Kanneganti T-D (2021) TLR2 senses the SARS-CoV-2 envelope protein to produce inflammatory cytokines. *Nat Immunol* 22: 829–838

Zhong NS, Zheng BJ, Li YM, Poon LLM, Xie ZH, Chan KH, Li PH, Tan SY, Chang Q, Xie JP, *et al* (2003) Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *Lancet* 362: 1353–1358

Zhou H, Ji J, Chen X, Bi Y, Li J, Wang Q, Hu T, Song H, Zhao R, Chen Y, *et al* (2021) Identification of novel bat coronaviruses sheds light on the evolutionary origins of SARS-CoV-2 and related viruses. *Cell* 184: 4380-4391.e14

Zhou Y, Lu K, Pfefferle S, Bertram S, Glowacka I, Drosten C, Pöhlmann S & Simmons G (2010) A Single Asparagine-Linked Glycosylation Site of the Severe Acute Respiratory Syndrome Coronavirus Spike Glycoprotein Facilitates Inhibition by Mannose-Binding Lectin through Multiple Mechanisms. *J Virol* 84: 8753–8764

Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, *et al* (2020) A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 382: 727–733

Zinkernagel RM & Doherty PC (1974) Restriction of in vitro T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature* 248: 701–702

Zumla A, Hui DS & Perlman S (2015) Middle East respiratory syndrome. *The Lancet* 386: 995–1007

Zwart MP & Elena SF (2015) Matters of Size: Genetic Bottlenecks in Virus Infection and Their Potential Impact on Evolution. *Annu Rev Virol* 2: 161–179

Appendix

Licenses

Figure 1: License Number 5518180087346, obtained March 29th, 2023 (Content Publisher: Elsevier)

Figure 2: License Number 1339392-1, obtained March 30th, 2023 (Rightsholder: Annual Reviews, Inc.)

Figure 5: License Number 5518190799690, obtained March 29th, 2023 (Content Publisher: Elsevier)

Figure 6: License Number 5518191108665, obtained March 29th, 2023 (Content Publisher: Springer Nature)

Figure 7: License Number 5518200077052, obtained March 29th, 2023 (Content Publisher: Springer Nature)

Figure 8, Publication and Supplementary Material: License Number OP-00145573, obtained March 21st, 2023 (Content Publisher: AAAS)

Curriculum Vitae

Academic CV

PERSONAL INFORMATION

Name	Jakob-Wendelin Genger
Date of Birth	23.03.1992, Berlin, Germany

EDUCATION

Since Sep 2018	**PhD Student in the Lab of Prof. Dr. Andreas Bergthaler** at the CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences and Medical University of Vienna, Vienna, Austria
Oct 2016 – Oct 2017	**M. Sc. Applied Biosciences and Biotechnology, Imperial College London**, London, United Kingdom Final grade: Distinction and Award for Outstanding Student Runner-Up
Oct 2015 – Sep 2016	**M. Sc. Molecular Biotechnology, Heidelberg University**, Heidelberg, Germany
Oct 2012 – Sep 2015	**B. Sc. Molecular Biotechnology, Heidelberg University**, Heidelberg, Germany Final grade: 1.7

RESEARCH EXPERIENCE

Since Sep 2018	**CeMM Research Center for Molecular Medicine and Medical University of Vienna**, Vienna, Austria *Research group: Viral Immunology,* ***Prof. Dr. Andreas Bergthaler*** Project: "Identification of Immunometabolites that Regulate the Activity of CD8+ T cells during viral infection"
Jan – Apr 2018	**Centre for Structural Systems Biology (CSSB)**, Hamburg, Germany *Research group: Structural Biology of Viruses,* ***Prof. Dr. Kay Grünewald*** Project: "The role of glycoproteins on the surface of HSV-1 in cell entry and virus particle formation"
May – Aug 2017	**Imperial College London**, London, United Kingdom *Research group: Synthetic Biology and Metabolic Engineering,* ***Dr. John Heap*** Master Thesis: "Rewiring the transcriptional network of *Synechocystis* sp.: A forward genetics approach to generate new phenotypes of industrial relevance"
Jan – Mar 2016 **Jul – Sep 2016**	**European Molecular Biology Laboratory (EMBL)**, Heidelberg, Germany *Research group: Structural and Computational Biology,* ***Dr. Martin Beck*** Project: "Establishment of a versatile protocol to study the process of protein complex assembly"
Mar – Aug 2015	**University of Heidelberg**, Heidelberg, Germany *Research group: Virus-Host Interactions* Bachelor Thesis: "Systematic comparison of regulatory elements in recombinant adeno-associated viral genomes - towards the next generation of designer vectors for human gene therapy"

