frontiers | Research Topics

Viral emerging and re-emerging diseases: Basic understanding and future intervention strategies

Edited by
Mohamad S. Hakim, Wenshi Wang, Marco Goeijenbier
and Shailendra Saxena

Published in
Frontiers in Microbiology

ISSN 1664-8714
ISBN 978-2-8325-4677-2
DOI 10.3389/978-2-8325-4677-2

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Viral emerging and re-emerging diseases: Basic understanding and future intervention strategies

Topic editors
Mohamad S. Hakim — Gadjah Mada University, Indonesia
Wenshi Wang — Xuzhou Medical University, China
Marco Goeijenbier — Spaarne Gasthuis, Netherlands
Shailendra Saxena — King George’s Medical University, India

Citation
Hakim, M. S., Wang, W., Goeijenbier, M., Saxena, S., eds. (2024). *Viral emerging and re-emerging diseases: Basic understanding and future intervention strategies*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4677-2

Table of contents

frontiers | Frontiers in Microbiology

TYPE Editorial
PUBLISHED 15 March 2024
DOI 10.3389/fmicb.2024.1395742

OPEN ACCESS

EDITED AND REVIEWED BY
Anna Kramvis,
University of the Witwatersrand, South Africa

*CORRESPONDENCE
Mohamad S. Hakim
✉ m.s.hakim@ugm.ac.id

RECEIVED 04 March 2024
ACCEPTED 06 March 2024
PUBLISHED 15 March 2024

CITATION
Hakim MS, Wang W, Goeijenbier M and Saxena SK (2024) Editorial: Viral emerging and re-emerging diseases: basic understanding and future intervention strategies. *Front. Microbiol.* 15:1395742. doi: 10.3389/fmicb.2024.1395742

Editorial: Viral emerging and re-emerging diseases: basic understanding and future intervention strategies

Mohamad S. Hakim[1,2,3]*, Wenshi Wang[4], Marco Goeijenbier[5,6] and Shailendra K. Saxena[7,8,9]

[1]Department of Microbiology, Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia, [2]Postgraduate School of Molecular Medicine, Erasmus MC-University Medical Center, Rotterdam, Netherlands, [3]Viral Infection Working Group (VIWG), International Society of Antimicrobial Chemotherapy (ISAC), London, United Kingdom, [4]Department of Pathogen Biology and Immunology, Jiangsu Key Laboratory of Immunity and Metabolism, Jiangsu International Laboratory of Immunity and Metabolism, Xuzhou Medical University, Xuzhou, China, [5]Department of Intensive Care, Erasmus MC-University Medical Center, Rotterdam, Netherlands, [6]Department of Intensive Care, Spaarne Gasthuis, Haarlem, Netherlands, [7]Centre for Advanced Research, Faculty of Medicine, King George's Medical University, Lucknow, India, [8]World Society for Virology, Northampton, MA, United States, [9]The Indian Virological Society (IVS), New Delhi, India

KEYWORDS
biology, control, emerging, vaccines, viral diseases

Editorial on the Research Topic
Viral emerging and re-emerging diseases: basic understanding and future intervention strategies

We have observed a recent pandemic of Coronavirus disease 2019 (COVID-19) due to an emerging virus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). SARS-CoV-2 is only one example of contagious virus transmitted to humans and significantly impacting on the global health and economy. In addition to SARS-CoV-2, many other viruses have recently emerged or re-emerged in the human population, including Middle East respiratory syndrome coronavirus (MERS-CoV) that is recently reported to continuously circulate in the kingdom of Saudi Arabia and the United Arab Emirates, Nipah virus (recently reported in India), Influenza A virus, Marburg virus (recently reported in Guinea and the United Republic of Tanzania), Measles, Dengue virus, among others. A close collaboration between the scientific community and public health authorities is highly required to better anticipate for the potential dangers of emerging and re-emerging viral diseases in the near future. A thorough understanding of the biology, immunology, and pathogenesis of viral infections is essential for the development of efficient prevention and treatment measures. Additionally, persistent genomic surveillance, identification of viral reservoirs, and vector management are critical.

The aim of the Research Topic was to present recent research on the growing subject of emerging and re-emerging viral diseases. In more detail, the Research Topic included comprehensive basic and translational aspects to better understand the genetics, diversity, immune pathogenesis, as well as the development of antiviral, vaccines, and diagnostic tools of emerging and re-emerging viral diseases. In addition to RNA viruses, we also focused on DNA viruses, such as monkeypox (mpox) virus and papillomaviruses.

By highlighting these recent studies, we hope to improve our comprehension of the complexity of viral-host interactions and to more adequately prepare for upcoming large-scale viral outbreaks and pandemics.

Influenza viruses have been the cause of repeating epidemics and pandemics in the last century. Rafique et al. reviewed the global spreading of the H5N8 influenza virus. The highly pathogenic avian influenza (HPAI) subtype H5N8 was first reported in wild birds in China. To minimize the risk of HPAI incidence in the near future, strengthening biosecurity practice and increased active surveillance of wild birds are required. Additionally, to stop future HPAI H5N8 outbreaks, poultry birds in high-risk countries should receive vaccinations.

Recently, we experienced an epidemic due to monkeypox virus (MPXV). The virus could persistently survive on the environmental surfaces or in wastewater, enhancing its potential transmission. Thus, Taha et al. reviewed the persistence of MPXV on a variety of environmental surfaces. They also discussed various methods of disinfection and wastewater treatment to prevent transmission.

Detecting novel viruses or novel virus variants circulating in animals is important to identify potentially emerging viruses with epidemic or pandemic potential. Utilizing the advanced next generation sequencing (NGS) techniques, Li, Xiao et al. reported two novel strains of papillomaviruses, PV-HMU-1 and PV-HMU-2, in the nasal and throat swab samples collected from belugas *(Delphinapterus leucas)* at Polar Ocean Parks in Dalian and Qingdao. Li, Du et al. identified GI.1aP-GI.2 recombinants of rabbit hemorrhagic disease virus (RHDV) in domestic rabbits in China. RHDV is a highly infectious agent that causes acute multi-organ hemorrhagic syndrome with high morbidity and fatality rates in rabbits. Importantly, animal studies demonstrated that the recombinant GI.1aP-GI.2 variant had a moderately increased pathogenicity. Viral discovery in our Research Topic is also highlighted by the study of Xiao et al. that reported the first identification of Canine circovirus (CanineCV) in cats. They showed that about 9% of dogs and 3.4% of cats were CanineCV-positive. The whole genome analysis showed that the first cat-derived CanineCV belonged to the genotype 3. Xu et al. investigated the molecular epidemiology and virulence of goose astroviruses (GAstV). Phylogenetic studies revealed that the representative viruses belonged to genotype 2. Furthermore, subsequent *in vitro* and *in vivo* experiments showed notable variations in the virulence and pathogenicity across susceptible cells and embryos. While Zhang et al. examined viral-associated diarrhea from porcine diarrheal samples (including small intestinal contents and tissue, as well as feces) collected from five provinces in China from 2021 to 2023. They found that the most frequently detected virus was porcine epidemic diarrhea virus (PEDV), followed by porcine rotavirus (PRoV), porcine delta coronavirus (PDCoV), swine acute diarrhea syndrome coronavirus (SADS-CoV), and transmissible gastroenteritis coronavirus (transmissible gastroenteritis virus, TGEV). Interestingly, co-infections were commonly detected. All these findings emphasized the need of continuous surveillance in animals to identify novel pathogens or novel variants with altered pathogenic potential. Studies in human cases are also essential, as shown by Huang et al. that investigated the clinical picture of a novel bunyavirus infection causing severe fever with thrombocytopenia syndrome (SFTS) in infected patients.

Viral adaptation in humans following zoonotic transmission of emerging viruses is also an important aspect of virus evolution. By using the molecular dynamics simulation, Elli et al. examined the interactions between N1 neuraminidase (NA), particularly amino acid (aa) at position 347, with its substrates. Sialyllacto-N-tetraoses 3’SLN-LC and 6’SLN-LC are sialoglycan molecules correspond to the neuraminidase substrates in birds and humans, respectively. Their simulation demonstrated that tyrosine at position 347 played an essential role in the preference of neuraminidase binding for the avian-type substrates. They also emphasized the importance of an alterations at this position as a marker of host tropism (host-shift) and adaptive evolution of influenza viruses.

Our Research Topic also includes some studies on SARS-CoV-2. It is known that bats are the natural reservoir of coronaviruses (CoVs). Bat CoVs can occasionally jump to other mammals, including humans. Li, Tian et al. built a deep learning (DL) method to predict the evolutionary adaptation of bat CoVs to other mammals. The model can be utilized to predict transmission of bat CoVs to the human population. As the first step in viral infection, interaction between viral factors and host receptors is crucial before viral replication to ensue. Xia reviewed the identification of viral-host interactions by identifying host receptors for viral entry. He discussed the identification of ACE2 as the host receptor for SARS-CoV-2, as well as other candidate receptors and cofactors, including neuropilin-1 and integrin αVβ3. He then discussed the potential of the host receptors to be utilized as potential drug targets. In line, Zhou et al. established a novel hACE2 knock-in mouse model to mimic intestinal and respiratory SARS-CoV-2 infection. The generated mouse models were highly susceptible to intranasal SARS-CoV-2 infection. Interestingly, the infected mice not only developed lung diseases, but also acquired intestinal infection. Thus, this model is valuable in studying SARS-CoV-2 infection, particularly in the intestine.

Wang et al. tracked the first outbreak of SARS-CoV-2 Omicron BA.5.1.3 in Sanya city, the island province of Hainan, China. From August 1, 2022 to early September 2022, more than 8,500 cases have been confirmed as a result of the outbreak. Importantly, they found unique mutations, including three non-synonymous mutations in *ORF1ab* gene [Y1064N, S2844G, and R3574K within the non-structural proteins (nsp) 3, nsp4, and nsp6, respectively] and one synonymous mutation in *ORF3a* gene (S74S). The unique mutations identified in their study may result in more productive viral replication of the Omicron variant, but it deserves further exploration.

Negi et al. performed a retrospective study to search for prognosis indicators of SARS-CoV-2 infection. They examined the Ct values to evaluate viral load across three different waves of COVID-19 pandemic and correlated with demographic and clinical profiles of COVID-19 patients. During the second wave (mostly due to Delta variant, B.1.617), they observed a notably high percentage of patients with raised viral loads, indicating a higher severity. Their study suggested that in the event of infectious viral epidemic, the Ct value can be utilized as a monitoring tool for viral load and community transmission.

For antiviral development, Reshamwala et al. examined the antiviral properties of willow (*Salix* spp.) bark hot water extracts against enteroviruses and CoVs. Employing various molecular and cellular techniques, they revealed that the bark extracts of *Salix* spp. encompass a number of virucidal compounds that have the potential to exert antiviral actions both in direct and synergistic manner.

In conclusion, we need to improve our understanding for early detection and prevention of future threats of viral infectious diseases. We need to continuously identify viral reservoirs in the animal population, develop safe and efficacious vaccines against emerging and re-emerging viruses, as well as develop fast and efficient diagnostic tools and effective antiviral treatments.

Author contributions

MH: Writing – original draft, Writing – review & editing. WW: Writing – review & editing. MG: Writing – review & editing. SS: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 05 May 2023
DOI 10.3389/fmicb.2023.1157608

OPEN ACCESS

EDITED BY
Wenshi Wang,
Xuzhou Medical University, China

REVIEWED BY
Yang Li,
Chinese Academy of Sciences (CAS), China
Zhijiang Miao,
Erasmus Medical Center, Netherlands

*CORRESPONDENCE
Yigang Tong
✉ tongyigang@mail.buct.edu.cn
Tao Jiang
✉ jiangtao@bmi.ac.cn
Jia-Fu Jiang
✉ jiangjf2008@gmail.com

RECEIVED 02 February 2023
ACCEPTED 03 April 2023
PUBLISHED 05 May 2023

CITATION
Li J, Tian F, Zhang S, Liu S-S, Kang X-P, Li Y-D, Wei J-Q, Lin W, Lei Z, Feng Y, Jiang J-F, Jiang T and Tong Y (2023) Genomic representation predicts an asymptotic host adaptation of bat coronaviruses using deep learning. *Front. Microbiol.* 14:1157608. doi: 10.3389/fmicb.2023.1157608

Genomic representation predicts an asymptotic host adaptation of bat coronaviruses using deep learning

Jing Li[1], Fengjuan Tian[2], Sen Zhang[1], Shun-Shuai Liu[1], Xiao-Ping Kang[1], Ya-Dan Li[1], Jun-Qing Wei[2], Wei Lin[2], Zhongyi Lei[2], Ye Feng[1], Jia-Fu Jiang[1]*, Tao Jiang[1]* and Yigang Tong[2]*

[1]State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, AMMS, Beijing, China, [2]Beijing Advanced Innovation Center for Soft Matter Science and Engineering (BAIC-SM), College of Life Science and Technology, Beijing University of Chemical Technology, Beijing, China

Introduction: Coronaviruses (CoVs) are naturally found in bats and can occasionally cause infection and transmission in humans and other mammals. Our study aimed to build a deep learning (DL) method to predict the adaptation of bat CoVs to other mammals.

Methods: The CoV genome was represented with a method of dinucleotide composition representation (DCR) for the two main viral genes, *ORF1ab* and *Spike*. DCR features were first analyzed for their distribution among adaptive hosts and then trained with a DL classifier of convolutional neural networks (CNN) to predict the adaptation of bat CoVs.

Results and discussion: The results demonstrated inter-host separation and intra-host clustering of DCR-represented CoVs for six host types: Artiodactyla, Carnivora, Chiroptera, Primates, Rodentia/Lagomorpha, and Suiformes. The DCR-based CNN with five host labels (without Chiroptera) predicted a dominant adaptation of bat CoVs to Artiodactyla hosts, then to Carnivora and Rodentia/Lagomorpha mammals, and later to primates. Moreover, a linear asymptotic adaptation of all CoVs (except Suiformes) from Artiodactyla to Carnivora and Rodentia/Lagomorpha and then to Primates indicates an asymptotic bats-other mammals-human adaptation.

Conclusion: Genomic dinucleotides represented as DCR indicate a host-specific separation, and clustering predicts a linear asymptotic adaptation shift of bat CoVs from other mammals to humans via deep learning.

KEYWORDS

bat coronavirus, asymptotic adaptation, deep learning, dinucleotide composition representation (DCR), convolutional neural networks

1. Introduction

RNA viruses from natural reservoir hosts continuously pose a threat to human health, such as coronaviruses (CoVs) from bats (Liu et al., 2021; Wang et al., 2022) and avian influenza viruses from birds (Liu et al., 2014; Sun et al., 2014; Deng et al., 2017). In particular, bat-originated CoVs either caused high-pathogenic but low-transmissible infections of Severe Acute Respiratory Syndrome (SARS) or Middle East Respiratory Syndrome (MERS) or launched a widespread pandemic of low-pathogenic human CoVs, such as HcoV-NL63, HcoV-229E, HcoV-OC43, and HKU1 (Su et al., 2016; Forni et al., 2017). The ongoing global spread of SARS-CoV-2 has not only caused huge damage to public health (WHO, 2022) but also radically changed social habits and lifestyles (West et al., 2020; El-Sayed and Kamel, 2021). Orthocoronavirinae, known as CoV, is one of the two subfamilies in Coronavirinae and consists of four genera of alpha-, beta-, gamma-, and delta-coronaviruses that infect mammalian or avian hosts, especially those specific to species (Woo et al., 2012). The two former CoV members only infect mammals, and the two latter CoVs dominantly infect birds, with some exceptions for mammalian infection (Woo et al., 2012; Ji et al., 2022). According to current CoV databases, almost all human CoVs, with HcoV-OC43 and HKU1 as the exceptional origins of rodents (Forni et al., 2017), have been indicated to have originated in bats (Cui et al., 2019; Ruiz-Aravena et al., 2022). SARS-CoV-2 likely originated from bats as well (Zhou et al., 2020). Additionally, approximately half of the 20 Alphacoronavirus or Betacoronavirus species were identified only in bats (Cui et al., 2019). Taken together, bats are most likely natural reservoirs of CoVs.

Bats are the second largest order of mammals after rodents, widely inhabiting all continents except Antarctica (Gentles et al., 2020), accounting for approximately one-third of CoV sequences before Coronavirus Disease 2019 (COVID-19) (Ruiz-Aravena et al., 2022). The natural reservoir role of bats for CoVs is attributed to host/virus co-existence in an equilibrium pattern, which is also interpreted as the virus adapting to the host, enabling effective bat infection and inter-bat transmission but with limited pathogenicity (Li et al., 2020, 2022) due to several factors. First, bats exhibit extraordinary immune tolerance, which maintains a moderate immune response to invading viruses such as CoVs, leading to limited viral replication and asymptomatic or mild CoV infections (Baker et al., 2013; Olival et al., 2017; Banerjee et al., 2018; Skirmuntt et al., 2020; Sia et al., 2022). Second, bats have a high body temperature, which resembles other mammals' febrile responses and infection immune responses, helping to keep virus infections at a tolerable level (O'Shea et al., 2014). Third, factors such as the large and closely aggregated population, sustained flight capability, and extreme roosting proximity of bats support the widespread and sustained existence of CoVs within the bat population (Maganga et al., 2014; Olival et al., 2017; Roes, 2020). Thus, the sustained infection and transmission in bats provide CoVs with a high probability of accumulating mutations, leading to variants with marginal adaptation to other mammalian hosts and causing spillover infections in humans and other mammals.

Numerous bat CoVs have been isolated and sequenced in recent years. A total of 78% (2,209/2,820) of the recorded CoV sequences in NCBI were uploaded since 2015, before the COVID-19 pandemic (https://www.ncbi.nlm.nih.gov/nuccore). However, it is challenging to assess the risk of new bat CoV isolates that cause infection or pandemics in human or other mammalian populations (Seyran et al., 2021). Traditional phylogenetic analysis can sufficiently evaluate the cross-species infection risk or any bat CoV (Lima et al., 2013; Seyran et al., 2021). More recently, machine learning or deep learning approaches based on big sequencing data have led to remarkable predictions of the host adaptation (Li et al., 2022; Nan et al., 2022), evolution (Hie et al., 2021), transmissibility (Fischhoff et al., 2021), virus–host interaction (Dey et al., 2020), and pathogenicity (Gussow et al., 2020) of SARS-CoV-2 and other viruses (Li et al., 2020). Host-specific compositional features in the virus genome have been indicated by the representation traits, such as dinucleotides (DNTs) (Li et al., 2020), DNT composition representation (DCR) (Li et al., 2022), and Uniform Manifold Approximation and Projection (UMAP) (Hie et al., 2021). Unfortunately, there is no pipeline or framework available to predict the adaptation of recorded or newly isolated bat CoVs to main mammalian hosts, such as Primates, Rodents, Artiodactyla, Suiformes, or Carnivora.

The present study aimed to represent the genome composition of the two main genes, i.e., *Spike*, the receptor binding glycoprotein, and *ORF1ab*, the RNA-dependent RNA polymerase complex, and then to predict the adaptive host of recorded or newly isolated CoVs. In this study, the viral genome representation and adaptive host prediction framework provide an intelligent approach to assessing the risk of cross-species infection and transmission for bat CoVs.

2. Methods

2.1. Data preprocessing and genomic compositional trait parsing of ssRNA viruses

Full genome sequences of coronaviruses were downloaded from the NCBI nucleotide database (https://www.ncbi.nlm.nih.gov/nuccore) and cleaned by removing records with multiple imprecise nucleotides or filtering with sequence length thresholds of 27,000 and 32,000 bp. Six types of adaptive host labels, including Chiroptera (CHI), Artiodactyla (ART), Suiformes (SUI), Rodents/Lagomorphs (Rodents, ROD), Carnivora (CAR), and Primates (PRI), were extracted from the "FEATURES"-"source"-"host" of each sequence record in the genebank sequence files and were manually checked one by one according to the host family (genus for porcine CoVs). The coding sequences of the two main CoV genes, *ORF1ab* and *Spike,* were parsed with the Biopython Python package. Genomic nucleotide composition traits of mononucleotide (NT), dinucleotide (DNT), DNT composition representation (DCR), trinucleotide (codon), codon pair, and amino acid (AA) were counted as a frequency value for each *ORF1ab* or *spike* sequence sample with a nucleotide counting script. The traits of NT, DNT, and DCR were counted as codon nucleotide-dependent sequences (Li et al., 2022). In sum, 12, 48, 1,536, 64, 3,721, and 20 features of the aforementioned six types of compositional traits were counted and utilized for genome composition analysis.

2.2. Clustering in genomic composition traits of coronaviruses

To visualize data distribution and clustering, dimension reduction was performed using Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE) for the full-dimension features of 12 NTs, 48 DNTs, 1,536 DCRs, 64 codons, 3,721 codon pairs, or 20 Aas for *ORF1ab* or *Spike*. PCA and t-SNE were performed using sklearn.decomposition.PCA (Jolliffe and Cadima, 2016) and sklearn.manifold.TSNE (https://scikit-learn.org/stable/about.html#citing-scikit-learn), respectively. Two main components (PCA1 and PCA2, or t-SNE1 and t-SNE2) were plotted with a host label for each data point using the Python Seaborn package. An unsupervised machine learning approach based on hierarchical clustering was used to observe the clustering and homology of CoVs with various adaptation host labels based on full-dimension features of each compositional trait. Euclidean distance was utilized as a hierarchical clustering scalar, and hierarchical clustering was performed using the sns.clustermap package. Additionally, to balance the biased sample number of CoVs with the six host labels, random down- and up-sampling were performed using the imblearn.over_sampling.SMOTE package before dimension reduction and visualization.

2.3. Building and training of deep learning predictors for adaptive hosts

To predict the adaptation of bat CoVs to other mammalian hosts, a deep learning predictor of convolutional neural networks (CNN) (Li et al., 2022) was built based on the 1,536 DCR features and five host labels (ART, SUI, ROD, CAR, and PRI). Five adaptive hosts were labeled as {0: 'SUI,' 1:'ART', 2:'ROD', 3:'CAR',4: 'PRI'}, respectively. Two packages, pandas.DataFrame.sample and imblearn.over_sampling.SMOTE, were utilized to perform down- and up-sampling to maintain the sample number balance of various host-originated CoVs. Two CNN models were built, one for *ORF1ab* and another for *Spike*. sklearn.model_selection import train_test_split was utilized for random training/test data splitting with a test data size of 25%. All bat CoV samples, either for *ORF1ab* or *Spike,* were not included in either the training or test data sets to avoid data leaks and were only utilized for the adaptive host prediction with trained models. The 1,536 DCR features of *ORF1ab* or *Spike* sequences were reshaped into an array of (6, 16, 16) for a 3D-CNN model of three convolutional layers. Out-channels of (8, 16, 32), a stride of (1, 1, 1), a padding of (0,1,1), and a kernel_size of (1, 3, 3) were set for the three layers of CNN. ReLU activation and average pooling were followed for each CNN layer. Two linear transformations were performed into the 192- and 5-dimensions, respectively, from the 768- and 192-dimensions of a fully connected layer. The sigmoid activation function was utilized for the 192-dimensions of the full-connected layer after one time of linear transformation to output prediction, and the Softmax function was utilized to output the prediction probability. A learning rate of 0.001 and a training epoch of 50 were set uniformly for the *ORF1ab* or *Spike* 3D-CNN model.

2.4. Evaluating the deep learning predictor

To evaluate the predictor's performance, the prediction of adaptive hosts and the adaptation probability for each host label were the outputs for each model. The confusion matrix (Townsend, 1971) and micro-average receiver operating characteristic (ROC) (Fawcett, 2005) with AUCs were plotted. A pair plot of the PCA-reduced, fully connected layer data (768-dimension) was performed with the two components to visualize the separation or clustering of the CoVs from different or the same host(s). The PCA1 value was also plotted and compared between/among these CoVs. Statistical significance in the PCA1 value of the PCA-reduced fully connected data was analyzed using an unpaired, non-parametric Mann–Whitney test, based on the hypothesis of non-Gaussian data distribution using GraphPad Prism 9.

2.5. Predict adaptive hosts for bat coronaviruses via the deep learning predictor

To evaluate the adaptive host(s) of bat CoVs, each of the bat CoV samples was predicted using the trained *ORF1ab* or *Spike* 3D-CNN model based on 1,536 *ORF1ab* or *Spike* DCR data. The adaptation and adaptation probability were output for each of the five hosts (ART, SUI, ROD, CAR, and PRI). The probability vector (five probability values) of all bat CoVs and the CoVs from other mammalian hosts were reduced to two main values by PCA and plotted, with each data point labeled with its host type or virus name.

3. Results

3.1. The architecture of genomic parsing and adaptation for predicting bat CoVs

To represent viral genome composition, full-length CoV sequences were selected and labeled with each of the six adaptive hosts (ART, CHI, CAR, ROD, PRI, and SUI). Complete open reading frames (ORFs) of *Spike* and *ORF1ab* were parsed for adaptation analysis. Six types of codon-dependent compositional traits of mononucleotides (NTs, $N_{\mathrm{NTdimension}} = 12$), amino acids (AAs, $N_{\mathrm{AAdimension}} = 20$), DNTs ($N_{\mathrm{DNTdimension}} = 48$), codons ($N_{\mathrm{codondimension}} = 64$), DCR ($N_{\mathrm{DCRdimension}} = 1{,}536$), and codon pairs (codonpairs, $N_{\mathrm{codonpairdimension}} = 3{,}721$) were embedded for *Spike* and *ORF1ab,* respectively, with previously reported approaches (Li et al., 2022) (Figure 1A). Unsupervised machine learning methods such as t-SNE, PCA, and hierarchical clustering were performed to visualize the separation and clustering of CoVs based on their abovementioned traits (Figure 1B). A DCR-based CNN (Li et al., 2022) was utilized to classify the CoVs based on each of the five adaptive host labels (Figure 1C). Finally, the adaptive host was predicted for bat CoVs, which were recorded in the database (Figure 1D) or were newly isolated and sequenced CoV strains (Figure 1E).

3.2. Representation and visualization of DCR and other compositional traits for CoVs

Dimension reduction was performed using t-SNE or PCA into two main components for each trait type of the CoVs. We then used Synthetic Minority Over-sampling Technique (SMOTE) to

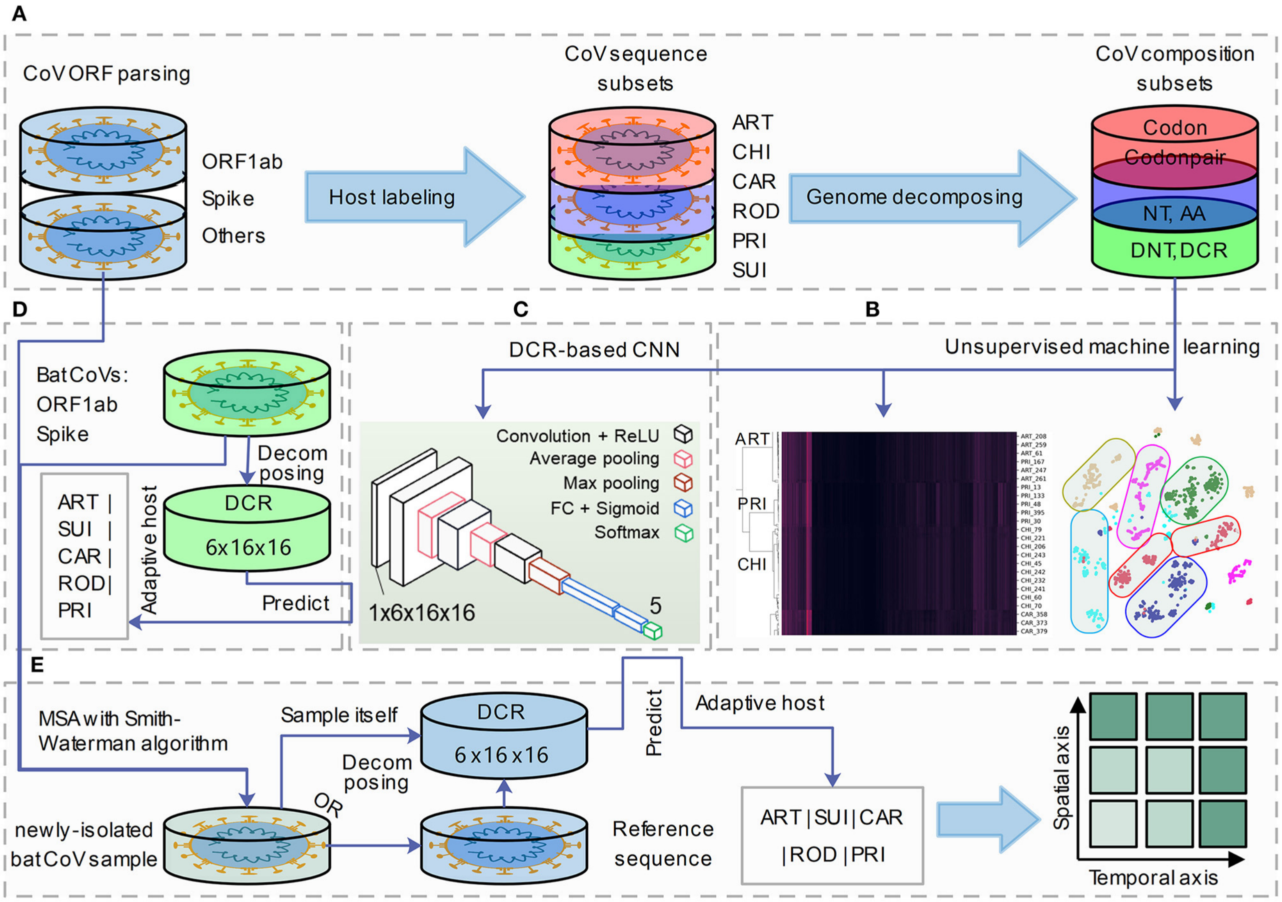


FIGURE 1
A pipeline for genome decomposition and adaptive host prediction of coronaviruses. The study was designed as the following pipeline of four sections. **(A)** Coronavirus (CoV) open reading frame (ORF) parsing (*ORF1ab* and *Spike*), adaptive host labeling (ART, Artiodactyla; CHI, Chiroptera; CAR, Carnivora; ROD, Rodents and Lagomorphs; PRI, Primates; SUI, Suiformes) and genome decomposition (into six types of NT, nucleotide; AA, amino acid; DNT, dinucleotide, codon; DCR, DNT compositional representation and codon pair); **(B, C)** unsupervised machine learning analysis (t-distributed stochastic neighbor embedding: t-SNE, Principal Component Analysis: PCA and hierarchical clustering) **(B)** and supervised deep learning (DCR-based Convolutional Neural Networks, CNN) prediction building of CoV genes; **(D)** field coronavirus sample collection, sequencing and Multiple Sequence Aligning (MSA); **(E)** host adaptation prediction of novel isolated CoV strains based on their decomposed DCR or based on their reference CoV sequences.

correct the data imbalance among host labels by up- and down-sampling. Given the high importance of SARS-CoV-2-related pangolin CoVs, we also added pangolin CoV data for unsupervised learning analysis. In the ORF1ab DNT trait, we observed a clear separation among CoVs with the five host labels in the two reduced t-SNE components (upper part, Figure 2A) and a much more diffuse distribution in the two reduced PCA components (lower part, Figure 2A). The intra-host clustering and the inter-host separation were also indicated using the hierarchical clustering of *ORF1ab* DNTs (Figure 2B). Similar clustering and separation of *ORF1ab* DCR were also observed post-t-SNE/PCA reduction and using hierarchical clustering (Figures 2C, D). The *Spike in* DNT and DCR also indicated intra-host clustering and inter-host separation in both DNT and DCR traits using the three types of unsupervised machine learning methods (Figures 2E–H). Interestingly, the pangolin CoVs were closely clustered with PRI CoVs, either for the reduced DNT or DCR features of *ORF1ab* (Figures 2A–D) of *Spike*. Moreover, the compositional traits of AAs and NTs for *ORF1ab* (Supplementary Figures S1a–f) and *Spike* (Supplementary Figures S1g–l) and the compositional traits of codons and codonpairs for *ORF1ab* (Supplementary Figures S2a–f) and *Spike* (Supplementary Figures S2g–l) were also observed. Additionally, some obvious disseminated distribution for ROD or CAR samples was mainly enlarged for abnormally disseminated samples using SMOTE sampling; the wide distribution of CHI samples had no association with data sampling and probably implied the wide host adaptation of CHI CoVs. Taking these results together, there was a host specificity in DCR and other compositional traits for the *ORF1ab* and *Spike* of CoVs.

Additionally, the other three genes, *E, M,* and *N,* were analyzed for the abovementioned six types of compositional traits. The severe mixture was observed for each type of trait in the two-dimensional space of t-SNE1 and t-SNE2 or of PCA1 and PCA2 (in order of amino acid, NTS, DNTS, DCR, codons, and codonpairs, respectively, for a-f, Supplementary Figures S3–S5).

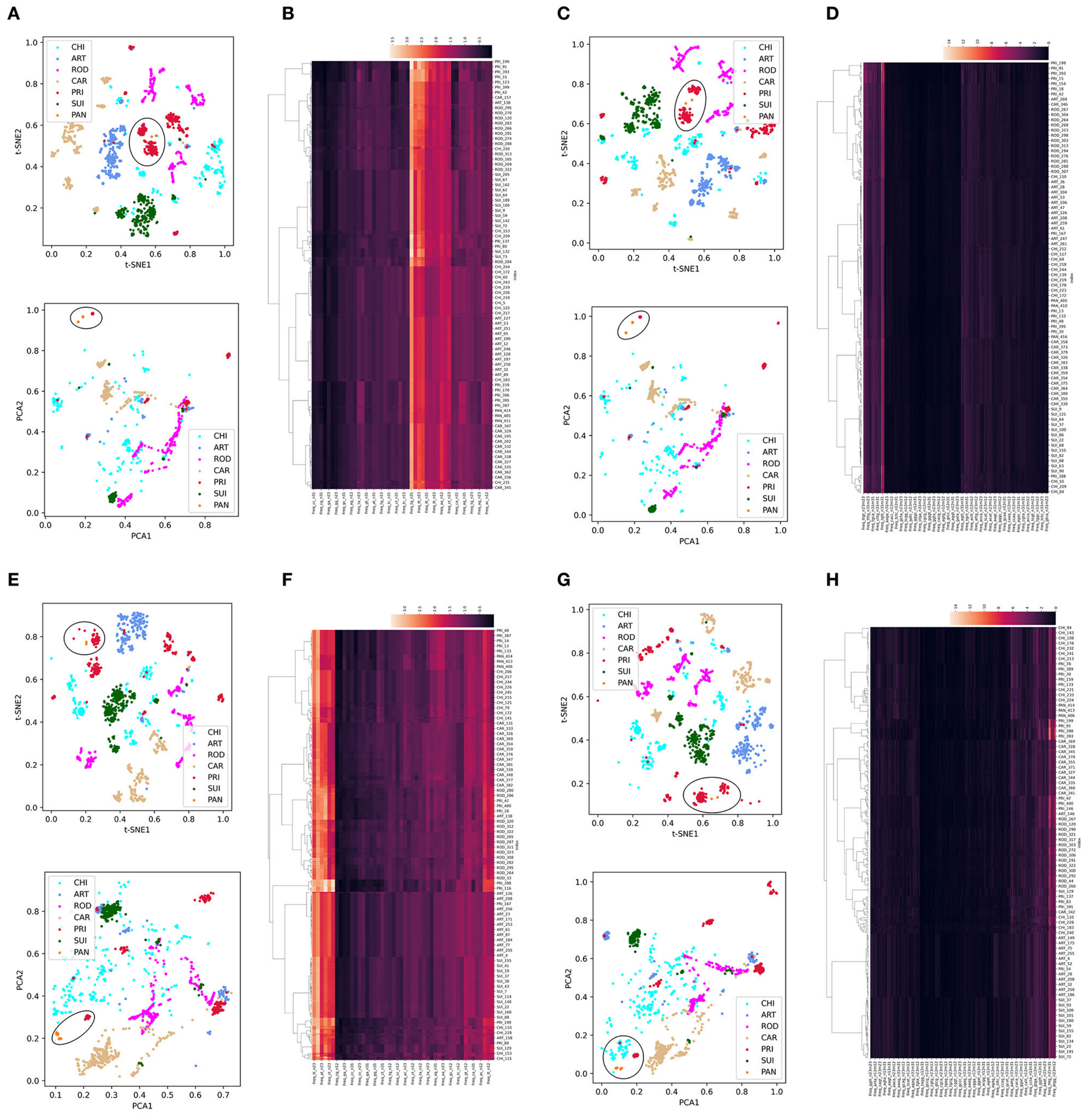


FIGURE 2
Distribution and clustering analysis of CoVs with host labels using unsupervised machine learning methods based on DNTs and DCR. Distribution of the two main components of compositional DNTs with t-SNE [**(A)** up] or PCA [**(A)** down] reduction and hierarchical clustering of 48 compositional DNTs of CoV *ORF1ab* **(B)**; **(C, D)** distribution of the two main t-SNE [**(C)** up] and PCA [**(C)** down] DCR components of *ORF1ab* and hierarchical clustering of all 1,536 DCR features **(D)**; **(E–H)** Similar unsupervised machine learning analysis of DNTs **(E, F)** and DCR features **(G, H)** of *Spike*.

3.3. Performance of the DCR-based CNN model to predict adaptive hosts CoVs

A deep learning model of CNN was built to predict the adaptation of bat CoVs to various types of mammalian hosts. The classification model with five labels (ART, CAR, ROD, PRI, and SUI) was trained using the 1,536-dimension DCR features of either *ORF1ab* or *Spike*. A training epoch-dependent performance elevation was observed for the classification of valid data based on DCR features of *ORF1ab* according to the confusion matrix (for epochs 10, 30, and 50, respectively, in Figures 3A–C; or for epochs 10–50, respectively, in Supplementary Figures S6a–e) or area under the receiver operating characteristic curve (ROC_AUC) (Figures 3D–F; Supplementary Figures S6f–j). Another model based on *spike* DCR features was also trained for the classification of CoV adaptive hosts. A high prediction accuracy was also obtained post-50-epoch training, as indicated by the confusion matrix (higher than 97% for epoch 50, Figures 3G–I; Supplementary Figures S6k–o) or ROC_AUC (Figures 3J–L; Supplementary Figures S6p–t). The training loss for either classifier descended quickly within the first 10 epochs and reached a plateau at approximately 20 epochs (respectively for *ORF1ab* and *Spike* Figures 3M, N; Supplementary Figures S6p–t).

To interpret the two trained classifiers, the reduction of the model's full-connected layer with PCA was visualized by plotting each pair of PCA1/PCA2 and PCA2/PCA1. The plotting results demonstrated that there was a sequential distribution of SUI, ART ROD, CAR, and PRI for the *ORF1ab* samples with five host labels for epochs 10, 30, and 50 (Figures 4A–C) or for 10–50 epochs (Supplementary Figures S7a–e). A significant separation of PRI (from other mammalian hosts) CoV samples was also observed from the distribution of the trained full-connected layer of *spike* DCR for epochs 10, 30, and 50 (Figures 4D–F) or for 10–50 epochs (Supplementary Figures S7f, g), with a different sequence of CAR, ART, SUI, and ROD for the other four host labels. The statistical analysis of the PCA1 values for each group indicated a significant difference between each neighboring pair of hosts in the *ORF1ab* samples ($P < 0.01$, except for ART vs. ROD with $P > 0.5$, Figure 4G). The difference was also significant for the neighboring ART/SUI or ROD/PRI *ORF1ab* samples ($P < 0.01$, Figure 4H).

3.4. DCR-based CNN predicts asymptotic bat-to-human adaptation of bat CoVs

To assess the adaptation of bat CoVs to other mammalian hosts, bat CoV sequences were fed to the two trained classifiers for *ORF1ab* and *Spike*. The results showed that 53% of CoV *ORF1ab* sequences were predicted as ART adaptive, while the percentages of adaptive samples for SUI, PRI, CAR, and ROD were 26, 11, 5, and 4%, respectively (Figure 5A). The average standardized probability of the predicted five groups of ART, PRI, SUI, CAR, and ROD were 0.640, 0.477, 0.276, 0.085, and 0.042, respectively (Figure 5B). The second classifier predicted almost the same percentage of Spike-adapted CV for ART hosts (54%). The percentages of adaptive samples for the other four types of hosts were 7, 12, 22, and 5%, respectively (Figure 5C), with an average standardized probability of 0.623, 0.451, 0.081, 0.456, and 0.048 for the five groups (Figure 5D). To further assess the distribution of bat CoVs and other mammalian CoVs in the adaptation space of mammalian hosts, five probability values for the five hosts were taken as a vector for each sample and were reduced to two main components with PCA. Interestingly, except for CoVs with an SUI host label, other mammalian but bat *ORF1ab* samples were almost linearly distributed, with ART samples on the lower left, CAR, and ROD samples in the middle, and PRI samples mainly on the upper right (Figure 5E), indicating a linear asymptotic adaptation shift from ART to CAR/ROD and then to PRI. Particularly, there was a linear-like distribution of all human CoV or human CoV-related ORF1ab samples in the two-dimensional space. MERS/bat MERS-related CoVs, SARS/bat SARS-like CoVs, and human CoVs of OC43, 229E, and others were successively distributed from the lower left to the upper right (Figure 5E). Similar linear asymptotic adaptation shifts of CoV *spike* samples were also observed (Figure 5F). Additionally, bat CoVs were disseminated in the adaptation space, with varied distances in PCA1 or PCA2 values for each of the five groups of CoVs (Figures 5E, F). Taken together, the two adaptation classifiers predicted a unanimous linear asymptotic adaptation shift from the ART host to humans.

4. Discussion

The present study aimed to predict the potential for viruses, such as influenza A viruses and coronaviruses, to cause infection and transmission in the human population.

Thus, we defined it as "the capability to infect humans easily, to transmit among populations efficiently, and to be virulent to some degree to humans" previously (Li et al., 2020, 2022). Genomic traits for virus adaptation have been biologically interpreted as shaping viral mRNA decay (Contu et al., 2021), methylation (Upadhyay et al., 2013), translation (Chen et al., 2020), replication efficiency (Forsberg, 2003; Bahir et al., 2009; Li et al., 2011), and antagonizing host anti-virus immune response (Xia, 2020), all of which reflect viral adaptive phenotypic traits to their hosts. Moreover, such adaptive genotypes were distinguishable and predictable with machine learning or deep learning approaches. Adaptation phenotypes of viruses to bats and other mammals are supported by parallel viral genotypes. A coarse-grained representation of the viral genome as compositional traits, such as DNT and DCR, is host-specific and predictable with machine learning or deep learning approaches for CoVs (Pollock et al., 2020; Li et al., 2022; Nan et al., 2022), influenza viruses (Taubenberger and Kash, 2010; Li et al., 2020), and other viruses (Bahir et al., 2009; Babayan et al., 2018; Chen et al., 2020). Fine-tuned sequential representation has been indicated to be sensitive to predicting the adaptation of SARS-CoV-2 Omicron sublineages with deep learning (Nan et al., 2022). In the present study, representative compositional traits of DCR and others confirmed the intra-host clustering and inter-host separability of various host-specific CoVs. Interestingly, there was a disseminated distribution of bat (CHI) CoVs into the areas of the CoVs with other host labels, indicating multiple adaptations to other hosts of bat CoVs. Additionally, the dispersed distribution of ROD samples was mainly caused by SMOTE up-sampling. Pangolin has been shown to play an intermediate role in the cross-species infection of SARS-CoV-2 viruses (Lam et al., 2020; Xiao et al., 2020) or MERS-CoV (Chen et al., 2020). The compositional traits indicated a close clustering of these pangolin CoVs with human CoVs, either for *ORF1ab* or *Spike* genes, implying a human adaptation. However, we did not set pangolin as an independent host label for supervised learning due to the small sample size of the whole genome and also due to the too-close clustering of pangolin viruses to human CoVs. Multiple genes other than *ORF1ab* and *Spike* might mediate the adaptation of CoVs to human and other mammalian hosts. However, the three other genes, *E*, *M*, and *N*, were mixed for CoVs of various host labels, suggesting less host specificity.

In the present study, the deep learning classifier with five host labels (ART, CAR, ROD, SUI, and PRI) targeting either the *ORF1ab* or *Spike* gene, accurately predicted the host of the five groups of CoVs. A complete landscape of mammalian CoV

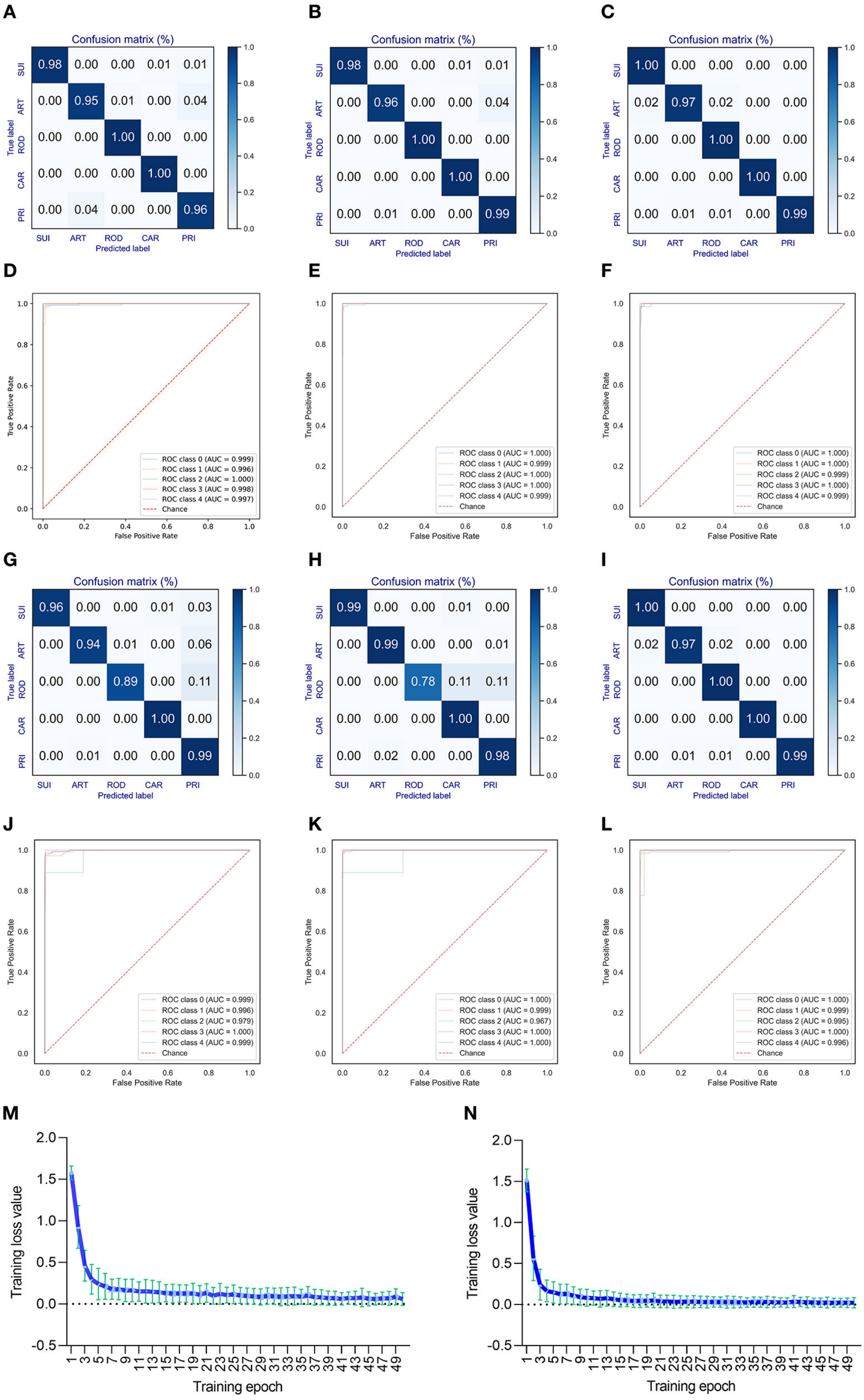


FIGURE 3
Performance of the DCR-based deep learning approach for predicting adaptive hosts of bat coronaviruses. The performance of the DCR-based deep learning predictor was evaluated with a confusion matrix [**(A–C)** for training epochs of 10, 30, and 50, respectively] and receiver operating characteristic ROC curve () [**(D–F)** for training epochs of 10, 30, and 50, respectively] for CoV *ORF1ab*; a similar evaluation was performed with a confusion matrix [**(G–I)**, respectively] and ROC [**(J–L)**, respectively] for CoV *Spike*. **(M, N)** Curve of the average training loss for validated data for the predictors for *ORF1ab* **(M)** and *Spike* **(N)**. ART, Artiodactyla; SUI, Suiformes; ROD, Rodents and Lagomorphs; CAR, Carnivora; PRI, Primates.

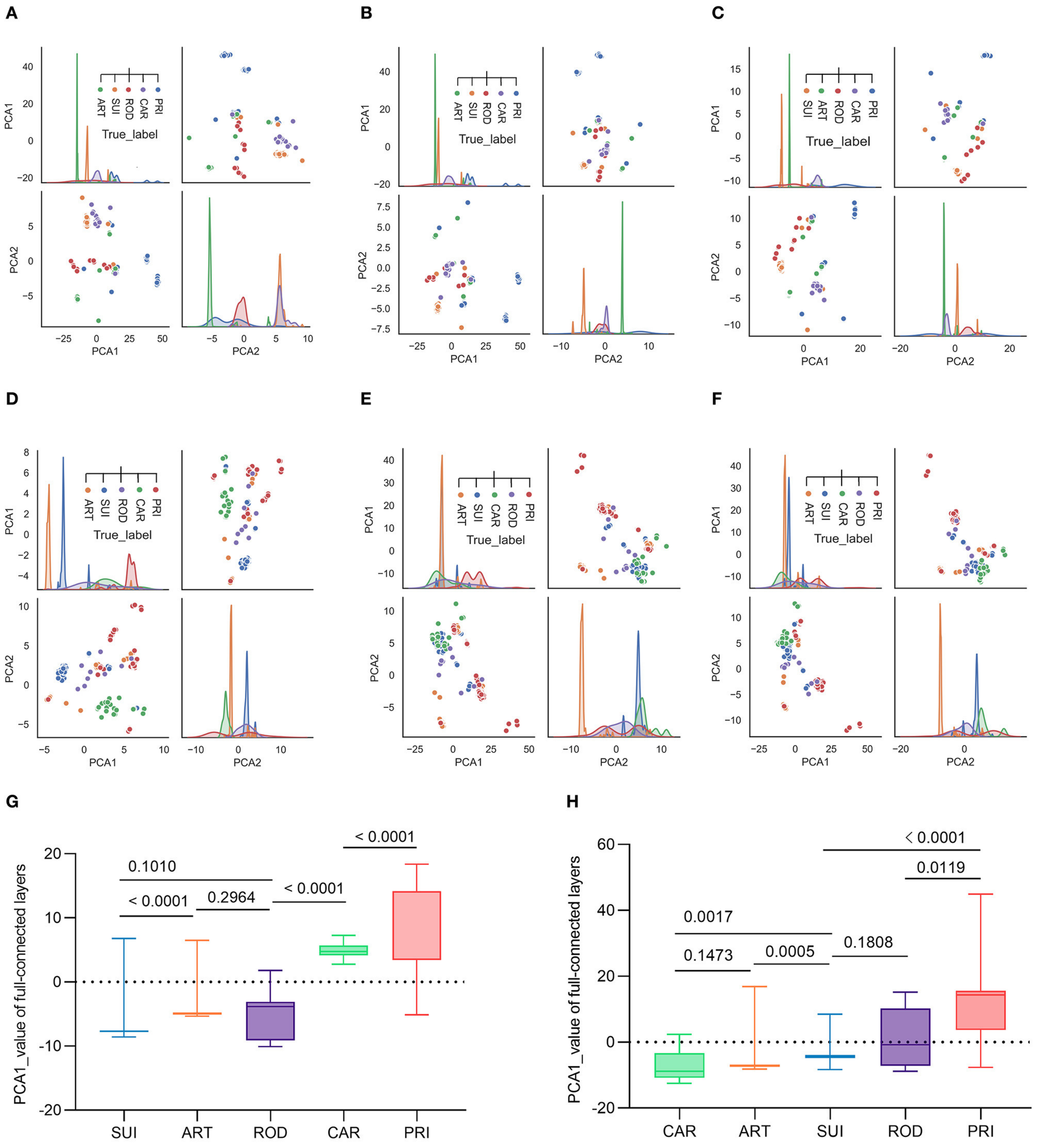


FIGURE 4
Visualization of the complete connection layer post-deep learning training of CoV *ORF1ab* and *Spike*. Full connection layers after three rounds of convolution were reduced by PCA into two principal components and plotted to visualize the distribution of CoV samples with the five host labels. Pair plots of reduced PCA1 and PCA2, post-training epochs of 10, 30, and 50, were plotted for *ORF1ab* **(A–C)** or *Spike* **(D–F)**, respectively. **(G, H)** Statistics of the reduced PCA1 value post-training epoch of 50 [**(G)** for *ORF1ab* and **(H)** for *Spike*].

samples in the predicted adaptation space constructed by the adaptation probability for the five hosts (Figure 5) unanimously showed a clearer intra-host clustering and inter-host separability of all CoV samples than the distribution of the original DCR features. Interestingly, a linear-like distribution of the CoV samples, except for the SUI CoVs, was observed in the adaptation space, suggesting CoV's asymptotic adaptation from ART to CAR/ROD and then to PRI hosts. Taking these results together, we proposed a possible niche distance-related landscape of host adaptation for bat CoVs (Figure 6): a dominant adaptation to the ART hosts, followed by

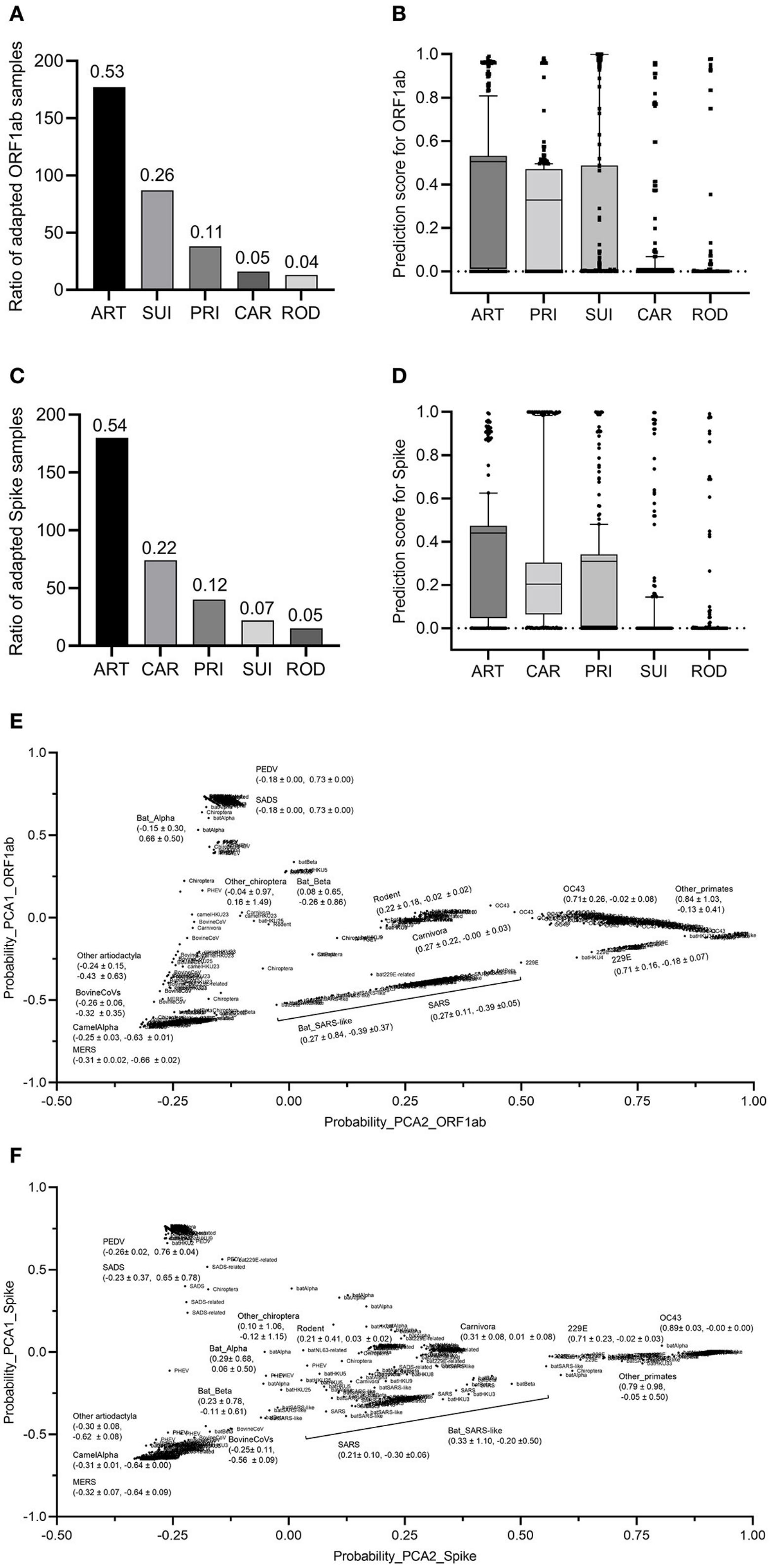


FIGURE 5

Prediction and prediction probabilities of bats and other coronaviruses. **(A, B)**: Prediction **(A)** and prediction probability **(B)** of *ORF1ab* for the adaptation to the five mammalian hosts of bat coronaviruses; **(C, D)** prediction **(C)** and prediction probability **(D)** of *Spike* for the adaptation to the five mammalian hosts of bat coronaviruses; **(E, F)** visualization of the PCA-reduced prediction probability of *ORF1ab* **(E)** and *Spike* **(F)** for bat and other coronaviruses.