SKILLS & ACHIEVEMENTS

Lab Skills

- *In vitro* work — Molecular cloning, protein biochemistry (SDS-PAGE, Western Blot assays, Protein purification), *In vitro* translation systems
- Cell Biology — Cell culture, FACS, AAV production, High-Throughput Screening
- Mouse work — Intraperitoneal injection, intravenous injection

IT Skills

- Machine Learning — Professional Certificate for Machine Learning and Artificial Intelligence – Imperial College London, 2022
- Pathogen Modelling — Creation of NextstrainAustria – a platform for phylodynamic modelling and scientific communication of SARS-CoV-2 transmission in populations (https://www.sarscov2-austria.org/cemm/nextstrain-austria/)
- Structural Biology — Experienced with Python, Perl, IMOD + PEET, Relion, ImageJ for cryo-electron microscopy

Scholarships & Awards

- DOC Fellowship of the Austrian Academy of Sciences ÖAW – 2020 to 2023
- Award for Outstanding Student Runner Up from Imperial College London
- Scholarship of the German National Academic Foundation ("Studienstiftung des deutschen Volkes") - March 2013 to September 2018
- 4th Prize in the national competition of the German student research competition ("Jugend forscht") 2011
- Laureate of the Berlin Nature Conservation Prize 2011 awarded by Stiftung Naturschutz Berlin

PUBLICATIONS

Emtenani, S., Martin, E.T., Gyoergy, A., Bicher, J., **Genger, J.-W.**, Köcher, T., Akhmanova, M., Guarda, M., Roblek, M., Bergthaler, A., et al. (2022). Macrophage mitochondrial bioenergetics and tissue invasion are boosted by an Atossa-Porthos axis in Drosophila. The EMBO Journal *41*. https://doi.org/10.15252/embj.2021109049.

Nicholson, M.D., Endler, L., Popa, A., **Genger, J.-W.**, Bock, C., Michor, F., and Bergthaler, A. (2021). Response to comment on "Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2." Science Translational Medicine *13*. https://doi.org/10.1126/scitranslmed.abj3222.

Agerer, B., Koblischke, M., Gudipati, V., Montaño-Gutierrez, L. F., Smyth, M., Popa, A., **Genger, J.-W.**, Endler, L., Florian, D. M., Mühlgrabner, V., Graninger, M., Aberle, S. W., Husa, A.-M., Shaw, L. E., Lercher, A., Gattinger, P., Torralba-Gombau, R., Trapin, D., Penz, T., et al. (2021). SARS-CoV-2 mutations in MHC-I-restricted epitopes evade CD8 + T cell responses. Science Immunology, *6*(57), 6461. https://doi.org/10.1126/sciimmunol.abg6461

Leber, W., Lammel, O., Redlberger-Fritz, M., Mustafa-Korninger, M.E., Glehr, R.C., Camp, J., Agerer, B., Lercher, A., Popa, A., **Genger, J.W**., et al. (2021). Rapid, early and accurate SARS-CoV-2 detection using RT-qPCR in primary care: A prospective cohort study (REAP-1). BMJ Open *11*. https://doi.org/10.1136/bmjopen-2020-045225.

Lercher, A., Bhattacharya, A., Popa, A. M., Caldera, M., Schlapansky, M. F., Baazim, H., Agerer, B., Gürtl, B., Kosack, L., Májek, P., Brunner, J. S., Vitko, D., Pinter, T., **Genger, J. W.**, Orlova, A., Pikor, N., Reil, D., Ozsvár-Kozma, M., Kalinke, U., et al. (2019). Type I Interferon Signaling Disrupts the Hepatic Urea Cycle and Alters Systemic Metabolism to Suppress T Cell Function. *Immunity*, *51*(6), 1074-1087.e9. https://doi.org/10.1016/j.immuni.2019.10.014