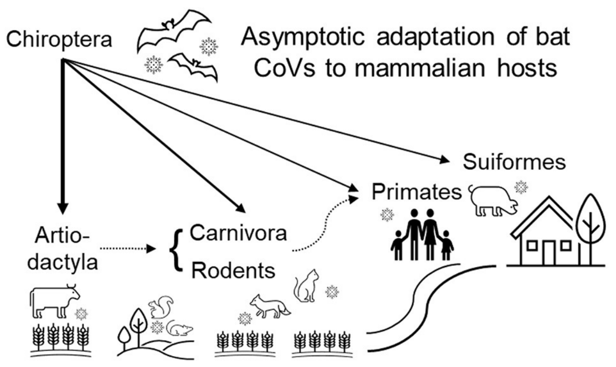


FIGURE 6
Summary diagram of adaptive hosts predicted by DCR-based deep learning. Adaption of Chiroptera CoVs to sequential ecological niches of ART/SUI, ROD/CAR, and then PRI are shown. ART, Artiodactyla; CAR, Carnivora; CHI, Chiroptera; PRI, Primates; ROD, Rodents and Lagomorphs; SUI, Suiformes.

a relatively less adaptation to CAR/ROD hosts, and finally to PRI hosts. Such asymptotic adaptation to the bat's close and far niche distances (Corman et al., 2018) reconfirmed the mediation of these natural hosts in the adaptation shift of bat CoVs to human beings. The ranked adaptation for bat CoVs provides more clues that CoVs might shift more probably from ART to a CAR/ROD host and then to humans than directly from CHI hosts, considering the closer niche distance between humans and these mediator hosts.

Additionally, the domestic pig in the SUI host type is the key mediator for the adaptation shift to a human host for the other major respiratory infectious agent, influenza A viruses (Neumann et al., 2009); however, the results in the present study indicated a significantly independent distribution of SUI CoVs from the linear-like and asymptotic distribution of the CoVs from other mammals. CoVs have been reported to cause infection and transmission in domestic pigs worldwide, such as porcine transmissible gastroenteritis virus (TGEV) (Brian and Baric, 2005), porcine enteric diarrhea virus (PEDV) (Lin et al., 2016), and swine acute diarrhea syndrome (SADS) CoV (Zhou et al., 2018). SUI CoVs are not likely to cause transmission in the human population, although the porcine delta coronavirus has been reported to infect malnourished Haitian children (Lednicky et al., 2021). SUI CoVs did not cause cross-species transmission in humans, as they were not closely related to human CoVs in the adaptation space predicted in this study. Therefore, we speculate that the risk of SUI CoVs threatening human populations is lower. However, it is important to note that overfitting can occur in machine learning or deep learning models to varying degrees. Additionally, there is a significant bias, with a smaller number of ROD CoVs and a much larger number of SUI or ART CoVs. The use of up-sampling for ROD CoVs and down-sampling for SUI and ART CoVs may lead to overfitting of the model and potentially explain the wide range of predicted adaptation probabilities (Figure 4).

In summary, the genomic dinucleotides represented as DCR indicate a host-specific separation and clustering that can predict a linear and asymptotic adaptation shift of bat CoVs from other mammals to humans through deep learning techniques.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Author contributions

JL, J-FJ, TJ, and YT conceived the study. JL, FT, and SZ contributed to the acquisition and interpretation of data. S-SL, X-PK, J-QW, and WL performed data cleaning and statistical analysis. JL, SZ, ZL, Y-DL, and YF performed genome parsing and unsupervised and supervised learning with the assistance of J-FJ, TJ, and YT. JL drafted the manuscript, coded all scripts for genome parsing, deep learning, and data visualization. All authors contributed to the critical revision of the manuscript for important intellectual content.

Funding

This study was supported by grants from the National Key Research and Development Program of China (Grant Nos. 2021YFC2302004, 2021YFC0863400, 2019YFC1200501, and 2018YFA0903000) and the National Natural Science Foundation of China (Grant No. 32070166).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1157608/full#supplementary-material

References

Babayan, S. A., Orton, R. J., and Streicker, D. G. (2018). Predicting reservoir hosts and arthropod vectors from evolutionary signatures in RNA virus genomes. *Science* 362, 577–580. doi: 10.1126/science.aap9072

Bahir, I., Fromer, M., Prat, Y., and Linial, M. (2009). Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Mol. Syst. Biol.* 5, 311. doi: 10.1038/msb.2009.71

Baker, M. L., Schountz, T., and Wang, L. F. (2013). Antiviral immune responses of bats: a review. *Zoonoses Public Health* 60, 104–116. doi: 10.1111/j.1863-2378.2012.01528.x

Banerjee, A., Misra, V., Schountz, T., and Baker, M. L. (2018). Tools to study pathogen-host interactions in bats. *Virus Res.* 248, 5–12. doi: 10.1016/j.virusres.2018.02.013

Brian, D. A., and Baric, R. S. (2005). Coronavirus genome structure and replication. *Curr. Top Microbiol. Immunol.* 287, 1–30. doi: 10.1007/3-540-26765-4_1

Chen, F., Wu, P., Deng, S., Zhang, H., Hou, Y., Hu, Z., et al. (2020). Dissimilation of synonymous codon usage bias in virus-host coevolution due to translational selection. *Nat. Ecol. Evol.* 4, 589–600. doi: 10.1038/s41559-020-1124-7

Contu, L., Balistreri, G., Domanski, M., Uldry, A., and Mühlemann, O. (2021). Characterisation of the Semliki Forest Virus-host cell interactome reveals the viral capsid protein as an inhibitor of nonsense-mediated mRNA decay. *PLoS Pathog.* 17, e1009603. doi: 10.1371/journal.ppat.1009603

Corman, V. M., Muth, D., Niemeyer, D., and Drosten, C. (2018). Hosts and sources of endemic human coronaviruses. *Adv. Virus Res.* 100, 163–188. doi: 10.1016/bs.aivir.2018.01.001

Cui, J., Li, F., and Shi, Z. (2019). Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* 17, 181–192. doi: 10.1038/s41579-018-0118-9

Deng, Y., Li, C., Han, J., Wen, Y., Wang, J., Hong, W., et al. (2017). Phylogenetic and genetic characterization of a 2017 clinical isolate of H7N9 virus in Guangzhou, China during the fifth epidemic wave. *Sci. China Life Sci.* 60, 1331–1339. doi: 10.1007/s11427-017-9152-1

Dey, L., Chakraborty, S., and Mukhopadhyay, A. (2020). Machine learning techniques for sequence-based prediction of viral-host interactions between SARS-CoV-2 and human proteins. *Biomed. J.* 43, 438–450. doi: 10.1016/j.bj.2020.08.003

El-Sayed, A., and Kamel, M. (2021). Coronaviruses in humans and animals: the role of bats in viral evolution. *Environ. Sci. Pollut. Res. Int.* 28, 19589–19600. doi: 10.1007/s11356-021-12553-1

Fawcett, T. (2005). An introduction to ROC analysis. *Patt. Recogn. Lett.* 27, 861–874. doi: 10.1016/j.patrec.2005.10.010

Fischhoff, I. R., Castellanos, A. A., Rodrigues, J., Varsani, A., and Han, B. A. (2021). Predicting the zoonotic capacity of mammals to transmit SARS-CoV-2. *Proc. Biol. Sci.* 288, 20211651. doi: 10.1098/rspb.2021.1651

Forni, D., Cagliani, R., Clerici, M., and Sironi, M. (2017). Molecular evolution of human coronavirus genomes. *Trends Microbiol.* 25, 35–48. doi: 10.1016/j.tim.2016.09.001

Forsberg, R. (2003). A codon-based model of host-specific selection in parasites, with an application to the influenza a virus. *Mol. Biol. Evol.* 20, 1252–1259. doi: 10.1093/molbev/msg149

Gentles, A. D., Guth, S., Rozins, C., and Brook, C. E. (2020). A review of mechanistic models of viral dynamics in bat reservoirs for zoonotic disease. *Pathog. Glob Health* 114, 407–425. doi: 10.1080/20477724.2020.1833161

Gussow, A. B., Auslander, N., Faure, G., Wolf, Y. I., Zhang, F., and Koonin, E. V. (2020). Genomic determinants of pathogenicity in SARS-CoV-2 and other human coronaviruses. *Proc. Natl. Acad. Sci. U S A.* 117, 15193–15199. doi: 10.1073/pnas.2008176117

Hie, B., Zhong, E. D., Berger, B., and Bryson, B. (2021). Learning the language of viral evolution and escape. *Science* 371, 284–288. doi: 10.1126/science.abd7331

Ji, W., Peng, Q., Fang, X., Li, Z., Li, Y., Xu, C., et al. (2022). Structures of a deltacoronavirus spike protein bound to porcine and human receptors. *Nat. Commun.* 13, 1467. doi: 10.1038/s41467-022-29062-5

Jolliffe, I. T., and Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosop. Transac. R. Soc. A.* 374, 20150202. doi: 10.1098/rsta.2015.0202

Lam, T. T., Jia, N., Zhang, Y. W., Shum, M. H., Jiang, J. F., Zhu, H. C., et al. (2020). Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature.* 583, 282–285. doi: 10.1038/s41586-020-2169-0

Lednicky, J. A., Tagliamonte, M. S., White, S. K., Elbadry, M. A., Alam, M. M., Stephenson, C. J., et al. (2021). Independent infections of porcine deltacoronavirus among Haitian children. *Nature* 600, 133–137. doi: 10.1038/s41586-021-04111-z

Li, J., Liu, B., Chang, G., Hu, Y., Zhan, D., Xia, Y., et al. (2011). Virulence of H5N1 virus in mice attenuates after in vitro serial passages. *Virol. J.* 8, 93. doi: 10.1186/1743-422X-8-93

Li, J., Wu, Y., Zhang, S., Kang, X., and Jiang, T. (2022). Deep learning based on biologically interpretable genome representation predicts two types of human adaptation of SARS-CoV-2 variants. *Brief Bioinform.* 23, bbac036. doi: 10.1093/bib/bbac036

Li, J., Zhang, S., Li, B., Hu, Y., Kang, X., Wu, X., et al. (2020). Machine learning methods for predicting human-adaptive influenza a viruses based on viral nucleotide compositions. *Mol. Biol. Evol.* 37, 1224–1236. doi: 10.1093/molbev/msz276

Lima, F. E., Campos, F. S., Kunert, F. H., Batista, H. B., Carnielli, P. J., Cibulski, S. P., et al. (2013). Detection of Alphacoronavirus in velvety free-tailed bats (Molossus molossus) and Brazilian free-tailed bats (Tadarida brasiliensis) from urban area of Southern Brazil. *Virus Genes* 47, 164–167. doi: 10.1007/s11262-013-0899-x

Lin, C. M., Saif, L. J., Marthaler, D., and Wang, Q. (2016). Evolution, antigenicity and pathogenicity of global porcine epidemic diarrhea virus strains. *Virus Res.* 226, 20–39. doi: 10.1016/j.virusres.2016.05.023

Liu, K., Pan, X., Li, L., Yu, F., Zheng, A., Du, P., et al. (2021). Binding and molecular basis of the bat coronavirus RaTG13 virus to ACE2 in humans and other species. *Cell* 184, 3438–3451. doi: 10.1016/j.cell.2021.05.031

Liu, W., Fan, H., Raghwani, J., Lam, T. T., Li, J., Pybus, O. G., et al. (2014). Occurrence and reassortment of avian influenza A (H7N9) viruses derived from coinfected birds in China. *J. Virol.* 88, 13344–13351. doi: 10.1128/JVI.01777-14

Maganga, G. D., Bourgarel, M., Vallo, P., Dallo, T. D., Ngoagouni, C., Drexler, J. F., et al. (2014). Bat distribution size or shape as determinant of viral richness in african bats. *PLoS ONE* 9:e100172. doi: 10.1371/journal.pone.0100172

Nan, B. G., Zhang, S., Li, Y. C., Kang, X. P., Chen, Y. H., Li, L., et al. (2022). Convolutional neural networks based on sequential spike predict the high human adaptation of SARS-CoV-2 Omicron Variants. *Viruses* 14, 1072. doi: 10.3390/v14051072

Neumann, G., Noda, T., and Kawaoka, Y. (2009). Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature* 459, 931–939. doi: 10.1038/nature08157

Olival, K. J., Hosseini, P. R., Zambrana-Torrelio, C., Ross, N., Bogich, T. L., and Daszak, P. (2017). Host and viral traits predict zoonotic spillover from mammals. *Nature* 546, 646–650. doi: 10.1038/nature22975

O'Shea, T. J., Cryan, P. M., Cunningham, A. A., Fooks, A. R., Hayman, D. T., Luis, A. D., et al. (2014). Bat flight and zoonotic viruses. *Emerg. Infect. Dis.* 20, 741–745. doi: 10.3201/eid2005.130539

Pollock, D. D., Castoe, T. A., Perry, B. W., Lytras, S., Wade, K. J., Robertson, D. L., et al. (2020). Viral CpG deficiency provides no evidence that dogs were intermediate hosts for SARS-CoV-2. *Mol Biol Evol* 37, 2706–2710. doi: 10.1093/molbev/msaa178

Roes, F. L. (2020). On the evolution of virulent zoonotic viruses in bats. *Biol. Theory* 15, 223–225. doi: 10.1007/s13752-020-00363-6

Ruiz-Aravena, M., McKee, C., Gamble, A., Lunn, T., Morris, A., Snedden, C. E., et al. (2022). Ecology, evolution and spillover of coronaviruses from bats. *Nat. Rev. Microbiol.* 20, 299–314. doi: 10.1038/s41579-021-00652-2

Seyran, M., Hassan, S. S., Uversky, V. N., Pal, C. P., Uhal, B. D., Lundstrom, K., et al. (2021). Urgent Need for Field Surveys of Coronaviruses in Southeast Asia to Understand the SARS-CoV-2 Phylogeny and Risk Assessment for Future Outbreaks. *Biomolecules* 11, 398. doi: 10.3390/biom11030398

Sia, W. R., Zheng, Y., Han, F., Chen, S., Ma, S., Wang, L. F., et al. (2022). Exploring the role of innate lymphocytes in the immune system of bats and virus-host interactions. *Viruses* 14, 150. doi: 10.3390/v14010150

Skirmuntt, E. C., Escalera-Zamudio, M., Teeling, E. C., Smith, A., and Katzourakis, A. (2020). The potential role of endogenous viral elements in the evolution of bats as reservoirs for zoonotic viruses. *Annu. Rev. Virol.* 7, 103–119. doi: 10.1146/annurev-virology-092818-015613

Su, S., Wong, G., Shi, W., Liu, J., Lai, A., Zhou, J., et al. (2016). Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends Microbiol.* 24, 490–502. doi: 10.1016/j.tim.2016.03.003

Sun, H., Sun, Y., Pu, J., Zhang, Y., Zhu, Q., Li, J., et al. (2014). Comparative virus replication and host innate responses in human cells infected with three prevalent clades (2.3.4, 2.3.2, and 7) of highly pathogenic avian influenza H5N1 viruses. *J. Virol.* 88, 725–729. doi: 10.1128/JVI.02510-13

Taubenberger, J. K., and Kash, J. C. (2010). Influenza virus evolution, host adaptation, and pandemic formation. *Cell Host Microbe.* 7, 440–451. doi: 10.1016/j.chom.2010.05.009

Townsend, J. T. (1971). Theoretical analysis of an alphabetic confusion matrix. *Percept. Psychophy.* 9, 40–50. doi: 10.3758/BF03213026

Upadhyay, M., Samal, J., Kandpal, M., Vasaikar, S., Biswas, B., Gomes, J., et al. (2013). CpG dinucleotide frequencies reveal the role of host methylation capabilities in parvovirus evolution. *J. Virol.* 87, 13816–13824. doi: 10.1128/JVI.02515-13

Wang, Z., Huang, G., Huang, M., Dai, Q., Hu, Y., Zhou, J., et al. (2022). Global patterns of phylogenetic diversity and transmission of bat coronavirus. *Sci. China Life Sci.* 66, 861–874. doi: 10.1007/s11427-022-2221-5

West, R., Michie, S., Rubin, G. J., and Amlot, R. (2020). Applying principles of behaviour change to reduce SARS-CoV-2 transmission. *Nat. Hum. Behav.* 4, 451–459. doi: 10.1038/s41562-020-0887-9

WHO (2022). *Coronavirus (COVID-19) Dashboard*. Available online at: https://covid19.who.int/ (accessed June 30, 2022).

Woo, P. C., Lau, S. K., Lam, C. S., Lau, C. C., Tsang, A. K., Lau, J. H., et al. (2012). Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. *J. Virol.* 86, 3995–4008. doi: 10.1128/JVI.06540-11

Xia, X. (2020). Extreme genomic CpG deficiency in SARS-CoV-2 and evasion of host antiviral defense. *Mol. Biol. Evol.* 37, 2699–2705. doi: 10.1093/molbev/msaa094

Xiao, K., Zhai, J., Feng, Y., Zhou, N., Zhang, X., Zou, J. J., et al. (2020). Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature.* 583, 286–289. doi: 10.1038/s41586-020-2313-x

Zhou, P., Fan, H., Lan, T., Yang, X. L., Shi, W. F., Zhang, W., et al. (2018). Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin. *Nature* 556, 255–258. doi: 10.1038/s41586-018-0010-9

Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273. doi: 10.1038/s41586-020-2012-7

Frontiers in Microbiology

TYPE Original Research
PUBLISHED 18 May 2023
DOI 10.3389/fmicb.2023.1183633

OPEN ACCESS

EDITED BY
Wenshi Wang,
Xuzhou Medical University, China

REVIEWED BY
Jianhui Wang,
Yale University, United States
Petr Sláma,
Mendel University in Brno, Czechia

*CORRESPONDENCE
Chuan Wang
✉ wangchuan@scu.edu.cn
Wentao Zhu
✉ wentaozhu@126.com

†These authors have contributed equally to this work

RECEIVED 10 March 2023
ACCEPTED 28 April 2023
PUBLISHED 18 May 2023

CITATION
Wang X, Zhu X, Lin Y, He L, Yang J, Wang C and Zhu W (2023) Tracking the first SARS-CoV-2 Omicron BA.5.1.3 outbreak in China.
Front. Microbiol. 14:1183633.
doi: 10.3389/fmicb.2023.1183633

Tracking the first SARS-CoV-2 Omicron BA.5.1.3 outbreak in China

Xiaoxia Wang[1,2†], Xiong Zhu[2†], Yujin Lin[2], Lvfen He[2], Jing Yang[3], Chuan Wang[1*] and Wentao Zhu[4*]

[1]West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, Sichuan, China, [2]Central and Clinical Laboratory of Sanya People's Hospital, Sanya, Hainan, China, [3]State Key Laboratory of Infectious Disease Prevention and Control, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China, [4]Department of Infectious Diseases and Clinical Microbiology, Beijing Institute of Respiratory Medicine and Beijing Chao-Yang Hospital, Capital Medical University, Beijing, China

The SARS-CoV-2 is still undergoing rapid evolution, resulting in the emergence of several variants of concern, especially the Omicron variants (B.1.1.529), which are surging worldwide. In this study, we tracked Omicron subvariant BA.5.1.3 as the causative agent in the Hainan Province wave in China, which started on 1 August 2022. This was China's first case of Omicron subvariant BA.5.1.3 and led to an indefinite total lockdown in Hainan with more than 8,500 confirmed cases. We obtained 391 whole genomes from positive nasopharyngeal swab samples in the city of Sanya in Hainan Province, which was the center of this outbreak. More than half of the infected cases were female (58%, 227/391) with a median age of 37.0years (IQR 23.0–53.0). Median Ct values were 24.9 (IQR 22.6–27.3) and 25.2 (IQR 22.9–27.6) for *ORF1ab* and *N* genes, respectively. The total single-nucleotide polymorphism (SNP) numbers of Omicron BA.5.1.3 sampled in Sanya (median 69.0, IQR=69.0–70.0) compared to those worldwide (median 63.0, IQR=61.0–64.0) showed a significant difference ($p<0.05$). Unique core mutations, including three non-synonymous mutations in *ORF1ab* (Y1064N, S2844G, and R3574K) and one synonymous mutation in *ORF3a* (S74S), were found. Phylogenetic analysis showed that virus from Sanya formed an independent sub-clade within the BA.5.1.3 subvariant, and could be divided into 15 haplotypes based on the *S* gene. The most recent common ancestor for the virus from Sanya was estimated as appearing on 5 July 2022, with 95% HPD ranging from 15 May to 20 September 2022. Thanks to our results, we were also able to delineate the mutational profile of this outbreak and highlight the importance of global genomic surveillance and data sharing.

KEYWORDS

SARS-CoV-2, Omicron BA.5.1.3, outbreak, mutation, Sanya

Introduction

In December 2019, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first isolated from several workers who were suffering from pneumonia in China (Roychoudhury et al., 2020). It was declared as a global pandemic on 11 March 2020 by the World Health Organization (WHO), which is one of the greatest public health emergencies in human history (Roychoudhury et al., 2020). Globally, the COVID-19 pandemic has resulted in more than 628 million confirmed cases, with over 3.8 million deaths as of 2 November 2022.[1] The spike (S) protein, this virus's most important structural protein, plays a dominant role in the viral invasion (Lan et al., 2020; Jackson et al., 2022). The process of SARS-CoV-2 entry is initiated by the interaction between the receptor-binding domain (RBD) in the C-terminus of the S protein sub-segment S1 and angiotensin-converting enzyme 2 (ACE2) receptor on the host cell (Roychoudhury et al., 2020, 2021). This is followed by a conformational change of S protein initiated by type 2 transmembrane serine protease (TMPRSS2) on host cell surface, which allows the virus to enter the cell (Roychoudhury et al., 2021). Then, SARS-CoV-2 hijacks the cellular machinery for viral replication. As the virus continued to evolve, many mutations, especially those in the *S* gene, have gradually increased over time, leading to the emergence of several SARS-CoV-2 variants, and expanding our understanding of the impact of virus variants (WHO, 2022). Viral variants with mutations in the receptor-binding domain (RBD) within the *S* gene that can enhance viral transmissibility and virulence or affect vaccine effectiveness (immune escape potential) and diagnostic testing are designated as variants of concerns (VOCs) by the World Health Organization (WHO; Kumar et al., 2022). Thus far, SARS-CoV-2 has evolved into four previously circulating VOCs (Alpha, Beta, Delta, and Gamma) as well as one currently circulating VOC (Omicron) that has rapidly spread throughout the world (Shrestha et al., 2022; Wang et al., 2022).

The ongoing Omicron (B.1.1.529) variant was first detected in South Africa in mid-November 2021 (Araf et al., 2022; Saxena et al., 2022), and categorized as a VOC subsequently by WHO (Maxmen, 2022). It quickly spread to 80 countries within 3 weeks (Viana et al., 2022). Currently, SARS-CoV-2 Omicron is still the dominant variant circulating worldwide (in more than 170 countries and territories), accounting for >98% of SARS-CoV-2 genomes uploaded to the GISAID database since February 2022 (WHO, 2022). The Omicron variants have continued to evolve, resulting in Omicron subvariants that have been classified into several lineages, namely BA.1, BA.2, BA.3, BA.4, BA.5, and descendent lineages (Shrestha et al., 2022). The first three lineages (BA.1, BA.2, and BA.3) were identified almost simultaneously, and they are different not only from each other but also from previously circulating VOCs (Berkhout and Herrera-Carrillo, 2022; Shrestha et al., 2022). More recently, two additional lineages (BA.4 and BA.5) were first discovered in South Africa and have subsequently spread worldwide (Tegally et al., 2021; Maxmen, 2022). The most recent common ancestor of BA.4 and BA.5 has been estimated to appear in mid-November 2021 (Tegally et al., 2022b).

Omicron is the most divergent variant so far, due to its unprecedented set of non-synonymous mutation sites including 60 substitutions, deletions, and insertions (Kumar et al., 2022). Eight substitutions (K856R, L2084I, A2710T, T3255I, P3395H, and I3758V for ORF1a, and P314L and I1566V for ORF1b) and two deletions (four amino acids) have been identified in ORF1a and ORF1b of the Omicron variant contains (He et al., 2021; Kumar et al., 2022). More than 30 mutations have been identified on S protein of this variant, with half of these mutations (G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, and Y505H) accumulated in the RBD, enhancing binding affinity to ACE2 as well as strengthening viral transmission speed (Kumar et al., 2022; Viana et al., 2022). Furthermore, Omicron possesses D614G in the spike protein that shared by five VOCs, correlating with higher viral loads in the upper respiratory tract and a higher rate of infection in younger persons (Korber et al., 2020; Plante et al., 2021). For envelope (E) and membrane (M) proteins, one substitution (T9I) and three substitutions (D3G, Q19E, and A63T) were identified, respectively (He et al., 2021). In addition, three substitutions (P13L, R203K, and G204R) and a deletion of three amino acid residues (ERS31-33del) in the nucleocapsid (N) proteins were also found (He et al., 2021).

Omicron subvariant BA.5.1.3 was first detected in Romania on 24 February 2022, and has since become the world's prevalent subvariant, identified in 59 countries and territories as of 2 October 2022. The densest concentration of cases has been mainly found in Germany, United States, Denmark, and United Kingdom. Compared with other subvariants, Omicron BA.5.1.3 is more contagious and spreads faster, thanks to the S:V289I mutation (Chen et al., 2021). The S protein of BA.5.1.3 has a common mutation pattern to the BA.4 and BA.5 lineages and is most similar to the BA.2 lineage, excluding 69–70 deletions, L452R, and F486V mutations, and the wild-type amino acid at Q493 (Tegally et al., 2022a).

From 1 August 2022 until early September 2022, a SARS-CoV-2 epidemic raged in Hainan Province in China. In this study, we performed whole-genome SARS-CoV-2 sequencing on RT-PCR positive nasopharyngeal swabs from the city of Sanya (which was the center of the outbreak), detecting the presence of the highly transmissible Omicron BA.5.1.3 subvariant. We then investigated the diversity of 384 genomes including their unique mutations, and compared the binding affinity of several mutations in the S protein. Finally, we conducted phylogenetic analysis and characterized the global distribution of BA.5.1.3.

Materials and methods

Sampling

From August 1st to 18th August 2022, as mentioned previously, SARS-CoV-2 positive (RT-PCR detection) oropharyngeal swab samples were preserved in 3 mL viral sample preservation solution (Inactivated) and collected as part of national COVID-19 routine surveillance from the city of Sanya in Hainan Province in China and preserved in viral transport medium (Zhu et al., 2021, 2022). The total RNA of was extracted from 200 μL sample using the QIAamp viral RNA mini kit (Qiagen, Germany) and eluted using 40 μL RNase-free water. The 5 μL eluted water of each sample was conducted to clinical one-step RT-PCR using detection kit for 2019-nCoV (PCR-Fluorescence) according to the manufacturer's instructions (DaAnGene, China) targeting the *ORF1ab* and *N* genes, respectively. All sampling work was done in collaboration with the national network for COVID-19 surveillance and approved by the Ethics Committee of Sanya People's Hospital (NO: SYPH-2022-043).

Whole-genome sequencing of SARS-CoV-2

The DNA of the positive oropharyngeal samples was removed using DNase I, and the concentration of total RNA was determined

---

[1] https://covid19.who.int/

using Qubit 2.0 (Invitrogen, United States). The sequencing libraries were constructed using a gene-specific multiplexed amplicon-based sequencing strategy (ATOPlex SARS-CoV-2) and sequenced on the MGISEQ-2000 (MGI, China) platform (paired reads 100). The raw reads were qualified using FastQC v0.11.9[2] (Andrews, 2010), and the clean reads were then aligned to the genome of the Wuhan-Hu-1 virus (MN908947.3) using Bowtie 2 (Langmead and Salzberg, 2012) and assembled by SPAdes v3.15.4 (Bankevich et al., 2012).

Lineage classification and variant calling

The lineages of whole SARS-CoV-2 genomes were classified using pangoLEARN v4.0[3] (O'Toole et al., 2021). The global Omicron BA.5.1.3 genomes (sampling location designated as non-China), along with their complete genomes and collection dates, were downloaded from the GISAID database (date of access 2 October 2022).[4] Those genomes with low coverage and > 0.01% of N were parsed. The mutations of genomes were called using SARS-CoV-2-freebayes with Wuhan-Hu-1 as the reference (Farkas et al., 2021), and the obtained VCF files were filtered and annotated using the SnpEff package (Cingolani et al., 2012). The frequency of each variant was calculated, which was then divided into two groups based on sampling location (non-China or China). The single nucleotide polymorphisms (SNPs) of the amino acids and genomes were summarized and visualized using ESPript v3.0 (Robert and Gouet, 2014) and snipit,[5] respectively.

Structural modeling

The native protein structure of hACE2 (1R42) was downloaded from the Protein Data Bank (PDB).[6] The SARS-CoV-2 RBD structures were modeled using SWISS-MODEL with PDB: 7a91.1 as the template (Waterhouse et al., 2018). Receptor (hACE2)-ligand (RBD) docking was conducted via the Hex protein docking algorithm, using the default parameter settings (shape dimension: 0.6, receptor range: 180, ligand range: 180, distance range: 40, and box size: 10; Macindoe et al., 2010).

Phylogenetic analysis

Whole genome sequences from this study were aligned using MAFFT v7.471 set to the default parameters (Katoh and Standley, 2013). The alignment was submitted to Nextstrain v12 using SARS-CoV-2 workflow and visualized using auspice (Hadfield et al., 2018). Additionally, the BA.5.1.3 globally-sourced and China-sourced genomes were aligned using MAFFT set to the fast option (FFT-NS-2) and trimmed using Gblocks v0.91b (Castresana, 2000), which were then used to build phylogenetic trees using FastTree v2.1 (Price et al., 2010). The haplotype phylogenetic network of *S* gene was clustered by sampling date and built using PopART v1.7 (Leigh and Bryant, 2015).

Molecular clock estimation

The genomes from this study and representatives of Nextstrain clades belonging to Omicron were aligned and used to infer its preliminary maximum-likelihood tree, which was also inspected for the presence of a temporal signal using TempEst v.1.5.3 (Rambaut et al., 2016). Bayesian time-calibrated tree was build using BEAST v1.10.4 (Suchard et al., 2018), with a strict molecular clock model, an exponential growth coalescent model, and an HKY + I substitution model (Tegally et al., 2021). The length of the Markov chain of the Monte Carlo (MCMC) tree was 1×10^8, with sampling every 1,000 steps. The result was evaluated using Tracer v1.7.1 (effective sample size >200) and summarized in TreeAnnotator v1.10.4, discarding 10% as burn-in (Rambaut et al., 2018; Suchard et al., 2018).

Phylogeographic analysis

To infer the phylogenetic diffusion, the genomic alignment and corresponding location traits were jointly inferred at 1×10^8 generations and sampled every 10,000 steps using BEAST v1.10.4 (Suchard et al., 2018). The geographic transition routes (country level) and posterior probability (PP) were summarized and visualized using SPREAD3 v0.9.7.1 with the Bayesian stochastic search variable selection (BSSVS) method (Bielejec et al., 2016).

Statistical analysis

Statistical plots were conducted using Origin Pro 2021 version. The significant difference was estimated by GraphPad Prism v8.3.0, using either the Mann–Whitney test or Kruskal–Wallis test with $p < 0.05$ indicating statistical significance.

Results

Outbreak investigation

On 1 August 2022, a SARS-CoV-2 outbreak erupted in Sanya, the southernmost city on the island province of Hainan in China. This epidemic lasted until early September 2022, resulting in 8,630 confirmed cases with 6,595 cases from the city of Sanya (Supplementary Figure S1). The nasopharyngeal swab specimen of the first case was analyzed using next-generation sequencing, indicating that the causative agent was SARS-CoV-2 Omicron BA.5.1.3 (Nextstrain clade 22B), which had spread to more than 60 countries and territories worldwide at the time this sample was collected (Figure 1A). The outbreak in Sanya has been attributed to China's first cases of the Omicron BA. 5.1.3. From 1 August 1 to 18 August, a total of 391 SARS-CoV-2 positive specimens with corresponding metadata were randomly collected to perform further analysis (Figure 1B). More than half of the confirmed cases were female (58%, 227/391; Figure 1B). The age of individuals ranged from 2 months to 83.0 years old, with a median of 37.0 (IQR 23.0–53.0) years old

---

[2] https://github.com/s-andrews/FastQC
[3] https://github.com/cov-lineages/pangoLEARN
[4] https://gisaid.org/
[5] https://github.com/aineniamh/snipit
[6] https://www.rcsb.org/

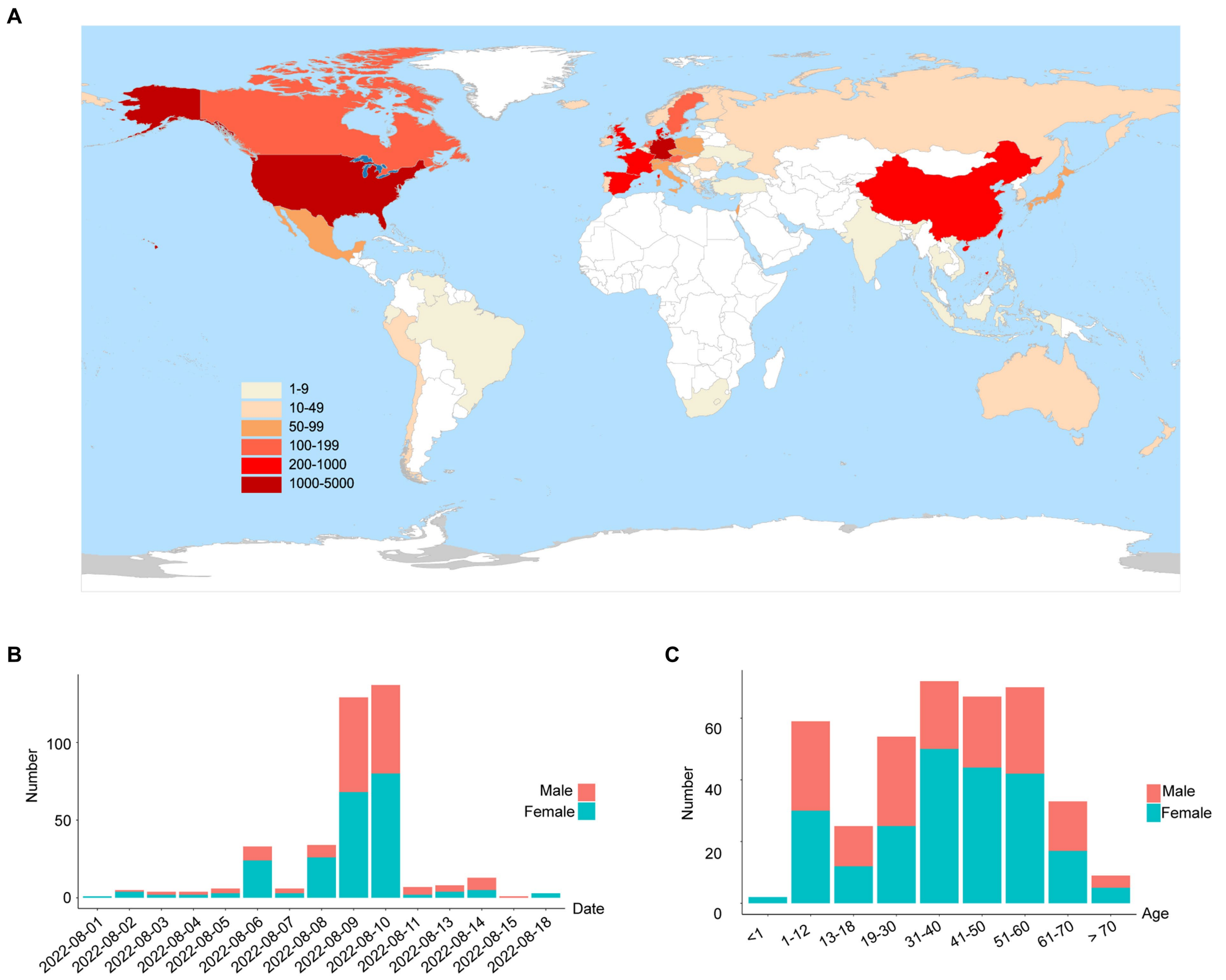


FIGURE 1
The global emergence of Omicron BA.5.1.3 and cases identified in this study. **(A)** Map showing the global distribution of Omicron BA.5.1.3. The locations of genomes from this study and GISAID (date of access 2 October 2022) were grouped by countries and territories. The shade of each color represents the corresponding confirmed cases per country. **(B)** The sampling date of confirmed cases in the city of Sanya. **(C)** The age range of confirmed cases in Sanya. Orange represents male cases and blue represents female cases.

(Figure 1C). The cycle threshold (Ct) value of *ORF1ab* gene ranged from 14.8 to 35.2 with median of 24.9 (IQR 22.6–27.3; Figures 2A,B). The Ct value of *N* gene ranged from 13.7 to 35.0, with a median of 25.2 (IQR 22.9–27.6; Figures 2A,C). The difference in Ct values between female and male cases was insignificant ($p > 0.05$) for both *ORF1ab* and *N* gene (Figures 2B,C).

Evolutionary analysis

After removing the low-quality genomes from the total genomes collected, a total of 384 SARS-CoV-2 genomes were retained for phylogenetic analysis. The results from Nextstrain and Nextclade analyses showed that all these SARS-CoV-2 genomes clustered together and belonged to Omicron BA.5.1.3 (Nextstrain clade 22B; Figure 3A). To assess their detailed phylogenetic relationship, all available BA.5.1.3 Omicron genomes (4,123 genomes retained after parsing) were retrieved from the GISAID database (2 October 2022) and used to build a maximum-likelihood tree. Results revealed that SARS-CoV-2 genomes from the city of Sanya formed a monophyletic cluster within all BA.5.1.3 genomes (Figure 3B). The *S* gene of Omicron BA.5.1.3 genomes from Sanya could be divided into 15 representative haplotypes, which indicated that the vast majority of genomes (83.3%, 320/384) belonged to haplotype 1 (H1), followed by H2 (8.6%, 33/384) and H3 (3.6%, 14/384; Figure 3C).

Mutational profile

In total, the genomes from China (in the city of Sanya) had marked hypermutation, including 70 to 90 mutations (containing deletions) with a median of 80 (IQR 80–81), showing a significant difference ($p < 0.05$) with non-China-sourced genomes (global; Figure 4A). During the outbreak, 25 core mutations (frequency > 98%) were identified in ORF1ab (Supplementary Table S1), which also included three unique mutations (Y1064N, S2844G, and R3574K) that were exclusive or non-core to global BA.5.1.3 genomes (Table 1; Figure 4B).

Obviously, the most mutations were located in the *S* gene, including 35 core mutations (frequency > 99% except for N501Y and

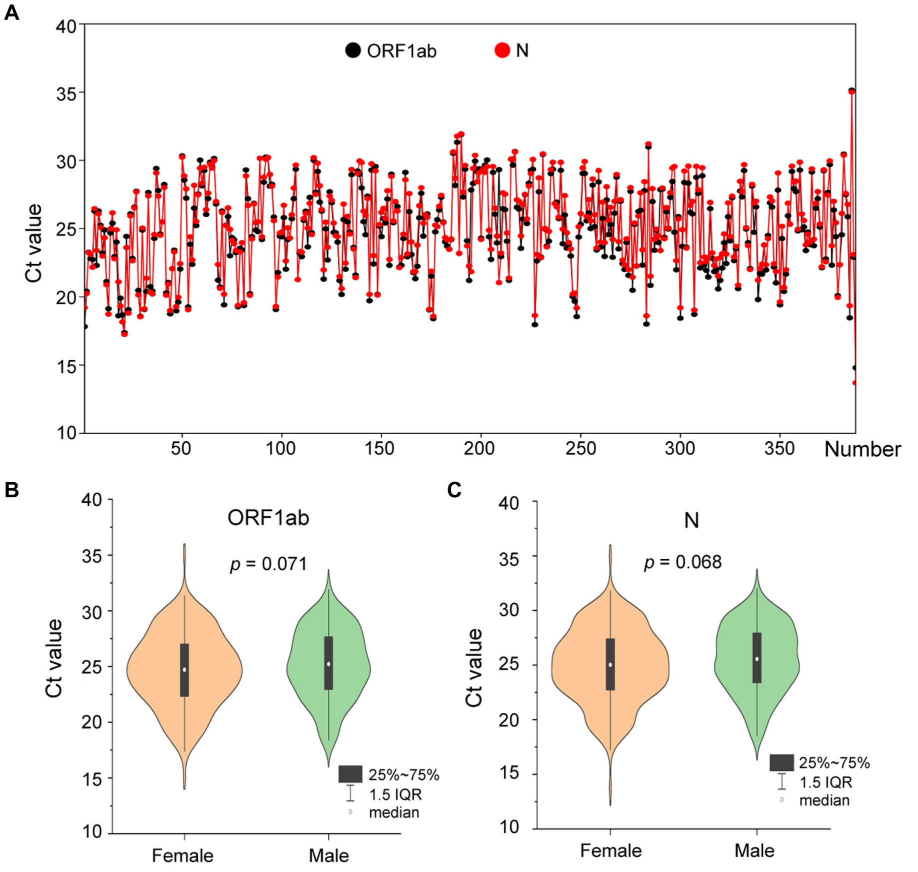


FIGURE 2
The Ct values of confirmed cases from the city of Sanya. **(A)** The Ct value of each case. The values of *ORF1ab* and *N* genes are labeled with black- and red-filled circles, respectively. **(B)** The comparison between female and male cases based on Ct value of ORF1ab. **(C)** The comparison between female and male cases based on Ct value of N.

Y505H) with one synonymous mutation (D1146D/C3438T). Half (17) of these core mutations of dominating genomes (H1) from the city of Sanya were located in the RBD domain (Figure 4C). The mutation profile of the *S* gene shared the core mutations of global Omicron BA.5.1.3 genomes (Table 2). In addition, all Sanya-sourced and globally sourced BA.5.1.3 genomes (>99.5%) shared the V289I mutation that was absent in the core mutations of any Omicron sub-lineages, while BA.5.1.3 did not contain the G142D mutation that shared by all other Omicron variants (BA.1, BA.2, BA.4, BA.5, BA.2.12.1, BQ.1, and XBB; Hodcroft, 2021). Interestingly, two revertant mutations with a relatively low frequency were found in position 501 (13.8%, 53/384) and position 505 (10.1%, 39/384), both belonging to the RBD domain (Table 2; Figure 4C).

We detected three (two synonymous and one non-synonymous) core (100%) mutations in ORF3a (T64T, S74S, and T223I), including one novel synonymous mutation (S74S/C222T; Table 1; Figure 4B). Additionally, 13 mutations, namely one non-synonymous mutation (T9I) in E, two non-synonymous mutations (D3N and Q19E) in M, one synonymous mutation (C15C/T45C) in ORF7a, one synonymous mutation (L18L) in ORF7b, seven non-synonymous mutations (P13L, ERS31-33del, RG203KR, and S413R) in N, and one non-synonymous mutation (L37F) in ORF10, were identified as core mutations. Meanwhile, four core mutations were identified in the non-coding regions, including C241T (5′UTR), C27889T (between ORF7b and ORF8), A28271T (between ORF8 and N), and 29734_29759del (3′UTR).

Receptor-ligand docking

The binding affinity between SARS-CoV-2 RBD and hACE2 accounted for variant infectivity and differed between variants (Kumar et al., 2022). Based on the haplotype analysis, the genome sequences of SY005, SY002, and SY034 were designated as representatives of H1, H2, and H3, respectively. We compared the docking energies of the RBD of H1 (SY005, EPI_ISL_15940718), H2 (SY002, EPI_ISL_15940715), and H3 (SY034, EPI_ISL_15940747) binding to hACE2 (Figure 4D), respectively. Results revealed that the docking energy of Wuhan-Hu-1 with hACE2 was −672.59, which was lower than that of the dominant strain SY005 (−676.89), indicating the higher infectivity and transmissibility of the virus in Sanya. The revertant mutation in position 501 (SY034) had a slightly lower docking energy (−673.29) compared with SY005, signifying that N501Y in the RBD promoted viral binding (Figure 4D). The two mutations in positions 501 and 505 reverted simultaneously (SY002) and showed more lower docking energy compared with SY005 and SY034, indicating that both N501Y and Y505H in the RBD might promote viral infectivity (Figure 4D).

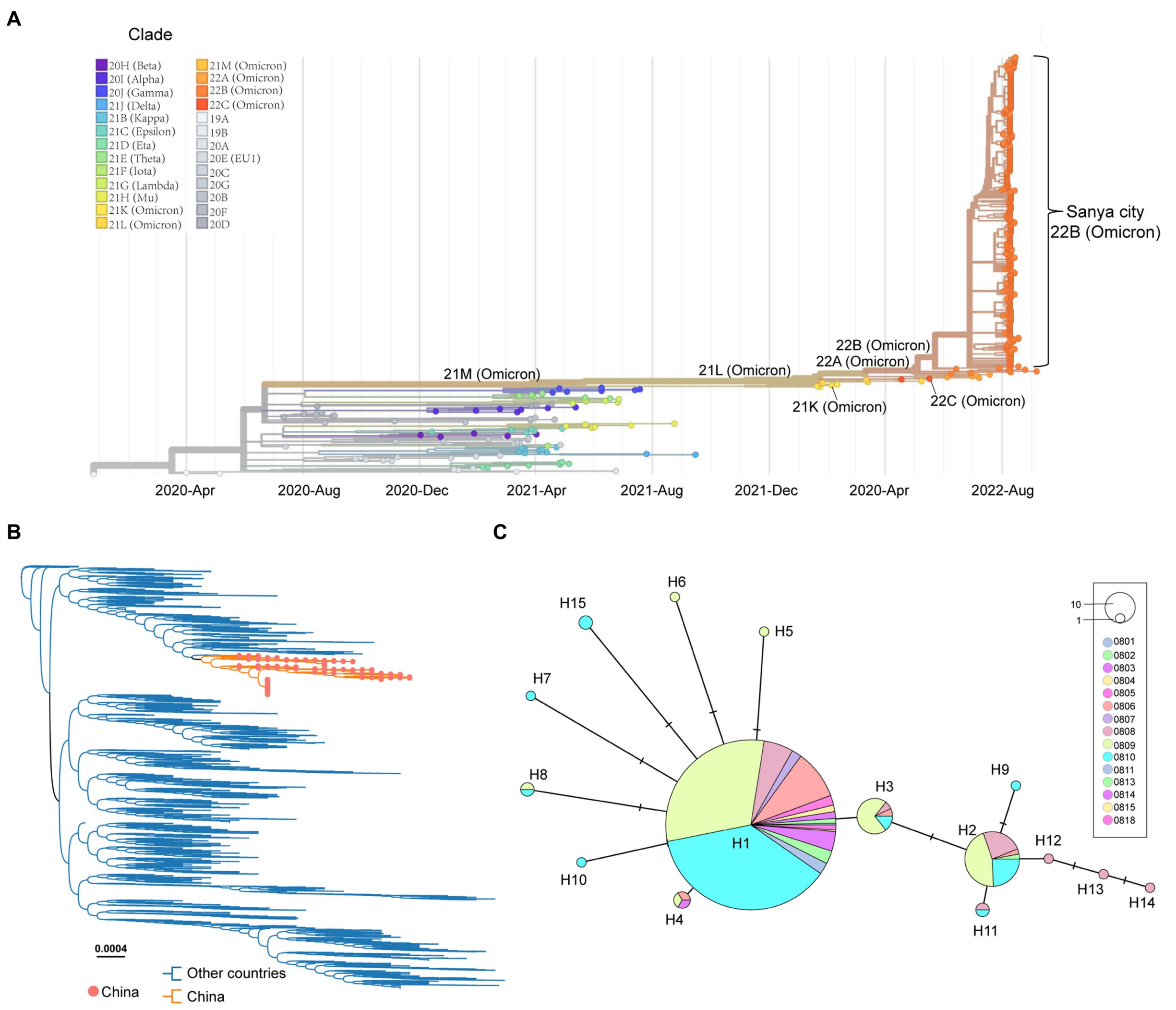


FIGURE 3
The overview of phylogenetic distribution of genomes identified from the city of Sanya. **(A)** Nextstrain's phylodynamic analysis of genomes from Sanya City. **(B)** The phylogenetic analysis, including BA.5.1.3genomes from the GISAID database and the city of Sanya. The sequences from this study are labeled with orange-filled circles and orange branches. **(C)** The haplotype network of the *S* genes identified in the city of Sanya. The network is built by PopART using the median-joining method. The size of the circles represents the number of corresponding haplotypes. The color indicates the corresponding sampling date.

Molecular clock analysis

To investigate the phylogenetic history of BA.5.1.3 genomes from Sanya, the representatives of Omicron genomes were used as background to conduct a molecular clock analysis. The temporal signal test showed that the sampling dates were correlated with the root-to-tip distance, implying a significant ($R^2 = 1$) temporal signature (Figure 5). The Bayesian analysis estimated the substitutions of Omicron BA.5.1.3 genomes at 8.896×10^{-4} nucleotide changes per site per year, which was a reliable rate compared to other lineages (Tegally et al., 2021). The time-calibrated phylogenetic tree tracked that the most recent common ancestor (tMRCA) of all BA.5.1.3 genomes from Sanya emerged in early July 2022 with 95% highest posterior density (HPD) ranging from the middle of May to the middle of September 2022 (Figure 5).

Spatial dynamics

To illustrate the population genetics, the Bayesian skyline analysis was conducted based on global BA.5.1.3 genomes, including those from SARS-CoV-2 Omicron BA.5.1.3 outbreak in Sanya. The result showed that the population size of Omicron BA.5.1.3 grew steadily since its emergence in February 2022 and reached the relatively peak in mid-July 2022 (Supplementary Figure S2). The population size reached the summit at the end of August 2022 and decreased starting from that time (Supplementary Figure S2). To infer their spatial spread, the Omicron BA.5.1.3 genomes were first split into different groups based on their corresponding countries and territories. We performed Bayesian phylogeographic analysis according to the BSSVS method, detecting the presence of 18 migration links in the transmission processes of BA.5.1.3 worldwide, with posterior probability (PP) values of 1 (Supplementary Figure S3).

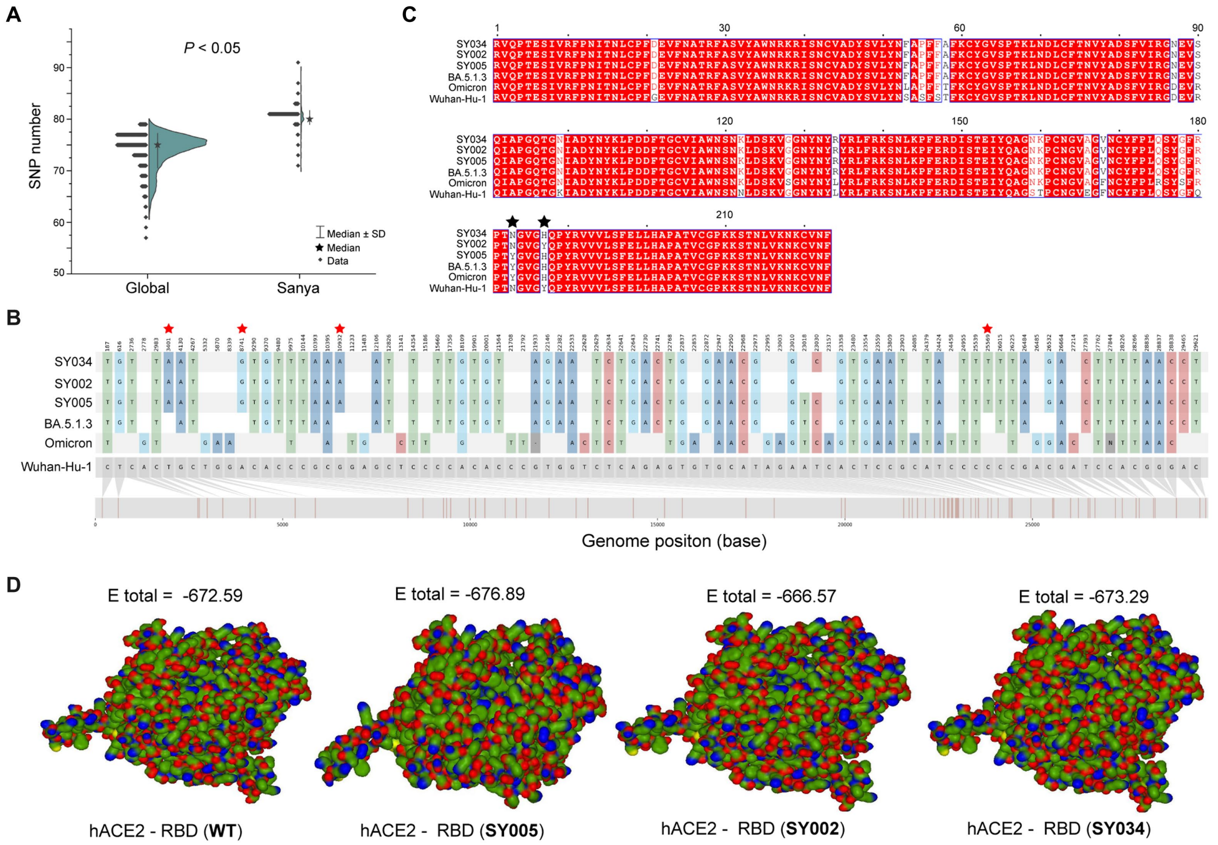


FIGURE 4
The mutation analyses of whole genomes identified in this study. **(A)** The comparison of SNP numbers of BA.5.1.3 between China (city of Sanya) and non-China (global) genomes. **(B)** The whole genome SNP map with Wuhan-Hu-1, Omicron, BA.5.1.3, SY005, SY002, and SY034. The red stars represent the position of unique mutations. **(C)** The comparison based on amino acid sequences of the RDB. The black stars represent the 501 and 505 positions of S proteins. **(D)** The receptor-ligand docking between hACE2 and the corresponding RBD. The E total value above each structure indicates the docking scores. The Wuhan-Hu-1, Omicron, and BA.5.1.3 represent the genomes of the original SARS-CoV-2 (MN908947.3), the first Omicron variant (EPI_ISL_6640917), and the first BA.5.1.3 variant (EPI_ISL_14026136), which are available in public databases, respectively. WT here stands for Wuhan-Hu-1.

TABLE 1 Unique mutations identified in BA.5.1.3 genomes from China.

Gene	Protein	Position (aa)	Mutation (nt)	Mutation (aa)	Frequency	Percentage (%)	Synonymous mutation
ORF1ab	nsp3	1,064	T3190A	Y1064N	384	100	N
ORF1ab	nsp4	2,844	A8530G	S2844G	384	100	N
ORF1ab	nsp6	3,574	G10721A	R3574K	384	100	N
ORF3a	-	74	C222T	S74S	384	100	Y

nsp, non-structural protein; N, No; Y, Yes.

Obviously, certain European and European-adjacent countries, especially Germany and United Kingdom, may have played a pivotal role in BA.5.1.3 transmission (Supplementary Figure S3; Supplementary Table S2).

Discussion

The failure to control the transmission of SARS-CoV-2 in many countries and territories around the world has led to the development of hypotheses suggesting that substantial evolutionary mutations accounting for the emergence of virus variants, particularly the Omicron (Fontanet et al., 2021; Mascola et al., 2021; Tegally et al., 2021). In this study, we investigated the China's first SARS-CoV-2 Omicron BA.5.1.3 outbreak, which has caused more than 8,500 confirmed cases and led to the lockdown of the tropical island known as Hainan Province.

The number of confirmed cases rose quickly in the initial stages of this outbreak. More than half these cases are female (58%), differing from those of the WT (<50%), Alpha (39%), Beta (27%), and Delta (49%) variants (Huang et al., 2020; Ong et al., 2022). In this cohort, we also unexpectedly observed three infants (≤1 years old) and 11

TABLE 2 Mutations found in S protein of BA.5.1.3 genomes from China.

Position (aa)	Mutation (nt)	Mutation (aa)	Frequency	Percentage of China (%)	Percentage of the world (%)	Synonymous mutation
19	C56T	T19I	381	99.2	97.0	N
24	deletion	deletion	382	99.5	97.2	N
25	deletion	deletion	382	99.5	97.2	N
26	deletion	deletion	382	99.5	97.2	N
27	deletion	deletion	382	99.5	97.2	N
69	deletion	deletion	380	99.0	97.0	N
70	deletion	deletion	380	99.0	97.0	N
213	T638G	V213G	384	100	99.3	N
289	G865A	V289I	384	100	99.2	N
339	G1016A	G339D	384	100	91.3	N
371	C1112T	S371F	384	100	87.8	N
373	T1117C	S373P	384	100	91.6	N
375	deletion	S375F	384	100	91.3	N
376	deletion	T376A	384	100	91.3	N
405	G1213A	D405N	384	100	92.4	N
408	A1224C	R408S	384	100	89.6	N
417	G1251T	K417N	384	100	91.1	N
440	T1320G	N440K	384	100	85.1	N
452	T1355G	L452R	384	100	92.3	N
477	G1430A	S477N	384	100	92.7	N
478	C1433A	T478K	384	100	92.7	N
484	A1451C	E484A	384	100	93.0	N
486	T1456G	F486V	384	100	93.1	N
498	A1493G	Q498R	384	100	91.6	N
501	A1501T	N501Y	331	86.2	92.4	N
505	T1513C	Y505H	345	89.8	92.8	N
614	A1841G	D614G	384	100	99.8	N
655	C1963T	H655Y	384	100	99.9	N
679	T2037G	N679K	384	100	99.3	N
681	C2042A	P681H	384	100	99.2	N
764	C2292A	N764K	384	100	95.8	N
796	G2386T	D796Y	384	100	99.4	N
954	A2862T	Q954H	384	100	99.5	N
969	T2907A	N969K	384	100	99.6	N
1,146	C3438T	D1146D	381	99.2	97.7	Y

N, No; Y, Yes.

persons of advanced age (≥70 years old), suggesting that people of all ages may be susceptible to Omicron BA.5.1.3. The viral loads in the upper respiratory tract caused by BA.5.1.3 among different age and sex groups show no obvious difference. Surprisingly, the BA.5.1.3 in this study seems to result in higher overall viral load (with a median Ct value of 25.2 for *N* gene from here) compared with BA.1 (median = 26.29) and BA.2 (median = 25.48) in symptomatic respondents (Whitaker et al., 2022). Meanwhile, the overall median Ct value of 26.4 from large-community surveillance study (Walker et al., 2021) indicated potentially faster transmission and enhanced viral fitness of BA.5.1.3.

The highly mutated SRAS-CoV-2 Omicron BA.5.1.3 have accumulated 35 core mutations in the S protein, higher than that found in the S proteins of Alpha (10), Beta (11), Delta (12), Gamma (9), and Omicron BA.2 (31), though lower than that found in the S protein of Omicron BA.1 (36; Elbe and Buckland-Merrett, 2017). Furthermore, this finding identically reflected the mutational pattern of BA.5.1.3 as it circulated around the world. Half of these mutations

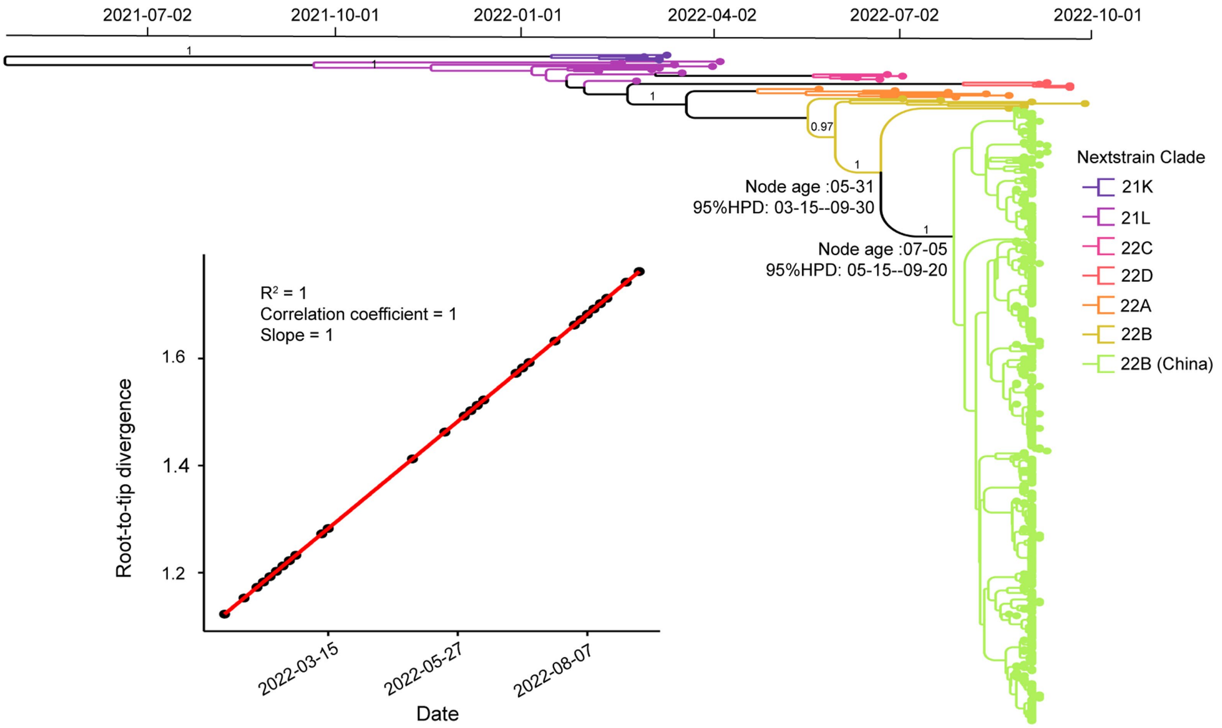


FIGURE 5
Temporal analysis of Omicron BA.5.1.3 genomes from China. The branches are labeled with different colors according to Nextstrain's clades. The posterior probabilities are indicated on corresponding branches. The tMRCA of genomes from Sanya is labeled on their corresponding nodes with 95% highest probability density (HPD) values. The molecular clock signal plot revealed the correlation between sampling dates and the root-to-tip distance. The tMRCA was based on collection dates.

are located in the RBD, resulting in higher binding affinity and selective advantage. The receptor-ligand docking estimation confirmed that N501Y and Y505H mutations of the S protein in SARS-CoV-2 enhance its binding affinity to hACE2, suggesting a higher rate of potential immune escape (Tian et al., 2021; Kumar et al., 2022). The S:V289I mutation is detected in all genomes from Sanya and > 99.5% BA.5.1.3 genomes from the GISAID database. Globally, genomes containing the S:V289I mutation accounted for 0.36% of all SARS-CoV-2 genomes, primarily dominated and prevalent in AY.47 (73.55%) and BA.5.1.3 (20.73%; Elbe and Buckland-Merrett, 2017; Chen et al., 2021). The S:G142D mutation is in about half (48.96%) of all SARS-CoV-2 genomes, mainly in BA.2 (18.00%), AY.4 (9.35%), BA.2.12.1 (4.08%), BA.5.2.1 (3.76%), BA.5.2 (3.37%), AY.103 (3.11%), BA.2.9 (3.11%), and BA.5.1 (3.05%; 2 October 2022; Chen et al., 2021). Thus, the absence of S:G142D mutation and presence of S:V289I mutation could be potential markers of BA.5.1.3 sub-lineage when compared with other sub-lineages (Chen et al., 2021; Hodcroft, 2021).

Three (Y1064N, S2844G, and R3574K) unique core (100%) mutations in ORF1ab are identified in BA.5.1.3 genomes from this study. Before 1 August 2022, there are relatively few genomes (94 belonging to 30 sub-lineages) harboring the ORF1a:Y1064N mutation, with only 5 genomes in BA.5.1.3, and 2,754 genomes (belonging to 46 sub-lineages) possessing the ORF1a:R3574K mutation with only 670 genomes in BA.5.1.3 from public database (Chen et al., 2021). However, the ORF1a:S2844G has not so far been detected in BA.5.13 genomes. Given that other core mutations are identical to publicly available BA.5.1.3 genomes, these unique core mutations of genomes from Sanya indicate that the Omicron BA.5.1.3 outbreak shares the same transmission chain. In addition, its three unique core mutations are within non-structural protein (nsp) 3, nsp4, and nsp6, respectively, which play different but equally essential roles in the viral life cycle (Wu et al., 2020; Armstrong et al., 2021; Wang and Cheng, 2022). As the largest non-structural protein, nsp3 plays a pivotal role in SARS-CoV-2 replication and transcription complexes assembly (Armstrong et al., 2021). When nsp4 binds with nsp3, this can cause membrane rearrangement and generate double-membrane vesicles with nsp6 protein as the connectors (Sakai et al., 2017; Ricciardi et al., 2022). Thus, the three unique core mutations may enhance the interaction of these non-structural proteins, resulting in more efficient viral replication.

The time-resolved maximum clade credibility (MCC) analysis estimated that the emergence of tMRCA was 1 month before the Hainan Province outbreak. Due to the earlier use of regular and large-scale nucleic acid amplification tests earlier in China, it seems impossible that the virus could circulate in Sanya City for so long without being detected, suggesting that this Omicron BA.5.1.3 strain might have evolved and adapted for some time before spreading to China. However, the detailed transmission routine of the Omicron BA.5.1.3 to China is still unclear.

This study also has some limitations. First, the clinical symptoms were unavailable to investigate the symptomatic differences from other lineages. Second, the nasopharyngeal swab specimens were only collected once, thus limiting us from monitoring long-term viral load changes. In addition, we were unable to sequence the whole genomes from all positive specimens, indicating of a potential underestimation of population size and mutations.

Conclusion

Our study tracks the BA.5.1.3 outbreak in the city of Sanya in Hainan Province in China, delineates its mutational profile (including unique mutation), and estimates the timing of its emergence. Our results highlight the importance of genomic surveillance and of carrying out collaboration and sharing of genomic data worldwide for an effective response to the COVID-19 pandemic.

Data availability statement

The datasets presented in this study have been deposited in the GISAID database (https://www.gisaid.org/) under accession numbers EPI_ISL_15940714 to EPI_ISL_15941103, and EPI_ISL_15942295.

Author contributions

WZ and CW conceived and designed this study. XW, XZ, YL, and LH collected samples and conducted the experiments. WZ analyzed the data and prepared all the figures. WZ and XW wrote the first manuscript. WZ, CW, and JY revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by Golden Seed Program of Beijing Chaoyang Hospital (CYJZ202220) and also supported by the innovation platform for Jianguo Xu Academicians of Hainan Province.

Acknowledgments

We would like to thank the authors from the originating laboratories that generated SARS-CoV-2 sequences used as reference dataset in this study, as well as the GISAID database (https://www.gisaid.org/) for providing their open-sharing platform.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1183633/full#supplementary-material

References

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data.* Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc

Araf, Y., Akter, F., Tang, Y. D., Fatemi, R., Parvez, M. S. A., Zheng, C., et al. (2022). Omicron variant of SARS-CoV-2: genomics, transmissibility, and responses to current COVID-19 vaccines. *J. Med. Virol.* 94, 1825–1832. doi: 10.1002/jmv.27588

Armstrong, L. A., Lange, S. M., Dee Cesare, V., Matthews, S. P., Nirujogi, R. S., Cole, I., et al. (2021). Biochemical characterization of protease activity of Nsp3 from SARS-CoV-2 and its inhibition by nanobodies. *PLoS One* 16:e0253364. doi: 10.1371/journal.pone.0253364

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Berkhout, B., and Herrera-Carrillo, E. (2022). SARS-CoV-2 evolution: on the sudden appearance of the omicron variant. *J. Virol.* 96:e0009022. doi: 10.1128/jvi.00090-22

Bielejec, F., Baele, G., Vrancken, B., Suchard, M. A., Rambaut, A., and Lemey, P. (2016). SpreaD3: interactive visualization of spatiotemporal history and trait evolutionary processes. *Mol. Biol. Evol.* 33, 2167–2169. doi: 10.1093/molbev/msw082

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334

Chen, C., Nadeau, S., Yared, M., Voinov, P., Xie, N., Roemer, C., et al. (2021). CoV-Spectrum: analysis of globally shared SARS-CoV-2 data to identify and characterize new variants. *Bioinformatics* 38, 1735–1737. doi: 10.1093/bioinformatics/btab856

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* 6, 80–92. doi: 10.4161/fly.19695

Elbe, S., and Buckland-Merrett, G. (2017). Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* 1, 33–46. doi: 10.1002/gch2.1018

Farkas, C., Mella, A., Turgeon, M., and Haigh, J. J. (2021). A novel SARS-CoV-2 viral sequence Bioinformatic pipeline has found genetic evidence that the viral 3' untranslated region (UTR) is evolving and generating increased viral diversity. *Front. Microbiol.* 12:665041. doi: 10.3389/fmicb.2021.665041

Fontanet, A., Autran, B., Lina, B., Kieny, M. P., Karim, S. S. A., and Sridhar, D. (2021). SARS-CoV-2 variants and ending the COVID-19 pandemic. *Lancet* 397, 952–954. doi: 10.1016/s0140-6736(21)00370-6

Hadfield, J., Megill, C., Bell, S. M., Huddleston, J., Potter, B., Callender, C., et al. (2018). Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 34, 4121–4123. doi: 10.1093/bioinformatics/bty407

He, X., Hong, W., Pan, X., Lu, G., and Wei, X. (2021). SARS-CoV-2 omicron variant: characteristics and prevention. *MedComm* 2, 838–845. doi: 10.1002/mco2.110

Hodcroft, E. (2021). CoVariants: SARS-CoV-2 mutations and variants of interest. Available at: https://covariants.org/.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395, 497–506. doi: 10.1016/s0140-6736(20)30183-5

Jackson, C. B., Farzan, M., Chen, B., and Choe, H. (2022). Mechanisms of SARS-CoV-2 entry into cells. *Nat. Rev. Mol. Cell Biol.* 23, 3–20. doi: 10.1038/s41580-021-00418-x

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., et al. (2020). Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cells* 182, 812–827.e19. doi: 10.1016/j.cell.2020.06.043

Kumar, S., Thambiraja, T. S., Karuppanan, K., and Subramaniam, G. (2022). Omicron and Delta variant of SARS-CoV-2: a comparative computational study of spike protein. *J. Med. Virol.* 94, 1641–1649. doi: 10.1002/jmv.27526

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., et al. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581, 215–220. doi: 10.1038/s41586-020-2180-5

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Leigh, J. W., and Bryant, D. (2015). Popart: full-feature software for haplotype network construction. *Methods Ecol. Evol.* 6, 1110–1116. doi: 10.1111/2041-210x.12410

Macindoe, G., Mavridis, L., Venkatraman, V., Devignes, M. D., and Ritchie, D. W. (2010). HexServer: an FFT-based protein docking server powered by graphics processors. *Nucleic Acids Res.* 38, W445–W449. doi: 10.1093/nar/gkq311

Mascola, J. R., Graham, B. S., and Fauci, A. S. (2021). SARS-CoV-2 viral variants-tackling a moving target. *JAMA* 325, 1261–1262. doi: 10.1001/jama.2021.2088

Maxmen, A. (2022). Are new omicron subvariants a threat? Here's how scientists are keeping watch. *Nature* 604, 605–606. doi: 10.1038/d41586-022-01069-4

O'Toole, Á., Scher, E., Underwood, A., Jackson, B., Hill, V., McCrone, J. T., et al. (2021). Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evol* 7:veab064. doi: 10.1093/ve/veab064

Ong, S. W. X., Chiew, C. J., Ang, L. W., Mak, T. M., Cui, L., Toh, M., et al. (2022). Clinical and Virological features of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants of concern: a retrospective cohort study comparing B.1.1.7 (alpha), B.1.351 (Colonia et al.), and B.1.617.2 (Delta). *Clin. Infect. Dis.* 75, e1128–e1136. doi: 10.1093/cid/ciab721

Plante, J. A., Liu, Y., Liu, J., Xia, H., Johnson, B. A., Lokugamage, K. G., et al. (2021). Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* 592, 116–121. doi: 10.1038/s41586-020-2895-3

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarization in Bayesian Phylogenetics using tracer 1.7. *Syst. Biol.* 67, 901–904. doi: 10.1093/sysbio/syy032

Rambaut, A., Lam, T. T., Max Carvalho, L., and Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly path-O-gen). *Virus Evol* 2:vew007. doi: 10.1093/ve/vew007

Ricciardi, S., Guarino, A. M., Giaquinto, L., Polishchuk, E. V., Santoro, M., di Tullio, G., et al. (2022). The role of NSP6 in the biogenesis of the SARS-CoV-2 replication organelle. *Nature* 606, 761–768. doi: 10.1038/s41586-022-04835-6

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42, W320–W324. doi: 10.1093/nar/gku316

Roychoudhury, S., Das, A., Sengupta, P., Dutta, S., Roychoudhury, S., Choudhury, A. P., et al. (2020). Viral pandemics of the last four decades: pathophysiology, health impacts and perspectives. *Int. J. Environ. Res. Public Health* 17:9411. doi: 10.3390/ijerph17249411

Roychoudhury, S., das, A., Sengupta, P., Dutta, S., Roychoudhury, S., Kolesarova, A., et al. (2021). Viral pandemics of twenty-first century. *J. Microbiol. Biotechnol. Food Sci.* 10, 711–716. doi: 10.15414/jmbfs.2021.10.4.711-716

Sakai, Y., Kawachi, K., Terada, Y., Omori, H., Matsuura, Y., and Kamitani, W. (2017). Two-amino acids change in the nsp4 of SARS coronavirus abolishes viral replication. *Virology* 510, 165–174. doi: 10.1016/j.virol.2017.07.019

Saxena, S. K., Kumar, S., Ansari, S., Paweska, J. T., Maurya, V. K., Tripathi, A. K., et al. (2022). Characterization of the novel SARS-CoV-2 omicron (B.1.1.529) variant of concern and its global perspective. *J. Med. Virol.* 94, 1738–1744. doi: 10.1002/jmv.27524

Shrestha, L. B., Foster, C., Rawlinson, W., Tedla, N., and Bull, R. A. (2022). Evolution of the SARS-CoV-2 omicron variants BA.1 to BA.5: implications for immune escape and transmission. *Rev. Med. Virol.* 32:e2381. doi: 10.1002/rmv.2381

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., and Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* 4:vey016. doi: 10.1093/ve/vey016

Tegally, H., Moir, M., Everatt, J., Giovanetti, M., Scheepers, C., Wilkinson, E., et al. (2022a). Emergence of SARS-CoV-2 omicron lineages BA.4 and BA.5 in South Africa. *Nat. Med.* 28, 1785–1790. doi: 10.1038/s41591-022-01911-2

Tegally, H., Moir, M., Everatt, J., Giovanetti, M., Scheepers, C., Wilkinson, E., et al. (2022b). Continued emergence and evolution of omicron in South Africa: new BA.4 and BA.5 lineages. medRxiv [Preprint].

Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., et al. (2021). Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* 592, 438–443. doi: 10.1038/s41586-021-03402-9

Tian, F., Tong, B., Sun, L., Shi, S., Zheng, B., Wang, Z., et al. (2021). N501Y mutation of spike protein in SARS-CoV-2 strengthens its binding to receptor ACE2. *Elife* 10: e69091. doi: 10.7554/eLife.69091

Viana, R., Moyo, S., Amoako, D. G., Tegally, H., Scheepers, C., Althaus, C. L., et al. (2022). Rapid epidemic expansion of the SARS-CoV-2 omicron variant in southern Africa. *Nature* 603, 679–686. doi: 10.1038/s41586-022-04411-y

Walker, A. S., Pritchard, E., House, T., Robotham, J. V., Birrell, P. J., Bell, I., et al. (2021). Ct threshold values, a proxy for viral load in community SARS-CoV-2 cases, demonstrate wide variation across populations and over time. *Elife* 10:e64683. doi: 10.7554/eLife.64683

Wang, L., and Cheng, G. (2022). Sequence analysis of the emerging SARS-CoV-2 variant omicron in South Africa. *J. Med. Virol.* 94, 1728–1733. doi: 10.1002/jmv.27516

Wang, L., Zhou, H. Y., Li, J. Y., Cheng, Y. X., Zhang, S., Aliyari, S., et al. (2022). Potential intervariant and intravariant recombination of Delta and omicron variants. *J. Med. Virol.* 94, 4830–4838. doi: 10.1002/jmv.27939

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46, W296–w303. doi: 10.1093/nar/gky427

Whitaker, M., Elliott, J., Bodinier, B., Barclay, W., Ward, H., Cooke, G., et al. (2022). Variant-specific symptoms of COVID-19 in a study of 1,542,510 adults in England. *Nat. Commun.* 13:6856. doi: 10.1038/s41467-022-34244-2

WHO (2022). Tracking SARS-CoV-2 variants. Available at: https://www.who.int/activities/tracking-SARS-CoV-2-variants.

Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., et al. (2020). Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe* 27, 325–328. doi: 10.1016/j.chom.2020.02.001

Zhu, W., Yang, J., Lu, S., Huang, Y., Jin, D., Pu, J., et al. (2022). Novel pegiviruses infecting wild birds and rodents. *Virol. Sin.* 37, 208–214. doi: 10.1016/j.virs.2022.01.013

Zhu, W., Yang, J., Lu, S., Lan, R., Jin, D., Luo, X. L., et al. (2021). Beta- and Novel Delta-coronaviruses are identified from wild animals in the Qinghai-Tibetan plateau, China. *Virol. Sin.* 36, 402–411. doi: 10.1007/s12250-020-00325-z

frontiers | Frontiers in Microbiology

TYPE Review
PUBLISHED 02 June 2023
DOI 10.3389/fmicb.2023.1200681

EDITED BY
Mohamad S. Hakim,
Gadjah Mada University, Indonesia

REVIEWED BY
Hao Xiaoli,
Yangzhou University, China
Min Liao,
Zhejiang University, China

*CORRESPONDENCE
Zhixun Xie
✉ xiezhixun@126.com

RECEIVED 05 April 2023
ACCEPTED 15 May 2023
PUBLISHED 02 June 2023

CITATION
Rafique S, Rashid F, Mushtaq S, Ali A, Li M, Luo S, Xie L and Xie Z (2023) Global review of the H5N8 avian influenza virus subtype. *Front. Microbiol.* 14:1200681. doi: 10.3389/fmicb.2023.1200681

Global review of the H5N8 avian influenza virus subtype

Saba Rafique[1], Farooq Rashid[2], Sajda Mushtaq[1], Akbar Ali[3], Meng Li[4,5,6], Sisi Luo[4,5,6], Liji Xie[4,5,6] and Zhixun Xie[4,5,6]*

[1]SB Diagnostic Laboratory, Sadiq Poultry Pvt. Ltd, Rawalpindi, Pakistan, [2]Department of Infectious Diseases, Chongqing Public Health Medical Center, Chongqing, China, [3]Poultry Research Institute, Rawalpindi, Pakistan, [4]Department of Biotechnology, Guangxi Veterinary Research Institute, Nanning, China, [5]Guangxi Key Laboratory of Veterinary Biotechnology, Nanning, China, [6]Key Laboratory of China (Guangxi)-ASEAN Cross-border Animal Disease Prevention and Control, Ministry of Agriculture and Rural Affairs of China, Nanning, China

Orthomyxoviruses are negative-sense, RNA viruses with segmented genomes that are highly unstable due to reassortment. The highly pathogenic avian influenza (HPAI) subtype H5N8 emerged in wild birds in China. Since its emergence, it has posed a significant threat to poultry and human health. Poultry meat is considered an inexpensive source of protein, but due to outbreaks of HPAI H5N8 from migratory birds in commercial flocks, the poultry meat industry has been facing severe financial crises. This review focuses on occasional epidemics that have damaged food security and poultry production across Europe, Eurasia, the Middle East, Africa, and America. HPAI H5N8 viral sequences have been retrieved from GISAID and analyzed. Virulent HPAI H5N8 belongs to clade 2.3.4.4b, Gs/GD lineage, and has been a threat to the poultry industry and the public in several countries since its first introduction. Continent-wide outbreaks have revealed that this virus is spreading globally. Thus, continuous sero- and viro-surveillance both in commercial and wild birds, and strict biosecurity reduces the risk of the HPAI virus appearing. Furthermore, homologous vaccination practices in commercial poultry need to be introduced to overcome the introduction of emergent strains. This review clearly indicates that HPAI H5N8 is a continuous threat to poultry and people and that further regional epidemiological studies are needed.

KEYWORDS

avian influenza virus, H5N8 subtype, epidemiology, surveillance, control and prevention

Introduction

Avian influenza viruses (AIVs) belong to the Orthomyxoviridae family and contain a segmented genome with eight single-stranded RNA segments and have negative polarity (Webster et al., 1992). Hemagglutinin (HA) gene and neuraminidase (NA) gene, two of the envelope proteins of these viruses, are used to classify them into different subtypes (Kawaoka et al., 1988). To date, 16 HA and 9 NA subtypes of AIVs have been identified in poultry and wild birds (Wang et al., 2022).

Low-pathogenic avian influenza (LPAI) viruses are naturally found in wild water birds such as swans, ducks, gulls, geese, swans, shorebirds, and terns (Krammer et al., 2018; Verhagen et al., 2021). LPAI viruses are transmitted to domestic birds, animals, and even humans from wild water birds. Influenza viruses with H5 HA have been circulating in wild birds and domestic poultry since 1995 (Harfoot and Webby, 2017). The Qinghai Lake-like H5N1 virus was first widely spread by migratory birds and caused huge damage to the poultry industry worldwide, but the origin of the virus remains unclear. The LPAI viruses of the H5 subtype, when infecting poultry, can evolve into HPAI viruses, causing severe mortality (Alexander and Brown, 2009). During July and August 2005, HPAI H5 clade 2.2 viruses were detected in poultry farms in Russia and Kazakhstan, where they caused high mortality (Coulombier et al., 2005). These viruses were genetically related to viruses detected in 2005 in Qinghai Lake in China (Chen et al., 2005). From July 2005 onward, HPAI H5 viruses were observed to cause outbreaks on poultry farms (Coulombier et al., 2005). The H5N1 virus became endemic in 2003 in Southern China, giving rise to several genotypes.

In mainland China, the H5N8 virus was detected in poultry between 2009 and 2010, which derived its HA gene from the Asian H5N1 lineage and its neuraminidase (NA), nucleoprotein (NP), and polymerase basic (PB1) genes from unidentified, non-H5N1 viruses. The H5N8 virus is highly pathogenic to chickens and moderately to extremely dangerous to mice (Zhao et al., 2013). In 2014, a novel reassortant HPAI H5N8 clade 2.3.3.4 virus with the HA gene was identified in South Korea (Jeong et al., 2014). Two types of H5N8 were found during these outbreaks, namely Gochang-like and Buan2-like. The predominant group, Buan2-like, afterward spread to Europe, East Asia, and North America by migratory waterfowl and formed three distinct subgroups (Jeong et al., 2014; Lee et al., 2014; Dalby and Iqbal, 2015; Lee et al., 2015; Verhagen et al., 2021). In autumn 2016, another High pathogenic AI H5N8 virus of clade 2.3.4.4 spread across different continents (Li et al., 2017) and showed sustained prevalence in Africa, Europe, and the Middle East (OIE, https://www.oie.int/en/animal-health-in-the-world/). In early 2020, HAPI H5N8 was continuously reported in Iraq, Kazakhstan, and Russia (Lewis et al., 2021). Furthermore, in December 2020 in Russia, seven poultry farm workers were infected with a clade 2.3.4.4b H5N8 virus (Pyankova et al., 2021). In June 2021, 2,782 outbreaks of H5N8 were reported, causing the mortality or destruction of approximately 38 million poultry in more than 25 countries (Cui et al., 2022).[1]

In conclusion, the spread of High pathogenic AI H5N8 viruses has raised serious issues for the security and conservation of animals, poultry, and even public health (Shi and Gao, 2021). All this evidence suggests that H5N8 viruses are likely to spread worldwide; therefore, continuous surveillance and vaccination of poultry are highly recommended. In this review, we describe the emergence of sporadic infection continentally, and the impacts are briefly described.

Intra and inter-continental transmission patterns of sporadic infection of HPAI H5N8

Asia and Africa

A number of emergence and re-emergence studies of HPAI H5N8 strains have been reported within & across Asia & Africa. One HPAI H5N8 virus (Dkk1203) was isolated from a poultry farm in mainland China during 2009–2010. The Dkk1203 isolate derived its HA gene from the Asian H5N1 lineage. Phylogenetic analysis of the HA gene revealed that this isolate was classified into the 2.3.4 clade. Compared to H5N5 viruses that were isolated between December 2008 and January 2009, this strain has longer branches. This strain was distantly related to Eurasian N8 genotype viruses and clustered with three H3N8 viruses with an origin in Eastern Asia. Therefore, the N5 and N8 NA genes of the Dkk1203 isolate are derived from Asian viruses; however, the exact origin is not known (Zhao et al., 2013).

In a breeding duck farm on January 16, 2014, in the Jeonbuk Province of South Korea, High-pathogenic AI clinical signs, such as reduced egg production by about 60% and slightly increased mortality rates, were discovered. Moreover, on January 17 of the same year, a farmer was also diagnosed with HPAI from breeder ducks in the Donglim Reservoir (Lee et al., 2014). Also, the Donglim Reservoir had 100 Baikal teal carcasses, all of which tested positive for the high pathogenic AI H5N8 virus (Lee et al., 2014).

A few months later, in April 2014, an outbreak of the HPAI virus with the genotype H5N8, A/chicken/Kumamoto/1–7/2014, occurred in Japan (Kanehira et al., 2015). The HA clade 2.3.4.4 membership of this virus was also made known. In particular, A/broiler duck/Korea/Buan2/2014 and A/baikal teal/Korea/Donglim3/2014, HPAI H5N8 that were isolated in Korea in January 2014, all eight genomic segments displayed substantial sequence similarity (Kanehira et al., 2015). The experimental work delineated that this isolate from Japan was lethal in chickens when a higher titer of virus was used for infection; however, the chickens were unaffected when challenged with lower viral doses (Kanehira et al., 2015).

In the same year (2014), three H5N8 viruses were reported from domestic geese in mainland China. The selected strains' sequence analyses revealed that all H5N8 viruses were direct progeny of the K1203 (H5N8)-like viruses discovered in China in 2010 and belonged to the Asian H5N1 HA lineage of clade 2.3.4.4. The recent common clade 2.3.4.4 H5N8 reassortants, which have severely damaged the poultry sector and pose a threat to public health, were created by K1203-like viruses, according to studies (Li et al., 2014).

Eight highly pathogenic H5N8 AIVs were discovered in Japan over the winter, particularly in a location where migratory birds overwinter. These isolates were divided into three groups based on genetic analysis, demonstrating that three genetic subgroups of H5N8 HPAIs circulated in these migratory birds. These findings also suggest that the migration of these birds next winter may result in the redistribution of H5N8 HPAI globally (Ozawa et al., 2015; Isoda et al., 2020).

In 2016 in Malard County of the Tehran Province and the Meighan wetland of Arak City, Markazi Province, the HA genes indicated categorization in the 2.3.4.4b subclade. Although being identified as an H5N8 2.3.4.4b virus, the A/Goose/Iran/180/2016 virus's cluster was split from the A/Chicken/Iran/162/2016 virus. This suggests that the entry of these viruses in Iran occurred through more than one window. The most recent HPAI-H5 outbreak in Iran happened in 2015 and was entirely caused by viruses from clade 2.3.2.1c. These findings underscore the necessity to continue proper monitoring activities in the target wild and domestic bird species for early HPAI identification and show that Iran is at high risk of the importation of HPAI H5 of the A/Goose/Guangdong/1/1996 lineage from East Asia. These activities would also allow the study of the genetic and antigenic evolution of H5 HPAI clade 2.3.4.4.viruses in the region and the world (Ghafouri et al., 2017). Furthermore, it appears that migrating wild aquatic birds carried these HPAI H5N8 strains into Iran *via* the West Asia-East African flyway (Motahhar et al., 2016).

An H5N8 influenza virus of clade 2.3.4.4 outbreak was reported in 2016 in the Republic of Tyva. The H5N8 clade 2.3.4.4 virus spread

---

[1] http://empres-i.fao.org/eipws3g/

over Europe in the fall. The reports provide a clear overview of the viral strains that were discovered in the Russian Federation during the spring and fall of 2016. The strains under investigation were extremely harmful to mice, and several of their antigenic and genetic characteristics were different from an H5N8 strain that was prevalent in Russia in 2014 (Marchenko et al., 2017).

The newly emerged H5N8 influenza virus was also isolated from green-winged teal ducks. The genomes of the HPAI H5N8 viruses from Egypt were also found to be related to recently identified reassortant H5N8 viruses of clade 2.3.4.4 recovered from several Eurasian nations, according to analyses of the viruses' genomes. The Egyptian H5N8 viruses had a number of genetic shifts that likely allowed for the spread and virulence of these viruses in mammals. Instead of human-like receptors, Egyptian H5N8 viruses prefer to bind to avian-like receptors. Likewise, amantadine and neuraminidase inhibitors had little effect on the Egyptian H5N8 viruses. It is important to continue monitoring waterfowl for avian influenza because it provides early warning of specific dangers to poultry and human health (Kandeil et al., 2017). The presence of this group and clade was also found in Qinghai Lake, China, in 2016, which resulted in the deaths of wild migratory birds (Li et al., 2017). An HPAI H5N8 virus of clade 2.3.4.4b has been detected in Egypt. PA and NP gene replacement identified the strain as A/duck/Egypt/F446/2017. The Russian 2016 HPAI H5N8 virus (A/great crested grebe/Uvs-Nuur Lake/341/2016 (H5N8)) was likely the source of Egyptian H5N8 viruses, according to Bayesian phylogeographic analysis and reassortment most likely took place prior to an incursion into Egypt (Yehia et al., 2018).

In Egypt, multiple introductions of different reassorted strains have been observed. The antigenic sites A and E of the HA gene have two new mutations. With various vaccination seeds, the HA nucleotide sequence identity ranges from 77 to 90%. To determine the main reassorted strain in Egypt, full-genome sequence analysis representing various governorates and sectors has been conducted. All viruses have been shown to be identical to the clade 2.3.4.4b reassorted strain that was discovered in Germany and other nations. Examination of these viruses revealed changes unique to Egyptian strains rather than the original virus identified in 2017 (A/duck/Egypt/F446/2017), and two strains of these viruses had the novel antiviral resistance marker V27A, which indicated amantadine resistance in the M2 protein. The findings showed that circulating H5N8 viruses were more variable than prior viruses analyzed in 2016 and 2017. An early 2017 strain served as the foundation for the main reassorted virus that circulated in 2017 and 2018. To track the development of circulating viruses, it is crucial to keep up this surveillance of AIVs (Yehia et al., 2020). The Democratic Republic of the Congo strains also belongs to the same clade, 2.3.4.4B. The emergence of this clade in central Africa threatens animal health and food security (Twabela et al., 2017).

The recovered HPAI A(H5N8) viruses in Pakistan during 2018–19 belonged to clade 2.3.4.4b and were most closely related to the Saudi Arabian A(H5N8) viruses, which were most likely introduced *via* cross-border transmission from nearby regions about 3 months before the virus was discovered in domestic poultry. It was also found that, prior to the first human A(H5N8) infection in Russian poultry workers in 2020, clade 2.3.4.4b viruses underwent rapid lineage expansion in 2017 and acquired significant amino acid mutations, including mutations correlated with increased hemagglutinin affinity to human-2,6 receptors. Our findings demonstrate the necessity of routine avian influenza surveillance in Pakistan's live bird markets in order to keep an eye out for any potential A(H5Nx) variants that might emerge from poultry populations (Ali et al., 2021). Every year, the Indus Flyway, also known as the Green Way, transports between 0.7 and 1.2 million birds from Europe, Central Asian countries, and India to Pakistan. [International Visitors: Birds Come Flying In. http://www.wildlifeofpakistan.com/PakistanBirdClub/birdcomeflyingin.html].

A thorough investigation was conducted to track the evolution of influenza viruses in poultry during the years 2020–2022 in China. A total of 35 influenza viruses, including 30 H5N8 viruses, 3 H5N1 viruses, and 2 H5N6 viruses, were isolated from chickens, ducks, and geese. The internal genes of H5N1 and H5N6 viruses shared different genetic heterogeneity with H5N8 viruses and had been reassorted with wild bird-origin H5N1 viruses from Europe. All HP H5N8 isolates were derived from clade 2.3.4.4b. The fact that practically all H5N8 viruses in China and Korea showed just one phylogenic cluster with H5N8 viruses of wild bird origin suggests that the H5N8 viruses in China were more stable. We also discovered that the main geographic source for the transmission of these H5N8 viruses to northern and eastern China is Korea. The majority of the co-circulation of H5N8 viruses took place within China, with central China serving as a seeding population during the H5N8 epidemic. Strong statistical evidence supported viral migration from wild birds to chickens and ducks, demonstrating that during 2020–2021, 2.3.4.4b H5N8 viruses with poultry origins were borne by wild birds. Multiple gene segments were also discovered to be involved in the development of severe disease due to H5N8 HPAI viruses, in mallards birds, which explains why no viral gene was found to be solely responsible for reducing the high virulence of an H5N8 virus but the PB2, M and NP segments significantly decreased mortality Our results give new insights into the dynamics of H5 subtype influenza virus evolution and transmission among poultry following the almost one-year invasion of China by novel H5N8 viruses (Leyson et al., 2021; Ye et al., 2022). In China, the re-emergence of the High Pathogenic H5N8 virus in domestic geese was also reported (Guo et al., 2021).

The establishment of novel H5N8 strains in China is frequently linked to the migration of migratory birds *via* the East Asian-Australasian Flyway. This flyway connects Siberia to Australia and includes various stopover spots in China where wild birds gather throughout their annual migration. These locations allow diverse bird species to interact and exchange influenza viruses (Li et al., 2022).

During, May 2020 in Iraq, H5N8 was reported in poultry. Complete genome sequencing delineated that a noval H5 2.3.3.4b variant had emerged. Furthermore, the long branch lengths for all segments indicated that undetected isolate was circulating for some period and possibly in galliform poultry (Lewis et al., 2021).

After outbreaks in Iraq in July 2020, H5N8 was detected in ducks, geese, and backyard chickens of Chelyabinskaya Oblast (Chelyabinsk), in southern central Russia. During August and September 2020, a total of 11 cases were detected in the Tyumen, Omsk, and Kurgan regions of Russia (Lewis et al., 2021). Wild birds were described as the cause of the incursion.

Concurrent with the H5N8 outbreak in Russia, the outbreak of H5N8 was also confirmed in several regions of Kazakhstan, including Kostanay, Akmola, and Pavlodar (Lewis et al., 2021). AI H5N8 diagnosis was confirmed by subtype-specific quantitative RT–PCR (Nagy et al., 2021). The AI H5N8 virus from Iraq and Kazakhstan shared a lot of genetic similarities, according to genetic analyses (Lewis et al., 2021).

Europe and the Americas

In 2014, European countries, such as Germany, the United Kingdom, the Netherlands, and Italy, reported several outbreaks of H5N8 in poultry. Two different Highly pathogenic viruses, H5N2 and H5N8, were found in the United States in December 2014 in wild birds and later in backyard birds in Washington State. This sparked concerns about potential connections with recent H5N2 outbreaks in Canada and H5N8 in Asia, which is now affecting poultry farms in Europe. The continuous spread of these Eurasian HPAI H5 viruses among wild birds has a significant impact that could arise and the ensuing consequences on American poultry and wildlife rehabilitation facilities. Tundra swans (*C. columbianus*), c ommon teal (*A. crecca*), spot-billed duck (*A. poecilorhyncha*), Eurasian wigeon (*A. penelope*) and mallard, that appeared to be in good health also tested positive for the HPAI H5N8 virus, which raises the possibility that wild birds may be contributing to the spread of this High Pathogenic H5 lineage in North America (Ip et al., 2014).

With a comprehensive review of the spatiotemporal expansion and genetic characteristics of HPAI Gs/GD H5N8 from Poland's 2019/20 epidemic, the Highly pathogenic H5 subtype of the Gs/GD lineage repeatedly invaded Poland from 2016 to 2020, posing a major threat to poultry globally. In nine Polish provinces during 2019 and 2020, 35 outbreaks in backyard and commercial poultry holdings as well as 1 incidence in a wild bird were confirmed. The majority of the outbreaks were found in the meat of ducks and turkeys. All sequenced viruses belonged to a previously unidentified genotype of HPAI H5N8 clade 2.3.4.4b and were closely related to one another. The main methods of HPAI dissemination were found to be human activity and wild birds. A review of current risk assessment techniques is necessary in light of the HPAI virus's unusually delayed emergence (Shin et al., 2019; Śmietanka et al., 2020).

Asia and Europe

A new wave of H5N8 outbreaks in domestic and wild birds was observed in several European nations in October 2020, including the United Kingdom, Denmark, Ireland, Germany, and the Netherlands. In August 2020, several outbreaks of the disease were confirmed from Russia in both domestic and wild birds, and the affected regions spread to Kazakhstan in mid-September. Moreover, H5N8 epidemics in domestic and/or wild birds appeared in East Asia (Japan and South Korea) and the Middle East (Israel). A unique variant between clade 2.3.4.4b and Eurasian LPAI viruses in wild birds was described as well as two different forms of HPAI H5N8 variants, one of which only belonged to clade 2.3.4.4b. The geographical areas affected have been steadily expanding, and at least 46 nations have documented highly pathogenic H5N8. with one of the human cases being related to poultry workers during an outbreak in poultry (Pyankova et al., 2021).

An influenza A (H5N8) clade 2.3.4.4b strain was recovered from a poultry worker during an outbreak of highly pathogenic H5N8 in chickens at a poultry farm in the Astrakhan region on the Volga River in southern Russia in December 2020, according to a study of a similar nature. Nasopharyngeal swabs were collected from seven poultry workers that tested positive, and two were confirmed by RT–PCR and sequencing. The seven individuals, five of whom were female and two of whom were male, ranged in age from 29 to 60. The HA gene of all five viruses obtained from birds and one from humans shared a significant degree of genetic similarity with other clades. From 2016 to 2021, viruses with the 2.3.4.4b gene were found in wild and domestic birds in Russia. human influenza A in some cases (H5) 2.3.4.4. A potential public health hazard is infections (Pyankova et al., 2021).

H5N8 clade 2.3.4.4b outbreaks were observed in Russia, the Middle East, Central Europe, and Ukraine in 2016. In the southern part of Ukraine, close to areas where migrating waterfowl congregate in large numbers, especially mute swans (*Cygnus olor*), an outbreak of HPAI strains was documented in domestic backyard poultry between 2016 and 2017. Upon sequence analysis, it was found that 2 novel H5N8 HPAI strains were isolated from domestic backyard chickens (*Gallus gallus*) and mallard duck (*Anasplatyrhynchos*). HPAI outbreaks in Ukraine underscore the ongoing need for AIV bio-monitoring, genomic sequencing, and mapping of wild bird flyways and their contacts with domestic poultry in Eurasia (Sapachova et al., 2021).

Long-distance migratory birds can play a significant role in the global spread of avian influenza viruses, notably through nesting regions in the sub-arctic. The investigation of H5N8 viral sequences, epidemiological studies, waterfowl migration, and chicken trade all revealed that wild birds can spread the virus to poultry *via* contact with infected water or surfaces. Furthermore, the chicken trade may contribute to the virus's spread. Clade 2.3.4.4 viral hemagglutinin was discovered to be extraordinarily promiscuous, producing reassortants with diverse subtypes and potentially boosting its ability to infect different species of birds and mammals. This promiscuity is likely to have a role in its ability to quickly adapt to various hosts and settings, potentially enhancing its pandemic potential (Lycett et al., 2009).

H5N8 evolution

Whole genome

Gammaviruses are characterized as low pathogenic (LP) viruses or highly pathogenic (HP) viruses based on virulence in chickens. HP viruses may emerge from LP viruses through genetic mutations in wild birds (Fouchier et al., 2005). In this context, AIV subtypes H5 and H7 are characterized as HP viruses. To date, AI viruses have 16 subtypes on the basis of the Hemagglutinin gene and 9 due to the Neuraminidase gene (Webster et al., 1992; Fouchier et al., 2005).

The entire genome of HPAI H5N8, is made up of eight single-stranded RNA segments. Each segment encodes a distinct gene that is essential for the virus's replication and infection. Polymerase Basic Protein 2 (PB2), which is roughly 2,341 nucleotides long and encodes the PB2 protein, is one of these segments. The Polymerase Basic Protein 1 (PB1) gene is approximately 2,341 nucleotides long and codes for the PB1 protein. The Polymerase Acidic Protein (PA) gene encodes the PA protein and is approximately 2,234 nucleotides long. The Hemagglutinin (HA) gene encodes the HA protein and is approximately 1,778 nucleotides long. The Nucleoprotein (NP) gene has a length of about 1,565 nucleotides and codes for the NP protein.

The Neuraminidase (NA) gene encodes the NA protein and is approximately 1,413 nucleotides long. The Matrix (M) gene encodes the M1 and M2 proteins and is approximately 1,027 nucleotides long. The Non-structural protein (NS) gene has around 890 nucleotides and encodes the NS1 and NS2 proteins. It is crucial to note that the lengths provided are approximations and may differ slightly across various H5N8 strains or isolates (Bouvier and Palese, 2008).

Hemagglutinin gene (HA)

HA gene sequence analysis was performed, and a phylogenetic tree was constructed by comparing sequences retrieved from the GISAID platform.[2] These HPAI H5 strains belong to different groups and lineages. Sequence analysis was performed by following H5 numbering, which uncovered the genetic diversity during evolution. The cleavage site motif of HPAI H5 includes the polybasic amino acids QGERRRKKR*GLF (Perdue et al., 1997; Siddique et al., 2012), whereas in the selected isolates reported globally during different years, maximum HPAI H5N8 evolved, and the cleavage site became LREKRRKKR*GLF. Studies have demonstrated that, although HPAI H5N8 attaches to avian-like receptors, it may also attach to human virus-like receptors in the human respiratory tract. HPAI showed more affinity for cats than dogs, which were more susceptible to HPAI. It is suggested that, due to its establishment in ducts, the transmission of HPAI H5N8 viruses may modify the genetic evolution of preexisting avian poultry strains (Kim et al., 2014).

On the basis of similarity, H5N8 viruses evolved into three groups (Li et al., 2014). Groups I and II contain the isolates belonging to clade 2.3.4.4b and the Eurasian continent, whereas group III contains isolates from the North American lineage, with apparent divergence from those in groups I and II. Moreover, the transmission pattern of this subtype was observed in depth by reviewing the continent wide distribution in Africa (A), Asia (B & C), Europe (D), North America and Oceania (E). In this regard, HA gene sequences of selected HPAI H5N8 viruses were retrieved from the GISAID database. Initially, Bayesian evolutionary analysis was performed using BEAST version 1.10.4, and then FigTree software (v1.4.4) was used for phylogenetic tree construction, as shown (Figure 1). Moreover, No isolation has been reported from Antarctica or South America. These continent-wide sporadic infection, further clarify that the domestic birds are reassortant hosts for the emergence of novel virus subtypes and are thought to be the reservoir of AIV. The spread of these viruses could endanger the health of both humans and birds.

In addition, asparagine-linked glycosylation sites have been observed among HPAI H5 strains, revealing that some are common during evolution, whereas a number of substitutions and deletions are also seen. Siddique et al. in 2012 reported the same sites along with additional glycosylation sites at the globular head of the HA gene, which is responsible for the prediction of high efficiency of replication (Bender et al., 1999; Siddique et al., 2012). Moreover, the conserved amino acids at positions 222 glutamine and glycine at position 224 of the HA gene are responsible for avian-like receptors at the binding site that is common among all the HPAI H5 proteins selected for analysis, and similar reports are available in this context (Matrosovich et al., 1999; Smith et al., 2006; Siddique et al., 2012).

Furthermore, a number of amino acid mutations have been observed at antigenic sites, including at amino acid position 39, where glutamic acid has been shown to have mutated into glycine, S141P, K169R, D171N, A172T, R178I/R, P197S, R205N/K, and N268Y. These sites have been designated as crucial residues of the antigenic site (Kaverin et al., 2004).

In NA, PB1, PB2, PA, NP, PA, M, and NS, almost 29 molecular signatures are present that are associated with replication, virulence, transmission, and adaptation in mammals (Hiromoto et al., 2000; Shaw et al., 2001; Chen et al., 2007; Gabriel et al., 2008; Long et al., 2008; Lycett et al., 2009; Spesock et al., 2011; Hui et al., 2017; Kamal et al., 2017; Yu et al., 2017; Pulit-Penaloza et al., 2020). In this regard, a maximum of 20molecular signatures were present in HPAI H5N1/483, whereas 4–6 were present in HPAI/LPAI H5N8 viruses. The PB2 gene contains the known marker 627 K for mammalian adaptation that has only been shown to be present in 2 HPAI H5 human isolates, HPAI subtype H5N1/483 and H5N6/39715. There are a number of other mutations in the NA gene at the 96A amino acid position and the Matrix 2 gene at the S31N site that are responsible for dual resistance against antivirals, including oseltamivir and amantadine (Cheung et al., 2006; Ilyushina et al., 2010). However, some other mutations, such as R118K in the NA gene, are associated with additional resistance to zanamivir (Intharathep et al., 2008; Orozovic et al., 2011). Due to these genetic changes, adamantanes and neuraminidase inhibitors may not be able to effectively prevent the replication of these viruses in the host in this situation.

The highly pathogenic AI H5 subtype has been spreading at an unprecedented rate since 2021, which is concerning given the disease's high mortality rate in wild birds and poultry as well as cases that have been observed in mammals and people. This could potentially lead to a future pandemic. Along with causing mass demise in a number of wild mammal species, H5 HPAI has the capacity to switch from infecting avian to mammalian hosts and develop the necessary characteristics for effective transmission from mammal to mammal. Therefore, enhanced surveillance of wild animals, large-scale animal farms, and humans handling them is urgently needed, along with improved biosecurity measures, reduction of poultry farm size and density, vaccination of poultry against HPAI, and avoidance of areas rich in water birds as a location for poultry farms. In addition, the medical sector and society need to prepare for the emergence of the human-to-human spread of H5 HPAI. It is crucial to include the community, communicate about risks, and counter intentional disinformation. The next pandemic, which could result from this AIV, should be prepared for using the lessons learned from the COVID-19 pandemic as a reference (Kuiken, 2023).

Risk assessment and mitigation strategies

During 2020–21, in Eurasia, Europe, and Africa, emergent strains were highly pathogenic subtypes of H5N8 belonging to clade 2.3.4.4b and had a significant impact on the poultry industry. In the current scenario, an emergency has been declared for the enhancement of sero-and viro-surveillance across the globe depending on the previous outbreaks in 2005 and 2016 (Alarcon et al., 2018; Adlhoch et al., 2020). For risk mitigation strategies, an effective risk assessment needs to

---

[2] http://platform.gisaid.org/

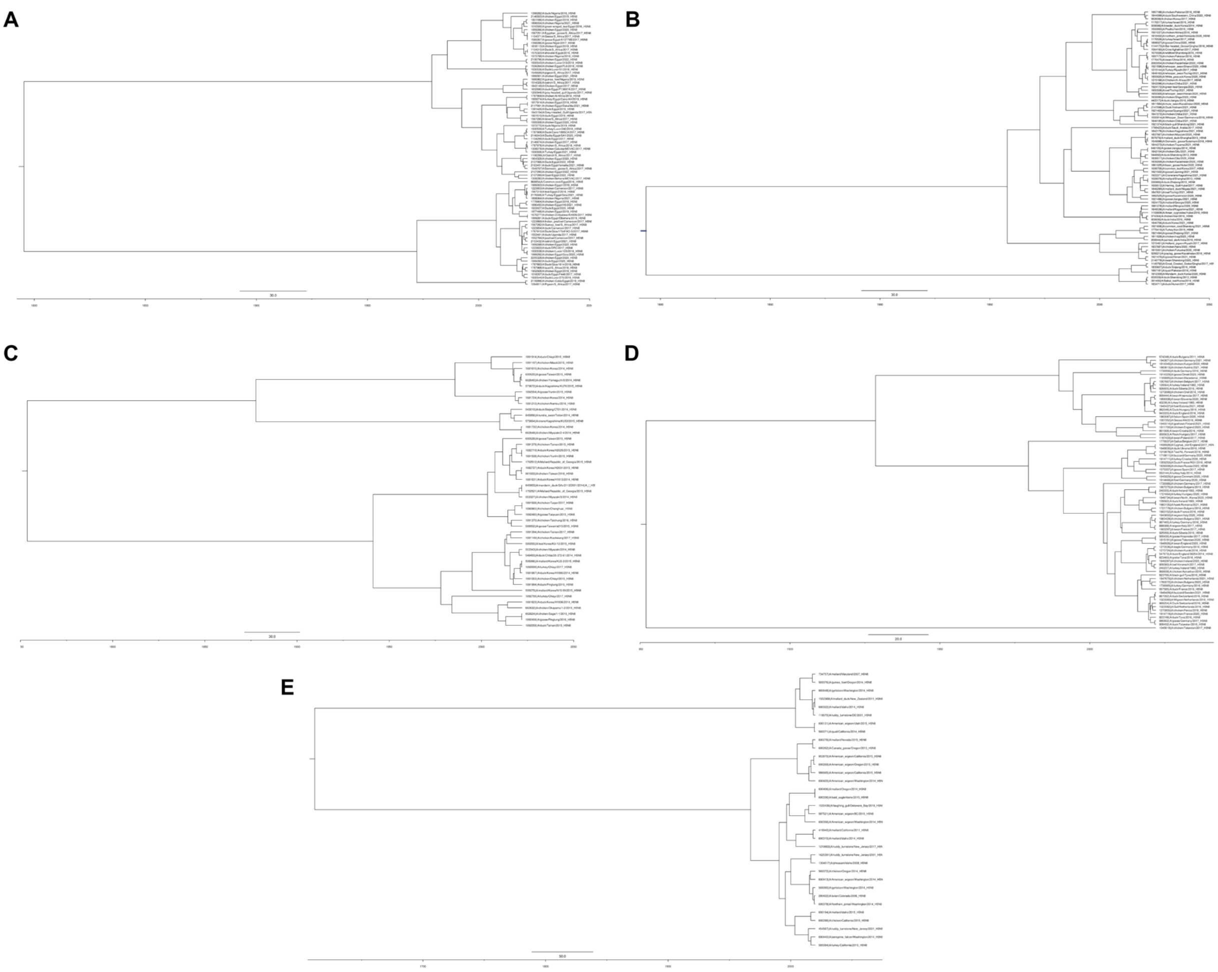


FIGURE 1
Phylogenetic analysis of the hemagglutinin gene of AIV subtype H5N8 inferred with BEAST software **(A)** Africa **(B)** Asia-group B **(C)** Asia-group C **(D)** Europe **(E)** North America and Oceania.

be performed in terms of tissue/host tropism, pathogenesis, and disease transmission and dissemination. Influenza A virus poses a continuous threat to poultry and the public due to its evolutionary mechanism through reassortment.

HPAI are extremely risky to poultry if not properly vaccinated. The low pathogenic H7N9 virus which emerged in 2013, was converted into high pathogenic due to mutations in early 2017, caused the death of millions of chickens to control the outbreak (Shi et al., 2017; Zeng et al., 2018). The use of H7N9 vaccines effectively controlled the circulation of this virus both in poultry and humans (Zeng et al., 2020). Since 2004 vaccines are in in use against H5 avian viruses in China (Zeng et al., 2020). Since the emergence of H7N9 in 2017, a bivalent inactivated vaccine against H5/H7 was developed to control both H5 and H7 in poultry in China (Shi et al., 2018; Zeng et al., 2018). Currently, a trivalent vaccine-H5/H7 which contain Re-11, Re-12 and H7-Re3 vaccine seed viruses are in use. This trivalent vaccine was generated by reverse genetics, and HA genes were derived from A/duck/Guizhou/S4184/2017(H5N6) (DK/GZ/S4184/17) (a clade 2.3.4.4 h virus), A/chicken/Liaoning/SD007/2017 (H5N1) (CK/LN/SD007/17) (a clade 2.3.2.1d virus), and A/chicken/Inner Mongolia/SD010/2019 (H7N9) (CK/IM/SD010/2019), respectively (Zeng et al., 2020; Cui et al., 2022). Although the newly emerged H5N8 viruses differ antigenically from currently used vaccines, poultry birds vaccinated in routine with current vaccines still completely protect against H5N8 virus challenge (Cui et al., 2022). In another recent study (Niqueux et al., 2023), the efficacy of three vaccines was determined against the HPAI A/decoy duck/France/161105a/2016 (H5N8), clade 2.3.3.4b. The first vaccine (Vac1), was derived from HA gene clade 2.3.4.4b A (H5N8) HPAI, the second vaccine (Vac2) used was a commercial bivalent adjuvanted vaccine that contained an expressed HA modified from clade 2.3.2 A (H5N1) HPAI. The third vaccine (Vac3) also incorporated a homologous 2.3.4.4b H5 HA gene. Vac2 partly decreased the respiratory and intestinal excretion of challenge virus, Vac3 completely abolished cloacal shedding while Vac1 abolished oropharyngeal and cloacal shedding to almost undetectable levels. These results provided significant insights in the immunogenicity of recombinant H5 vaccines in mule ducks (Niqueux et al., 2023). Since the H5N8 viruses have been detected in a wide range of wild birds across the globe, therefore it could spread worldwide and can be very lethal to poultry. Therefore, homologous vaccination practices need to be introduced for the control and transmission of the disease, as the exact information on the disease and transmission is still not clear. The Iraqi-like strains are dispersed through poultry or indirect transmission in central Asia. In 2014–2017, there was little evidence of reassortment of HPAI H5N8 and H5N1 viruses in wild birds, as dispersal was unclear, but later, evidence of reassortment was found to be substantive, whereas in Europe in 2020, the emerging HPAI H5N8 strain was clearly a combination of sub-Saharan African viruses with a Eurasian LPAIV origin. Despite the implementation of biosecurity measures, several outbreaks of HPAI H5N8 strains were reported in France during 2016–2017, possibly due to airborne viral transmission. The area around the poultry facilities, almost 50–110 m, is considered contaminated with varied viral concentrations (Scoizec et al., 2018). In case of outbreaks, depopulation methods need to be wisely implemented to further control the air-borne contamination of influenza viruses, which could result in instant mass culling.

Conclusions and future perspectives

This study backs up the hypothesis that asymptomatic migrating birds may have assisted viral development and reassortment as well as regional transmission of HPAI subtype H5N8. Another evidence that rapid and active mutation and reassortment of H5 subtypes may occur in these hosts comes from the HPAI subtypes H5N1 and H5N8 coinfecting and cocirculating in migratory ducks. Therefore, intersectoral alliance and coaction for mitigating avian influenza outbreaks based on the One Health approach that is worthwhile and advisable. This review discusses knowledge of the disease's nature, distribution, epidemiology, applied surveillance techniques, diagnosis, and control approaches as they related to Sahelian Africa and its surrounding suburbs. Understanding of the influenza virus and its footprint on the well-being of humans and animals would aid in better preparing for the erratic/capricious challenges posed by this infectious disease.

Continuous vigilance, strengthening biosecurity, and intensifying surveillance in wild birds are needed to better manage the risk of HPAI occurrence in the future. Moreover, high-risk countries should vaccinate their poultry birds to prevent further outbreaks of HPAI H5N8. This review clearly indicates that HPAI H5N8 is a threat from a poultry standpoint and public perspective and that continuous surveillance and further epidemiological studies are needed.

Author contributions

ZX: conceptualization, supervision, and funding acquisition. SR: wrote the manuscript. FR, SM, and AA: edited and proof read the manuscript. ML, SL, and LX: collected data and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This project was and funded by grants from Guangxi Science and Technology Project (no. AB21076004), Guangxi BaGui Scholars Program Foundation (2019A50).

Acknowledgments

This study uses data from the Centre for Health Protection websites as well as the GISAID database, the authors thank everyone who contributed to the study by collecting and sharing of data.

Conflict of interest

SR and SM were employed by Sadiq Poultry Pvt. Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Adlhoch, C., Fusaro, A., Kuiken, T., Niqueux, E., Staubach, C., Terregino, C., et al. (2020). Avian influenza overview February–may 2020. *EFSA J.* 18:e06194. doi: 10.2903/j.efsa.2020.6194

Alarcon, P., Brouwer, A., Venkatesh, D., Duncan, D., Dovas, C. I., Georgiades, G., et al. (2018). Comparison of 2016–17 and previous epizootics of highly pathogenic avian influenza H5 Guangdong lineage in Europe. *Emerg. Infect. Dis.* 24, 2270–2283. doi: 10.3201/eid2412.171860

Alexander, D. J., and Brown, I. H. (2009). History of highly pathogenic avian influenza. *Rev. Sci. Tech.* 28, 19–38. doi: 10.20506/rst.28.1.1856

Ali, M., Yaqub, T., Shahid, M. F., Wong, F. Y., Mukhtar, N., Naeem, M., et al. (2021). Genetic characterization of highly pathogenic avian influenza a(H5N8) virus in Pakistani live bird markets reveals rapid diversification of clade 2.3.4.4b viruses. *Viruses* 13:1633. doi: 10.3390/v13081633

Bender, C., Hall, H., Huang, J., Klimov, A., Cox, N., Hay, A., et al. (1999). Characterization of the surface proteins of influenza a (H5N1) viruses isolated from humans in 1997–1998. *Virology* 254, 115–123. doi: 10.1006/viro.1998.9529

Bouvier, N. M., and Palese, P. (2008). The biology of influenza viruses. *Vaccine* 26, D49–D53. doi: 10.1016/j.vaccine.2008.07.039

Chen, H., Bright, R. A., Subbarao, K., Smith, C., Cox, N. J., Katz, J. M., et al. (2007). Polygenic virulence factors involved in pathogenesis of 1997 Hong Kong H5N1 influenza viruses in mice. *Virus Res.* 128, 159–163. doi: 10.1016/j.virusres.2007.04.017

Chen, H., Smith, G. J. D., Zhang, S. Y., Qin, K., Wang, J., Li, K. S., et al. (2005). H5N1 virus outbreak in migratory waterfowl. *Nature* 436, 191–192. doi: 10.1038/nature03974

Cheung, C. L., Rayner, J. M., Smith, G. J. D., Wang, P., Naipospos, T. S. P., Zhang, J., et al. (2006). Distribution of amantadine-resistant H5N1 avian influenza variants in Asia. *J. Infect. Dis.* 193, 1626–1629. doi: 10.1086/504723

Coulombier, D., Paget, W. J., Meijer, A., and Ganter, B. (2005). Highly pathogenic avian influenza reported to be spreading into western Russia. *Weekly Releases (1997–2007)* 10:E050818.1. doi: 10.2807/esw.10.33.02776-en

Cui, P., Zeng, X., Li, X., Li, Y., Shi, J., Zhao, C., et al. (2022). Genetic and biological characteristics of the globally circulating H5N8 avian influenza viruses and the protective efficacy offered by the poultry vaccine currently used in China. *Sci. China Life Sci.* 65, 795–808. doi: 10.1007/s11427-021-2025-y

Dalby, A. R., and Iqbal, M. (2015). The European and Japanese outbreaks of H5N8 derive from a single source population providing evidence for the dispersal along the long distance bird migratory flyways. *PeerJ* 3:e934. doi: 10.7717/peerj.934

Fouchier, R. A. M., Munster, V., Wallensten, A., Bestebroer, T. M., Herfst, S., Smith, D., et al. (2005). Characterization of a novel influenza a virus hemagglutinin subtype (H16) obtained from black-headed gulls. *J. Virol.* 79, 2814–2822. doi: 10.1128/jvi.79.5.2814-2822.2005

Gabriel, G., Herwig, A., and Klenk, H.-D. (2008). Interaction of polymerase subunit PB2 and NP with importin α1 is a determinant of host range of influenza a virus. *PLoS Pathog.* 4:e11. doi: 10.1371/journal.ppat.0040011

Ghafouri, S. A., GhalyanchiLangeroudi, A., Maghsoudloo, H., KH Farahani, R., Abdollahi, H., Tehrani, F., et al. (2017). Clade 2.3.4.4 avian influenza a (H5N8) outbreak in commercial poultry, Iran, 2016: the first report and update data. *Trop. Anim. Health Prod.* 49, 1089–1093. doi: 10.1007/s11250-017-1302-z

Guo, J., Yu, H., Wang, C., Yuan, S., Sun, H., Yang, J., et al. (2021). Re-emergence of highly pathogenic avian influenza a(H5N8) virus in domestic goose, China. *J. Infect.* 83, 709–737. doi: 10.1016/j.jinf.2021.10.006

Harfoot, R., and Webby, R. J. (2017). H5 influenza, a global update. *J. Microbiol.* 55, 196–203. doi: 10.1007/s12275-017-7062-7

Hiromoto, Y., Yamazaki, Y., Fukushima, T., Saito, T., Lindstrom, S. E., Omoe, K., et al. (2000). Evolutionary characterization of the six internal genes of H5N1 human influenza a virus. *Microbiology* 81, 1293–1303. doi: 10.1099/0022-1317-81-5-1293

Hui, K. P. Y., Chan, L. L. Y., Kuok, D. I. T., Mok, C. K. P., Yang, Z.-F., Li, R.-F., et al. (2017). Tropism and innate host responses of influenza a/H5N6 virus: an analysis of ex vivo and in vitro cultures of the human respiratory tract. *Eur. Respir. J.* 49:1601710. doi: 10.1183/13993003.01710-2016

Ilyushina, N. A., Seiler, J. P., Rehg, J. E., Webster, R. G., and Govorkova, E. A. (2010). Effect of neuraminidase inhibitor–resistant mutations on pathogenicity of clade 2.2 a/Turkey/15/06 (H5N1) influenza virus in ferrets. *PLoS Pathog.* 6:e1000933. doi: 10.1371/journal.ppat.1000933

Intharathep, P., Laohpongspaisan, C., Rungrotmongkol, T., Loisruangsin, A., Malaisree, M., Decha, P., et al. (2008). How amantadine and rimantadine inhibit proton transport in the M2 protein channel. *J. Mol. Graph. Model.* 27, 342–348. doi: 10.1016/j.jmgm.2008.06.002

Ip, H. S., Torchetti, M. K., Crespo, R., Kohrs, P., DeBruyn, P., Mansfield, K. G., et al. (2014). Novel Eurasian highly pathogenic avian influenza a H5 viruses in wild birds, Washington, USA. *Emerg. Infect. Dis.* 21, 886–890. doi: 10.3201/eid2105.142020

Isoda, N., Twabela, A. T., Bazarragchaa, E., Ogasawara, K., Hayashi, H., Wang, Z.-J., et al. (2020). Re-invasion of H5N8 high pathogenicity avian influenza virus clade 2.3.4.4b in Hokkaido, Japan, 2020. *Viruses* 12:1439. doi: 10.3390/v12121439

Jeong, J., Kang, H.-M., Lee, E.-K., Song, B.-M., Kwon, Y.-K., Kim, H.-R., et al. (2014). Highly pathogenic avian influenza virus (H5N8) in domestic poultry and its relationship with migratory birds in South Korea during 2014. *Vet. Microbiol.* 173, 249–257. doi: 10.1016/j.vetmic.2014.08.002

Kamal, R., Alymova, I., and York, I. (2017). Evolution and virulence of influenza a virus protein PB1-F2. *Int. J. Mol. Sci.* 19:96. doi: 10.3390/ijms19010096

Kandeil, A., Kayed, A., Moatasim, Y., Webby, R. J., McKenzie, P. P., Kayali, G., et al. (2017). Genetic characterization of highly pathogenic avian influenza a H5N8 viruses isolated from wild birds in Egypt. *J. Gen. Virol.* 98, 1573–1586. doi: 10.1099/jgv.0.000847

Kanehira, K., Uchida, Y., Takemae, N., Hikono, H., Tsunekuni, R., and Saito, T. (2015). Characterization of an H5N8 influenza a virus isolated from chickens during an outbreak of severe avian influenza in Japan in April 2014. *Arch. Virol.* 160, 1629–1643. doi: 10.1007/s00705-015-2428-9

Kaverin, N. V., Rudneva, I. A., Ilyushina, N. A., Lipatov, A. S., Krauss, S., and Webster, R. G. (2004). Structural differences among hemagglutinins of influenza a virus subtypes are reflected in their antigenic architecture: analysis of H9 escape mutants. *J. Virol.* 78, 240–249. doi: 10.1128/jvi.78.1.240-249.2004

Kawaoka, Y., Chambers, T. M., Sladen, W. L., and Gwebster, R. (1988). Is the gene pool of influenza viruses in shorebirds and gulls different from that in wild ducks? *Virology* 163, 247–250. doi: 10.1016/0042-6822(88)90260-7

Kim, Y.-I., Pascua, P. N. Q., Kwon, H.-I., Lim, G.-J., Kim, E.-H., Yoon, S.-W., et al. (2014). Pathobiological features of a novel, highly pathogenic avian influenza a(H5N8) virus, Emerg. *Microbes Infect.* 3, e75–e13. doi: 10.1038/emi.2014.75

Krammer, F., Smith, G. J. D., Fouchier, R. A. M., Peiris, M., Kedzierska, K., Doherty, P. C., et al. (2018). Influenza. *Nat. Rev. Dis. Primers.* 4:3. doi: 10.1038/s41572-018-0002-y

Kuiken, T. (2023). Ron a M Fouchier, Marion P G Koopmans, being ready for the next influenza pandemic? *Lancet Infect. Dis.* 23, 398–399. doi: 10.1016/S1473-3099(23)00117-2

Lee, Y.-J., Kang, H.-M., Lee, E.-K., Song, B.-M., Jeong, J., Kwon, Y.-K., et al. (2014). Novel reassortant influenza a(H5N8) viruses, South Korea. *Emerg. Infect. Dis.* 20, 1086–1089. doi: 10.3201/eid2006.140233

Lee, D.-H., Torchetti, M. K., Winker, K., Ip, H. S., Song, C.-S., and Swayne, D. E. (2015). Intercontinental spread of Asian-origin H5N8 to North America through Beringia by migratory birds. *J. Virol.* 89, 6521–6524. doi: 10.1128/jvi.00728-15

Lewis, N. S., Banyard, A. C., Whittard, E., Karibayev, T., Al Kafagi, T., Chvala, I., et al. (2021). Emergence and spread of novel H5N8, H5N5 and H5N1 clade 2.3.4.4 highly pathogenic avian influenza in 2020, Emerg. *Microbes Infect.* 10, 148–151. doi: 10.1080/22221751.2021.1872355

Leyson, C. M., Youk, S., Ferreira, H. L., Suarez, D. L., and Pantin-Jackwood, M. (2021). Multiple gene segments are associated with enhanced virulence of clade 2.3.4.4 H5N8 highly pathogenic avian influenza virus in mallards. *J. Virol.* 95:e0095521. doi: 10.1128/jvi.00955-21

Li, J., Gu, M., Liu, D., Liu, B., Jiang, K., Zhong, L., et al. (2014). Phylogenetic and biological characterization of three K1203 (H5N8)-like avian influenza a virus reassortants in China in 2014. *Arch. Virol.* 161, 289–302. doi: 10.1007/s00705-015-2661-2

Li, M., Liu, H., Bi, Y., Sun, J., Wong, G., Liu, D., et al. (2017). Highly pathogenic avian influenza a(H5N8) virus in wild migratory birds, Qinghai lake, China. *Emerg. Infect. Dis.* 23, 637–641. doi: 10.3201/eid2304.161866

Li, X., Lv, X., Li, Y., Xie, L., Peng, P., An, Q., et al. (2022). Emergence, prevalence, and evolution of H5N8 avian influenza viruses in central China, 2020. *Emer. Microb. Infections* 11, 73–82. doi: 10.1080/22221751.2021.2011622

Long, J.-X., Peng, D.-X., Liu, Y.-L., Wu, Y.-T., and Liu, X.-F. (2008). Virulence of H5N1 avian influenza virus enhanced by a 15-nucleotide deletion in the viral nonstructural gene. *Virus Genes* 36, 471–478. doi: 10.1007/s11262-007-0187-8

Lycett, S. J., Ward, M. J., Lewis, F. I., Poon, A. F. Y., Pond, S. L. K., and Brown, A. J. L. (2009). Detection of mammalian virulence determinants in highly pathogenic avian influenza H5N1 viruses: multivariate analysis of published data. *J. Virol.* 83, 9901–9910. doi: 10.1128/jvi.00608-09

Marchenko, V. Y., Susloparov, I. M., Komissarov, A. B., Fadeev, A., Goncharova, N. I., Shipovalov, A. V., et al. (2017). Reintroduction of highly pathogenic avian influenza a/H5N8 virus of clade 2.3.4.4. In Russia. *Arch. Virol.* 162, 1381–1385. doi: 10.1007/s00705-017-3246-z

Matrosovich, M., Zhou, N., Kawaoka, Y., and Webster, R. (1999). The surface glycoproteins of H5 influenza viruses isolated from humans, chickens, and wild aquatic birds have distinguishable properties. *J. Virol.* 73, 1146–1155. doi: 10.1128/jvi.73.2.1146-1155.1999

Motahhar, M., Keyvanfar, H., Shoushtari, A., Mehrabadi, M. H. F., and Brujeni, G. N. (2016). The arrival of highly pathogenic avian influenza viruses H5N8 in Iran through two windows. *Virus Genes* 58, 527–539. doi: 10.1007/s11262-022-01930-8

Nagy, A., Černíková, L., Kunteová, K., Dirbáková, Z., Thomas, S. S., Slomka, M. J., et al. (2021). A universal RT-qPCR assay for "one health" detection of influenza a viruses. *PLoS One* 16:e0244669. doi: 10.1371/journal.pone.0244669

Niqueux, E., Flodrops, M., Allée, C., Lebras, M.-O., Pierre, I., Louboutin, K., et al. (2023). Evaluation of three hemagglutinin-based vaccines for the experimental control of a panzootic clade 2.3.4.4b A(H5N8) high pathogenicity avian influenza virus in mule ducks. *Vaccine* 41, 145–158. doi: 10.1016/j.vaccine.2022.11.012

Orozovic, G., Orozovic, K., Lennerstrand, J., and Olsen, B. (2011). Detection of resistance mutations to antivirals oseltamivir and zanamivir in avian influenza a viruses isolated from wild birds. *PLoS One* 6:e16028. doi: 10.1371/journal.pone.0016028

Ozawa, M., Matsuu, A., Tokorozaki, K., Horie, M., Masatani, T., Nakagawa, H., et al. (2015). Genetic diversity of highly pathogenic H5N8 avian influenza viruses at a single overwintering site of migratory birds in Japan, 2014/15. *Eur. Secur.* 20:21132. doi: 10.2807/1560-7917.es2015.20.20.21132

Perdue, M. L., García, M., Senne, D., and Fraire, M. (1997). Virulence-associated sequence duplication at the hemagglutinin cleavage site of avian influenza viruses. *Virus Res.* 49, 173–186. doi: 10.1016/s0168-1702(97)01468-8

Pulit-Penaloza, J. A., Brock, N., Pappas, C., Sun, X., Belser, J. A., Zeng, H., et al. (2020). Characterization of highly pathogenic avian influenza H5Nx viruses in the ferret model. *Sci. Rep.* 10:12700. doi: 10.1038/s41598-020-69535-5

Pyankova, O. G., Susloparov, I. M., Moiseeva, A. A., Kolosova, N. P., Onkhonova, G. S., Danilenko, A. V., et al. (2021). Isolation of clade 2.3.4.4b a(H5N8), a highly pathogenic avian influenza virus, from a worker during an outbreak on a poultry farm, Russia, December 2020. *Euro. Surveill.* 26:2100439. doi: 10.2807/1560-7917.es.2021.26.24.2100439

Sapachova, M., Kovalenko, G., Sushko, M., Bezymennyi, M., Muzyka, D., Usachenko, N., et al. (2021). Phylogenetic analysis of H5N8 highly pathogenic avian influenza viruses in Ukraine, 2016–2017. *Vector-Borne Zoonotic Dis.* 21, 979–988. doi: 10.1089/vbz.2021.0031

Scoizec, A., Niqueux, E., Thomas, R., Daniel, P., Schmitz, A., and Le Bouquin, S. (2018). Airborne detection of H5N8 highly pathogenic avian influenza virus genome in poultry farms, France. *Front. Vet. Sci.* 5:15. doi: 10.3389/fvets.2018.00015

Shaw, M., Cooper, L., Xu, X., Thompson, W., Krauss, S., Guan, Y., et al. (2001). Molecular changes associated with the transmission of avian influenza a H5N1 and H9N2 viruses to humans. *J. Med. Virol.* 66, 107–114. doi: 10.1002/jmv.2118

Shi, J., Deng, G., Kong, H., Gu, C., Ma, S., Yin, X., et al. (2017). H7N9 virulent mutants detected in chickens in China pose an increased threat to humans. *Cell Res.* 27, 1409–1421. doi: 10.1038/cr.2017.129

Shi, J., Deng, G., Ma, S., Zeng, X., Yin, X., Li, M., et al. (2018). Rapid Evolution of H7N9 Highly Pathogenic Viruses that Emerged in China in 2017. *Cell Host Microbe.* 24, 558–568. e7. doi: 10.1016/j.chom.2018.08.006

Shi, W., and Gao, G. F. (2021). Emerging H5N8 avian influenza viruses. *Science* 372, 784–786. doi: 10.1126/science.abg6302

Shin, D.-L., Siebert, U., Lakemeyer, J., Grilo, M., Pawliczka, I., Wu, N.-H., et al. (2019). Highly pathogenic avian influenza a(H5N8) virus in gray seals, Baltic Sea. *Emerg. Infect. Dis.* 25, 2295–2298. doi: 10.3201/eid2512.181472

Siddique, N., Naeem, K., Abbas, M. A., Ahmed, Z., and Malik, S. A. (2012). Sequence and phylogenetic analysis of highly pathogenic avian influenza H5N1 viruses isolated during 2006–2008 outbreaks in Pakistan reveals genetic diversity. *Virol. J.* 9:300. doi: 10.1186/1743-422x-9-300

Śmietanka, K., Świętoń, E., Kozak, E., Wyrostek, K., Tarasiuk, K., Tomczyk, G., et al. (2020). Highly pathogenic avian influenza H5N8 in Poland in 2019–2020. *J. Vet. Res.* 64, 469–476. doi: 10.2478/jvetres-2020-0078

Smith, G. J. D., Naipospos, T. S. P., Nguyen, T. D., de Jong, M. D., Vijaykrishna, D., Usman, T. B., et al. (2006). Evolution and adaptation of H5N1 influenza virus in avian and human hosts in Indonesia and Vietnam. *Virology* 350, 258–268. doi: 10.1016/j.virol.2006.03.048

Spesock, A., Malur, M., Hossain, M. J., Chen, L.-M., Njaa, B. L., Davis, C. T., et al. (2011). The virulence of 1997 H5N1 influenza viruses in the mouse model is increased by correcting a defect in their NS1 proteins. *J. Virol.* 85, 7048–7058. doi: 10.1128/jvi.00417-11

Twabela, A. T., Tshilenge, G. M., Sakoda, Y., Okamatsu, M., Bushu, E., Kone, P., et al. (2017). Highly pathogenic avian influenza a(H5N8) virus, democratic republic of the Congo. *Emerg. Infect. Dis.* 24, 1371–1374. doi: 10.3201/eid2407.172123

Verhagen, J. H., Fouchier, R. A. M., and Lewis, N. (2021). Highly pathogenic avian influenza viruses at the wild–domestic bird interface in Europe: future directions for research and surveillance. *Viruses* 13:212. doi: 10.3390/v13020212

Wang, Y., Wang, M., Zhang, H., Zhao, C., Zhang, Y., He, G., et al. (2022). Emergence, evolution, and biological characteristics of H10N4 and H10N8 avian influenza viruses in migratory wild birds detected in eastern China in 2020. *Microbiol. Spectr.* 10:e0080722. doi: 10.1128/spectrum.00807-22

Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M., and Kawaoka, Y. (1992). Evolution and ecology of influenza a viruses. *Microbiol. Rev.* 56, 152–179. doi: 10.1128/mr.56.1.152-179.1992

Ye, H., Zhang, J., Sang, Y., Shan, N., Qiu, W., Zhong, W., et al. (2022). Divergent reassortment and transmission dynamics of highly pathogenic avian influenza a(H5N8) virus in birds of China during 2021. *Front. Microbiol.* 13:913551. doi: 10.3389/fmicb.2022.913551

Yehia, N., Hassan, W. M. M., Sedeek, A., and Elhusseiny, M. H. (2020). Genetic variability of avian influenza virus subtype H5N8 in Egypt in 2017 and 2018. *Arch. Virol.* 165, 1357–1366. doi: 10.1007/s00705-020-04621-7

Yehia, N., Naguib, M. M., Li, R., Hagag, N., El-Husseiny, M., Mosaad, Z., et al. (2018). Multiple introductions of reassorted highly pathogenic avian influenza viruses (H5N8) clade 2.3.4.4b causing outbreaks in wild birds and poultry in Egypt. *Infect. Genet. Evol.* 58, 56–65. doi: 10.1016/j.meegid.2017.12.011

Yu, Y., Zhang, Z., Li, H., Wang, X., Li, B., Ren, X., et al. (2017). Biological characterizations of H5Nx avian influenza viruses embodying different neuraminidases. *Front. Microbiol.* 8:1084. doi: 10.3389/fmicb.2017.01084

Zeng, X., Chen, X., Ma, S., Wu, J., Bao, H., Pan, S., et al. (2020). Protective efficacy of an H5/H7 trivalent inactivated vaccine produced from Re-11, Re-12, and H7-Re2 strains against challenge with different H5 and H7 viruses in chickens. *J. Integr. Agric.* 19, 2294–2300. doi: 10.1016/S2095-3119(20)63301-9

Zeng, X., Tian, G., Shi, J., Deng, G., Li, C., and Chen, H. (2018). Vaccination of poultry successfully eliminated human infection with H7N9 virus in China. *Sci. China. Life Sci.* 61, 1465–1473. doi: 10.1007/s11427-018-9420-1

Zhao, K., Gu, M., Zhong, L., Duan, Z., Zhang, Y., Zhu, Y., et al. (2013). Characterization of three H5N5 and one H5N8 highly pathogenic avian influenza viruses in China. *Vet. Microbiol.* 163, 351–357. doi: 10.1016/j.vetmic.2012.12.025

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 06 June 2023
DOI 10.3389/fmicb.2023.1175188

OPEN ACCESS

EDITED BY
Marco Goeijenbier,
Spaarne Gasthuis, Netherlands

REVIEWED BY
Ahmet Kursat Azkur,
Kırıkkale University, Türkiye
Hongliang Wang,
Xi'an Jiaotong University, China

*CORRESPONDENCE
Jintao Li
✉ ljtqms@tmmu.edu.cn
Yuwei Gao
✉ yuwei0901@outlook.com
Hui Wang
✉ geno0109@vip.sina.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 27 February 2023
ACCEPTED 15 May 2023
PUBLISHED 06 June 2023

CITATION
Zhou X, Sun W, Zhang Y, Gu H, Wang R, Xie P, Zhu Y, Qiu M, Ding X, Wang H, Gao Y and Li J (2023) A novel hACE2 knock-in mouse model recapitulates pulmonary and intestinal SARS-CoV-2 infection.
Front. Microbiol. 14:1175188.
doi: 10.3389/fmicb.2023.1175188

A novel hACE2 knock-in mouse model recapitulates pulmonary and intestinal SARS-CoV-2 infection

Xiaoyang Zhou[1†], Weiyang Sun[2†], Yu Zhang[1†], Hongjing Gu[3†], Ruixuan Wang[1], Peng Xie[1], Yunkai Zhu[1], Minyue Qiu[1], Xiaoyan Ding[1], Hui Wang[3*], Yuwei Gao[2*] and Jintao Li[1*]

[1]Department of Biosafety, School of Basic Medicine, Army Medical University, Chongqing, China, [2]Changchun Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Changchun, China, [3]State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, AMMS, Beijing, China

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission is responsible for the coronavirus disease 2019 (COVID-19) pandemic. SARS-CoV-2 uses the angiotensin-converting enzyme 2 (ACE2) receptor to enter the host, and the gastrointestinal tract is a potential infection site as this receptor is expressed on it. Multiple studies have indicated that an increasing number of COVID-19 patients presented with gastrointestinal symptoms that are highly associated with disease severity. Moreover, emerging evidence has demonstrated that alterations in the gut immune microenvironment induced by intestinal SARS-CoV-2 infection can regulate respiratory symptoms. Therefore, targeting the intestines may be a candidate therapeutic strategy in patients with COVID-19; however, no mouse model can serve as an appropriate infection model for the development of fatal pneumonia while mimicking intestinal infection. In this study, a novel human ACE2 knock-in (KI) mouse model (or hACE2-KI) was systemically compared with the popular K18-hACE2 mice; it showed differences in the distribution of lung and intestinal infections and pathophysiological characteristics. These newly generated hACE2-KI mice were susceptible to intranasal infection with SARS-CoV-2, and not only developed mild to severe lung injury, but also acquired intestinal infection. Consequently, this model can be a useful tool for studying intestinal SARS-CoV-2 infection and developing effective therapeutic strategies.

KEYWORDS

mouse model, intestinal infection, SARS-CoV-2, immune cells, pneumonia

1. Introduction

The coronavirus disease 2019 (COVID-19) has been a global health crisis for the past few years. It is well established that the causative pathogen, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) belonging to the order *Nidovirales*, family *Coronaviridae*, and genus *Coronavirus*, uses human angiotensin-converting enzyme 2 (hACE2) as the main receptor for initiating the infection (Hoffmann et al., 2020). Most studies have focused on clinical symptoms in the respiratory tract, as viral infection can cause acute respiratory distress syndrome and respiratory failure, which are considered the main causes of death in COVID-19 patients (Huang et al., 2020; Empson et al., 2022; Ramadori, 2022; Wang and Perlman, 2022).

However, ACE2 is also highly expressed in the human intestinal tract, and a increasing number of patients presented with gastrointestinal symptoms (Liang et al., 2020; Lin et al., 2020; Nobel et al., 2020; Muus et al., 2021; Blackett et al., 2022; Freedberg and Chang, 2022; Jin et al., 2022). Furthermore, direct evidence of active SARS-CoV-2 replication in the intestines has been found in clinical samples (Xiao et al., 2020; Qian et al., 2021; Zuo et al., 2021; Cuicchi et al., 2022; Freedberg and Chang, 2022).

Multiple studies have indicated that gastrointestinal symptoms may be closely associated with disease severity in COVID-19 patients. From a clinical perspective, patients with gastrointestinal symptoms have lower mortality than those without gastrointestinal symptoms (Hajifathalian et al., 2020; Nobel et al., 2020; Livanos et al., 2021). According to previous studies, bidirectional cross-talk between the intestinal immune microenvironment and respiratory tract infections occurs during influenza infection (Trompette et al., 2018; Sencio et al., 2020). Therefore, it was hypothesized that intestinal immune status after SARS-CoV-2 infection may contribute to the regulation of COVID-19 severity. Thus, a successful small animal model capable of recapitulating the respiratory and intestinal pathology observed in COVID-19 patients is expected to be useful to investigate the role of intestinal infection in modulating COVID-19.

As SARS-CoV-2 is unable to engage mouse ACE2 as a receptor for infection, mice expressing hACE2 have been developed and widely used for COVID-19 research (Bao et al., 2020; Jiang et al., 2020; Wan et al., 2020; Dong et al., 2022; Sefik et al., 2022). However, the currently available mouse models used to recapitulate the major pathological characteristics of COVID-19 patients do not mimic the intestinal infection well. Mice transduced with replication-defective adenoviruses only transiently express hACE2 in the lungs, and virtually no viral RNA is detected in the intestines (Hassan et al., 2020; Sefik et al., 2022). Some transgenic mouse models express hACE2 in the intestines driven by strong promoters (epithelial cell cytokeratin-18 [K18], lung-ciliated epithelial cell-specific HFH4/FOXJ1, and CAG); however, hACE2 expression is not physiologically distributed in these models (Jiang et al., 2020; Asaka et al., 2021; Dong et al., 2022). Notably, these transgenic mice show high brain tropism of SARS-CoV-2, and the symptoms that lead to mortality are mostly due to central nervous system dysfunction, whereas COVID-19 is primarily a respiratory disease in humans. Recently, hACE2 knock-in (KI) mouse models have received considerable attention because they have a clear genetic background, good reproductive performance, and most importantly, appropriate hACE2 tissue distribution under the control of an endogenous promotor. However, no obvious clinical symptoms or mortality are observed, and pathological changes in the lung tissues are minimal in these models. Consequently, suboptimal virus replication in these mice may lead to a failure in establishing intestinal infection (Sun et al., 2020; Sefik et al., 2022; Winkler et al., 2022).

In this study, a novel humanized hACE2 KI mouse model was established using clustered regularly interspaced palindromic repeats (CRISPR)/Cas9. This model is susceptible to intranasal infection with SARS-CoV-2, and mild to severe lung injury as well as clear evidence of intestinal infection was observed. Hence, the hACE2-KI mouse model described in this study can be of great value to find new clues to uncover COVID-19 pathogenesis and test new therapeutics for combating SARS-CoV-2.

2. Materials and methods

2.1. Ethics statement

All procedures involving infectious viruses were conducted in a Biosafety Level 3 laboratory. All experimental procedures involving mice in this study were approved by the Laboratory Animal Welfare and Ethics Committee of the Third Military Medical University (AMUWE20201373).

2.2. Establishment of hACE2-KI mice

To generate hACE2 KI mouse models, eggs were collected from mated female BALB/c mice and subjected to pronuclear microinjection with a complex of Cas9 mRNA, a pair of single guide RNAs (sgRNAs) sgRNA1 and sgRNA2 (Table 1), and an hACE2 KI vector plasmid at concentrations of 20, 10, and 10 ng/μL, respectively. The microinjected zygotes were subsequently implanted into pseudopregnant recipient mice. F0 founder mice were identified using two pairs of primers hACE2-F2/R2 and hACE2-F1/R1 (Table 1), which covered the 5′ and 3′ junction regions of hACE2 KI allele, respectively. Animals with positive PCR products were further confirmed by sequencing with primers R5 and R3 (Table 1) for 5′ and 3′ junction regions, respectively. The expected size of amplicons was 3,800 bp and 3,900 bp, respectively. Genotype-positive founders were backcrossed with wild-type (WT) BALB/c mice to produce generation F1, which were screened and positive individuals were self-bred. DNA samples of homozygous mice were subjected to Southern blotting to confirm correct insertion using 5′ and 3′ probes. Two endonucleases *Bam*HI (New England Biolabs, Ipswich, MA, United States; R0136) and *Mfe*I (New England Biolabs; R0589) were used to digest the DNA to identify correct insertion and no random recombinant. The expected fragment sizes for Southern blotting were 5′ probe-*Bam*HI: 4.54 kb-WT, 3.00 kb-mutation (MT) and 3′ probe-*Mfe*I: 10.83 kb-WT, 4.25 kb-MT, respectively. Primer pairs of 5′ probe (5′ probe-F/R) and 3′ probe (3′ probe-F/R) are shown in Table 1.

2.3. Experimental animals and study design

As K18-hACE2 transgenic mice were widely used in studies of SARS-CoV-2 infection, we included this mouse model for a comparative study. All mice aged 8–10 weeks were divided into three groups: homozygous hACE2-KI transgenic mice ($n = 19$),

---

Abbreviations: COVID-19, Coronavirus disease 2019; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; hACE2, human angiotensin-converting enzyme 2; KI, knock-in; K18, epithelial cell cytokeratin-18; CRISPR, clustered regularly interspaced palindromic repeats; sgRNAs, single guide RNAs; WT, wild-type; MT, mutation; PFU, plaque-forming units; dpi, days post-infection; qPCR, quantitative PCR; E, envelope gene; S, spike protein; H&E, hematoxylin and eosin; CDS, coding sequence.

TABLE 1 Primers and probes used for PCR or real-time qPCR analysis.

Primer/Probe/sgRNA name	Primer/Probe/sgRNA sequence
hACE2-F2	5′-GCTTCCACTCCTTATTAGCCTAGTG-3′
hACE2-R2	5-TATCCTCACTTTGATGCTTTGGTC-3′
hACE2-F1	5′-TGAATAATGCTGGGGACAAATGG-3′
hACE2-R1	5′-GAGGATAGAATTGGTTCTTAGGAAGG-3′
5′ Probe-F	5′-TCTTACACTCTGGGAATGAGGACACG-3′
5′ Probe-R	5′-GATGCTTCCTGTGTGGCTTTGGTAA-3′
3′ Probe-F	5′-AATGTGCCTTTGGCCTCACAGTCTA-3′
3′ Probe-R	5′-GATGTCTGGCTTCTTTCTCCGTTGA-3′
sgRNA1	5′-CTTGGCATTTTCCTCGGTGAGGG-3′
sgRNA2	5′-TCTGAGCATCATCACTGTTTTGG-3′
Sequencing primer (R5)	5′-TAGTGGATACATTTGGGCAAGTG-3′
Sequencing primer (R3)	5′-AAGAGATGTCAAATCCTTAGGCAG-3′
qPCR-hACE2-F	5′-ACAGTCCACACTTGCCCAAAT-3′
qPCR-hACE2-R	5′-TGAGAGCACTGAAGACCCATT-3′
qPCR-mACE2-F	5′-TCCAGACTCCGATCATCAAGC-3′
qPCR-mACE2-R	5′-TGCTCATGGTGTTCAGAATTGT-3′
qPCR-mGAPDH-F1	5′-AGGTCGGTGTGAACGGATTTG-3′
qPCR-mGAPDH-R1	5′-GGGGTCGTTGATGGCAACA-3′
E-leader	5′-CGATCTCTTGTAGATCTGTTCTC-3′
E-reverse	5′-ATATTGCAGCAGTACGCACACA-3′
E-probe	5′-FAM-ACACTAGCCATCCTTACTGCGCTTCG-BHQ1-3′

K18-hACE2 transgenic mice ($n = 9$), and WT BALB/c mice ($n = 19$). hACE2-KI transgenic mice were prepared and bred in our laboratory and K18-hACE2 transgenic mice and WT BALB/c mice were obtained from GemPharmatech Co. Ltd. (Nanjing, China) and the Laboratory Animal Center of Army Medical University, respectively.

Mice in these groups were anesthetized and intranasally inoculated with SARS-CoV-2 delta (B.1.617.2, national number: CCPM-B-V-049-2,105-08) at a dose of 1×10^5 plaque-forming units (PFU)/mouse. All mice were housed in groups, fed standard chow diets, monitored and weighed daily. Three mice in each group were euthanized at 3, 5, and 7 days post-infection (dpi) and tissues were collected for further analysis.

2.4. Immunochemistry assay

Collected tissues were fixed with 4% paraformaldehyde, embedded in paraffin, and sectioned at a thickness of 5 μm. Sections were dewaxed using xylene and rehydrated using a graded series of ethanol solutions. To detect hACE2 distribution in the organs of the hACE2-KI transgenic mouse model, an anti-hACE2 monoclonal antibody (Abcam, Cambridge, MA, United States; ab108209) was used as the primary antibody, and the sections were incubated with secondary antibody (Absin Bioscience, Shanghai, China; abs957), followed by visualization with a DAB Detection Kit (ZSGB Biotech, Beijing, China; ZLI-9018).

2.5. RNA extraction and real-time quantitative PCR

Total RNA was extracted from tissue homogenates of organs using RNAiso Plus (Takara Bio, Shiga, Japan; 9,108), and reverse transcription was performed to produce cDNA using the PrimeScript RT Reagent Kit (Takara Bio; RR047A) according to the manufacturer's recommendations. Real-time quantitative PCR (qPCR) was performed using TB Green Premix Ex Taq II (Takara Bio; RR820A) with the LightCycler 96 System (Roche Diagnostics, Indianapolis, IN, United States). Each reaction was performed in triplicate. Relative expression levels of hACE2 were determined using PCR pairs of primers qPCR-hACE2-F, qPCR-hACE2-R and qPCR-mACE2-F, qPCR-mACE2-R (Table 1). Amplicons were normalized to *GAPDH* expression, and primer pairs qPCR-mGAPDH-F1/R1 used for *GAPDH* are shown in Table 1.

2.6. Protein extraction and expression analysis

Frozen tissues (30 mg) were homogenized in RIPA Lysis Buffer (Beyotime Biotechnology, Shanghai, China; P0013B) supplemented with protease and phosphatase inhibitors (Beyotime Biotechnology; P1050). The homogenates were kept on ice for 30 min and then spun at 13,000 × *g* for 15 min at 4°C. The supernatant was collected and protein concentrations were determined using a BCA kit (Beyotime Biotechnology; P0010S). Protein expression and total protein were measured with the automated capillary-based Jess system (ProteinSimple, San Jose, CA, United States). Briefly, each sample was mixed with a master mix (1× fluorescent standard, 1× sample buffer, and 40 mM dithiothreitol) and heated at 95°C for 5 min to denature the samples. Subsequently, 3 μL of the denatured proteins and 10 μL each of protein normalization solution, chemiluminescent substrate, primary antibodies, and horseradish peroxidase-conjugated anti-rabbit secondary antibodies were pipetted into the appropriate wells of the assay plate. A biotinylated ladder cartridge (12–230 kDa) was integrated into each assay. Subsequently, the plate and the capillaries were moved to the Jess machine for automated protein electrophoresis, blocking, antibody incubation, and signal detection. Chemiluminescent reactions were analyzed by Compass for SW software ver. 6.1.0 (ProteinSimple). Relative protein amounts were assessed using the corrected area of chemiluminescent peaks. The primary antibodies were human ACE2 antibody (Sino Biological, Beijing, China; 10,108-RP01) and mouse ACE2 antibody (R&D Systems, Minneapolis, MN, United States; MAB34372), and both were used at 1:50 dilution.

2.7. Measurements of viral loads

Viral RNA quantification was performed using real-time qPCR targeting the SARS-CoV-2 subgenomic RNA transcript for the envelope (E) gene, as previously reported (Wölfel et al., 2020). Briefly, RNA was isolated from the tissue homogenates using Viral RNA Extraction Kit (Takara Bio, 9,766). Real-time qPCR was performed using the One Step PrimeScript III RT-PCR Kit (Takara Bio, RR600A) with the primer pairs E-Leader, E-reverse, and E-probe (Table 1). The cycling conditions were one cycle at 52°C for 5 min, then 95°C for 10 s, followed by 45 cycles at 95°C for 10 s and 60°C for 30 s. pUC19-2019-nCoV-E plasmid was synthesized by Sangon Biotech Co., Ltd. (Shanghai, China) as an E gene DNA standard and the inserted base sequence reported in a previous study was used (Wölfel et al., 2020). An E gene DNA sample was also run at the same time for conversion of cycle threshold value to genomic copies using the standard curve-based method.

2.8. Histopathological analysis

The processed and sectioned tissues were stained with hematoxylin and eosin (H&E) according to the standard procedures of the H&E Staining Kit (Solarbio Science & Technology, Beijing, China; G1120). The histopathological changes in the tissues were examined by M8 Microscope and Scanner (PreciPoint, Freising, Germany) and analyzed by two experienced pathologists who were blinded to the samples.

2.9. Immunofluorescence assay

The processed sections were incubated with 3% H_2O_2 for 20 min to quench endogenous peroxidase. Antigen repair was performed in boiled citrate buffer (pH 6) for 10 min and non-specific staining was blocked using normal goat serum for 2 h at 37°C. The sections were then incubated with primary antibodies at 4°C overnight and subsequently incubated with secondary antibodies for 2 h. Finally, the sections were stained with 4′,6-diamidino-2-phenylindole (DAPI) (Sigma-Aldrich, St. Louis, MO, USA; 32,670) for 7 min. The primary antibodies used in this study included SARS-CoV-2 spike (S; Sino Biological Japan, Kanagawa, Japan; 40,592-R004), Ly6G (Cell Signaling Technology, Danvers, MA, USA; 87,048), CD68 (Abcam, ab125212), CD3 (Cell Signaling Technology, 78,588), and CD19 (Cell Signaling Technology, 90,176). Secondary antibodies included Alexa Fluor 594 goat anti-rabbit IgG (Thermo Fisher Scientific, Waltham, MA, United States; A11037) and Alexa Fluor 488 goat anti-rabbit IgG (Thermo Fisher Scientific, A11008). Representative sections were scanned using a digital slide scanner (PANNORAMIC MIDI, 3DHISTECH, Budapest, Hungary) with CaseViewer software (3DHISTECH).

2.10. Statistical analysis

All data analyzes were performed using the R package (version 4.2.1; R Foundation for Statistical Computing, Vienna, Austria). Data are presented as mean ± standard error of the mean. The log-rank test was used for survival analysis, and the statistical significance of immunofluorescence staining analysis among different mice groups was assessed by unpaired Student's *t*-tests or Wilcoxon rank tests. Before statistical significance was assessed, the Shapiro–Wilk normality test was used to determine whether the data had a normal distribution. $p < 0.05$ was specified as the threshold of significance.

3. Results

3.1. Establishment of a humanized hACE2 KI mouse model using CRISPR/Cas9

Here, we aimed to establish a human ACE2 mouse model using CRISPR/Cas9 KI technology. As shown in Figure 1A, the mouse endogenous ACE2 coding sequence (CDS) was replaced with the human ACE2 coding sequence (CDS) without the sequence coding signal peptide via CRISPR/Cas9-mediated DNA homologous recombination by double cutting with a pair of sgRNAs, namely sgRNA1 and sgRNA2. The pair of sgRNAs precisely cut 3′ downstream of the start codon (ATG) and 5′ upstream of the termination stop codon (TAG) of mouse endogenous *ACE2*, and the hACE2 CDS was used to precisely replace the mouse endogenous *ACE2* CDS *in situ* and was placed under the drive of the mouse endogenous *ACE2* promoter. Human ACE2 CDS (without the signal peptide) was placed precisely between the start codon and stop codon of the endogenous mouse ACE2 gene, so that the inserted human ACE2 gene could be physiologically distributed in these humanized mice. As described in the Materials and Methods section, transgenic founder mice were subjected to genetic screening by PCR using primer pairs of hACE2-F2/R2 (Figure 1B) and hACE2-F1/R1 (Figure 1C), which covered the 5′ and 3′ junction regions of hACE2 KI allele, respectively. Animals with positive PCR products were further confirmed by sequencing and the sequence chromatography images of the 5′ and 3′ junction regions are shown in Figures 1B,C, respectively. The mice carrying the correct gene were backcrossed with WT mice, the F1 offspring were screened, and positive individuals were self-bred. Homozygous hACE2 KI individuals were subjected to Southern blotting to confirm correct insertion and off-targeting. The genomic DNA of the homozygous individuals was completely digested by the endonucleases *Bam*HI and *Mfe*I, and then subjected to Southern blotting using the 5′ and 3′ probes shown in Figure 1A, respectively. The sizes of Southern blot bands with 5′ probe were 4.54 kb in the WT allele and 3.00 kb in the hACE2 KI allele, and those with 3′ probe were 10.83 kb in the WT allele and 4.25 kb in the hACE2 KI allele (Figure 1D). As expected, none of the three homozygous mice showed random insertion and off-targeting (Figure 1D).

To investigate hACE2 expression patterns, real-time qPCR was used to measure the mRNA levels in different organs. As shown in Figure 2A, hACE2 was mainly expressed in the intestines, lungs, and trachea. The expression of human ACE2 protein in mouse intestine and lung tissues was detected using the capillary-based Jess system (ProteinSimple) and Total Protein Normalization analysis. Immunoblotting of the proteins showed

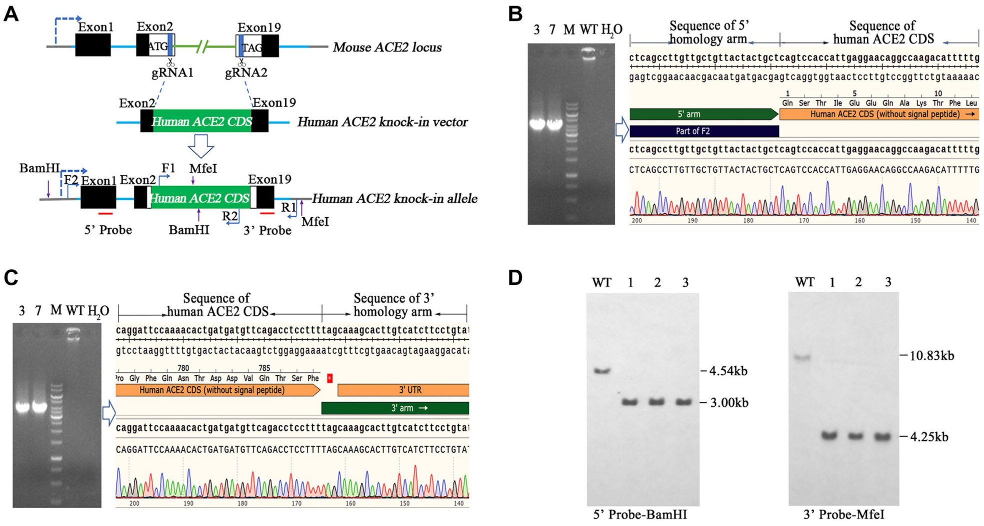


FIGURE 1
Generation and genetic characterization of human angiotensin-converting enzyme 2 (hACE2) knock-in (KI) mice. **(A)** The strategy of generation of hACE2 KI mice. A pair of single guide RNAs (sgRNAs) cutting exons 2 and 19 of mouse ACE2, respectively, was used to replace the translated exons of mouse ACE2 with the human ACE2 coding sequence (CDS) without the sequence coding signal peptide, and the human ACE2 CDS with signal peptide was placed downstream the sequence coding mouse ACE2 signal peptide as a result. Black boxes: untranslated exonic sequences of mouse ACE2; white boxes: translated exonic sequences of mouse ACE2; green box: human ACE2 CDS without signal peptide; blue lines: introns of mouse ACE2; the arrow in dotted line: the transcription direction; F1, F2, R1, and R2 show the positions of primers used for genetic screening of hACE2 KI mice; red lines indicate the locations of probes for Southern blotting. **(B,C)** Transgenic founder mice were subjected to genetic screening by PCR using primer pairs F2/R2 (B) and F1/R1 **(C)**, which covered the 5′ and 3′ junction regions of hACE2 KI allele, respectively. The PCR products were sequenced and the sequence chromatography images are shown. **(D)** Southern blotting of homozygous hACE2 KI individuals. The genomic DNAs of the homozygous individuals were completely digested by the endonucleases *Bam*HI and *Mfe*I, and then subjected to Southern blotting using the 5′ and 3′ probes shown in A, respectively. The sizes of Southern blot bands with the 5′ probe were 4.54kb in the wild-type (WT) allele and 3.00kb in the hACE2 KI allele, and those with 3′ probe were 10.83kb in the WT allele and 4.25kb in the hACE2 KI allele. WT: wild-type mouse genomic DNA; ~1–3: homozygous KI individuals.

that hACE2 was highly expressed in the lungs and intestines of hACE2-KI mice, while mouse ACE2 protein was detected in WT mice but not in hACE2-KI mice. Furthermore, the expression level of human ACE2 in the intestinal tracts of hACE2-KI mice was about 15-fold higher than that in the lungs (Figure 2B). Immunohistochemical analysis revealed robust hACE2 expression in the intestinal brush border, lung bronchi, and tracheal epithelial cells (Figure 2C). Thus, a mouse model that highly expresses hACE2 in the intestines, lungs, and trachea was established successfully.

3.2. Intranasal inoculation of SARS-CoV-2 induces pulmonary and intestinal infection

To further estimate the susceptibility of hACE2-KI mice to SARS-CoV-2, groups of hACE2-KI mice ($n = 19$) and K18-hACE2 mice ($n = 9$) were intranasally challenged with 1×10^5 PFU of SARS-CoV-2. WT BALB/c mice ($n = 19$) that received the same viral dose were used as negative controls. All mice were monitored daily for survival or euthanized at various dpi for tissue collection to measure viral copy numbers and observe histopathological changes (Figure 3A). It was found that 30% (3/10) of hACE2-KI mice died at 7–10 dpi and 100% of K18-hACE2 transgenic mice died at 5–7 dpi (6/6), whereas all WT mice survived until the end of follow-up at 10 dpi (Figure 3B). These results were consistent with previous findings that state that the K18-hACE2 mouse is a lethal model. In contrast to K18-hACE2 mice, our established hACE2-KI model mice showed lower mortality. Next, we measured the viral infection pattern. As shown in Figure 3C, viral RNA replication was observed in the lungs and intestines of hACE2-KI mice, and RNA copies were below the detection limit in the brain, heart, kidneys, liver, and spleen. As expected, the S protein was detected with immunofluorescence staining in the lung sections of both hACE2-KI and KI-18 mice. Furthermore, only the hACE2-KI mice showed S protein-positive cells in the intestinal sections (Figure 3D). These results indicate that the intranasal inoculation of SARS-CoV-2 could induce pulmonary and intestinal infection in our mouse model.

3.3. Histopathological changes in lungs and intestines after SARS-CoV-2 infection

Regarding the histopathological changes in lung tissues, H&E staining results showed that hACE2-KI mice acquired different levels of pneumonia in a time-dependent manner, characterized by a large amount of inflammatory cell infiltration, alveolar wall, and alveolar septa thickening (Figure 4A). As shown in the pathological sections, hACE2-KI mice displayed the most severe pneumonia at 3 dpi. In addition to inflammatory cell infiltration in the alveolar septa, alveolar congestion, peripheral parenchymal collapse, and ruptured septa were also observed. At 5 dpi, the damage progressed into more diffuse lesions with extensive infiltration of inflammatory cells into the

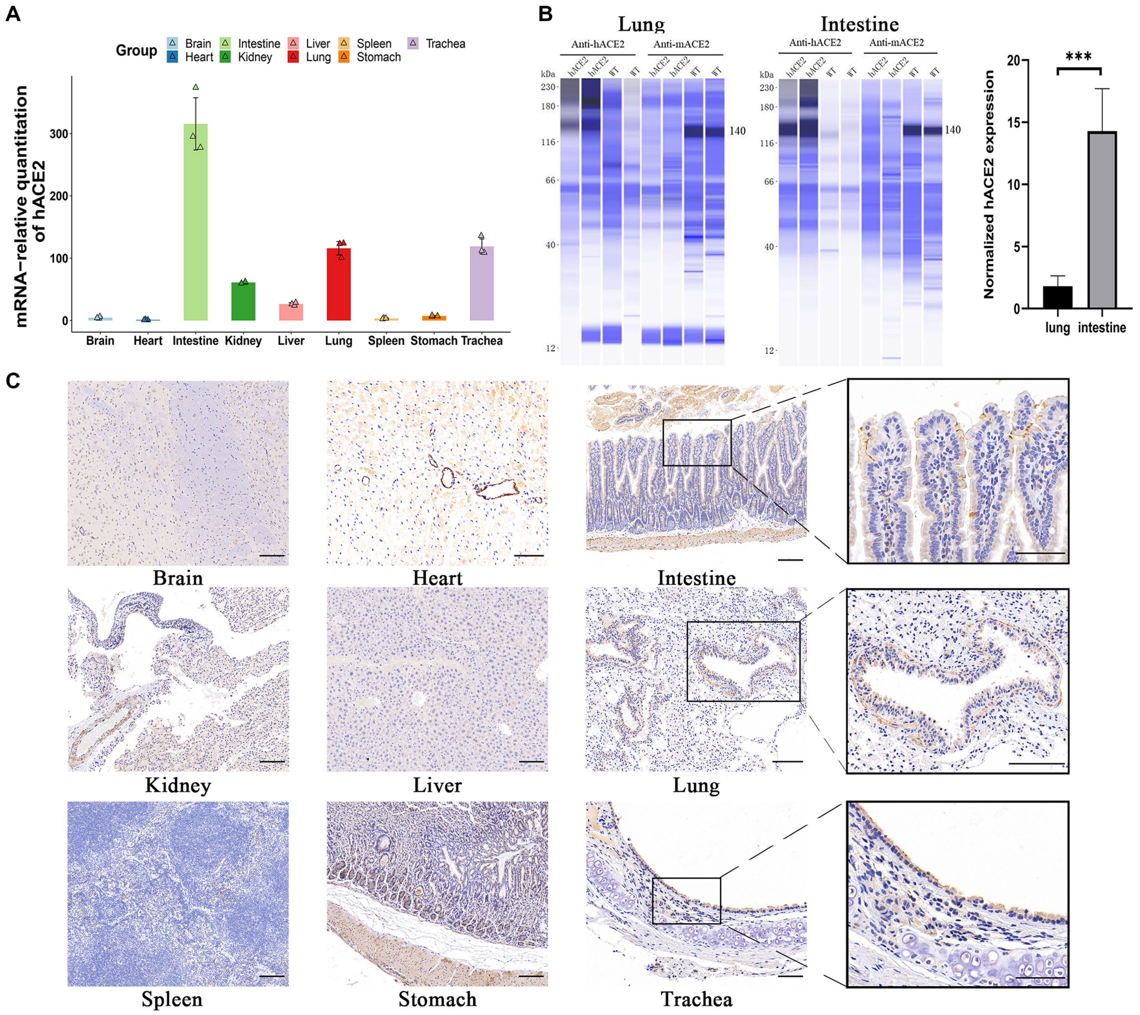


FIGURE 2

Expression of hACE2 in hACE2-KI mouse model. **(A)** The expression pattern of hACE2 mRNA in different organs including brain, heart, intestines, kidneys, liver, lungs, spleen, stomach, and trachea was detected by real-time quantitative PCR (qPCR) assay. **(B)** Representative immunoblotting image shows the expression of the ACE2 protein in mouse intestine and lung tissues normalized to total protein. Human ACE2 was highly expressed in the lungs and intestines of hACE2-KI mice, while mouse ACE2 protein was detected in WT mice but not in hACE2-KI mice. **(C)** Immunochemistry results of the hACE2 expression in different organs. The intestines, lungs, and trachea sections are shown in high-power magnification; scale bar: 100µm. Statistical analysis was performed using unpaired Student's *t*-test or Wilcoxon rank test. NS: not significant; $*p<0.05$; $**p<0.01$; $***p<0.001$.

alveolar cavities. This implies that the immune response remained high at this stage. At 7 dpi, mice that recovered from the infection were characterized by descending inflammatory cells in the alveolar septa and interstitial locations. Notably, the histopathological changes in hACE2-KI mice were comparable with those in K18-hACE2 transgenic mice during the same infection course. According to the histopathological observations in lung tissues, we confirmed that our infected mouse model developed typical pneumonia characteristics comparable to the K18-hACE2 mice. Regarding intestinal histopathological changes, mild inflammatory cell infiltration was identified in the intestinal lamina propria. Additionally, thinning of the muscular layer and intestinal epithelial injury were observed in our established model (Figure 4B). These findings imply that the hACE2-KI mice acquired intestinal injury after SARS-CoV-2 infection, which was consistent with the detected viral load in the intestine.

3.4. Immune responses in the lungs and intestines of hACE2-KI mice

To investigate the infiltration of specific inflammatory cells, immunofluorescence assays were performed to identify CD3$^+$ T lymphocytes, CD4$^+$ T lymphocytes, CD8$^+$ T lymphocytes, CD19$^+$ B lymphocytes, CD68$^+$ macrophages, and Ly6G$^+$ neutrophils. Among inflammatory cells, T lymphocytes were diffusely distributed in the lungs, whereas neutrophils were mainly found in the peri-bronchus (Figure 5A). These results demonstrate that both hACE2-KI and K18-hACE2 mice exhibited elevated immune responses in the lung tissues compared with the WT group (Figure 5B). Interestingly, there were no remarkable differences in inflammatory cell counts between hACE2-KI and K18-hACE2 mice, indicating that SARS-CoV-2 infection in hACE2-KI mice induced pneumonia, similar to that described in K18-hACE2 mice.

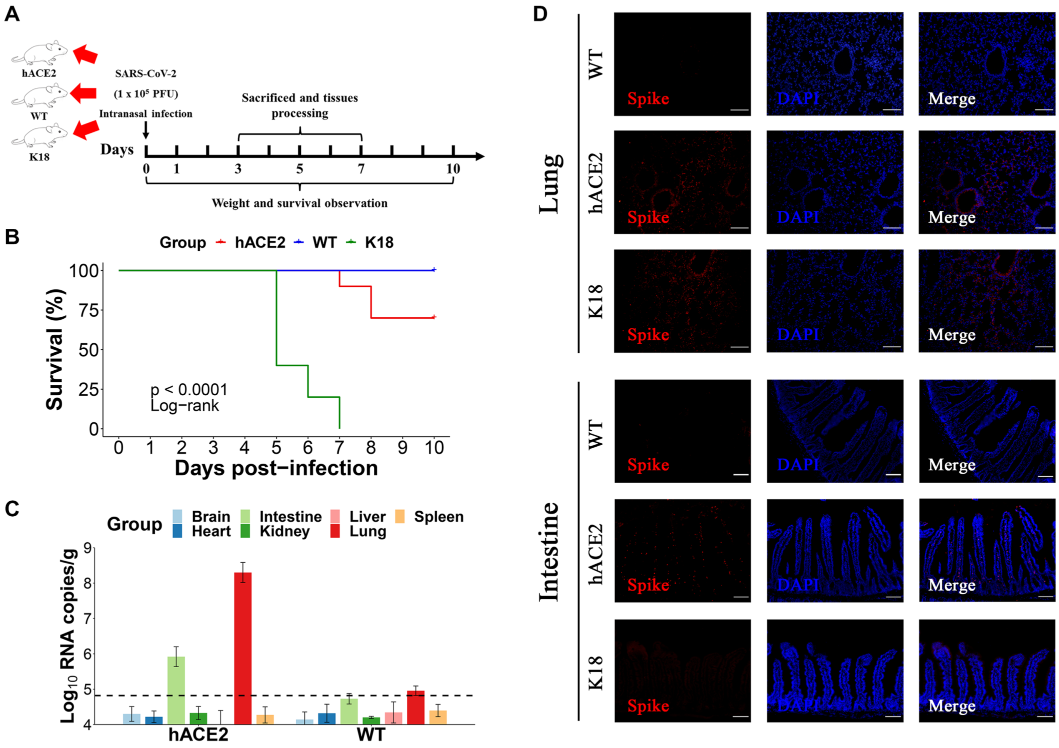


FIGURE 3
Intranasal infection of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in hACE2-KI mice. **(A)** The hACE2-KI, WT, and K18-hACE2 mice received intranasal SARS-CoV-2 infection at a dose of 1×10^5 plaque-forming units (PFU)/50μL per mouse, and three mice from each group were euthanized at 3, 5, and 7days post-infection (dpi) for tissue collection. **(B)** Survival of hACE2-KI (red), WT (blue), and K18-hACE2 (green) mice were observed for 10days. The statistical difference was evaluated by log-rank test. **(C)** Viral mRNA levels in brain, heart, kidney, lung, intestines, liver, and spleen are shown. The dotted lines indicate the detection limit. **(D)** Immunofluorescence staining analysis for the spike protein (S) (red) of SARS-CoV-2 in lungs and intestines; scale bar: 100μm.

In the intestines, the main infiltrated immune cells were T lymphocytes (Figure 6A); however, there were no notable differences in CD3$^+$, CD4$^+$ and CD8$^+$ T lymphocyte counts among the hACE2-KI, K18-hACE2, and WT mice groups (Figure 6B). When comparing infiltrated immune cell counts, only hACE2-KI mice exhibited elevated CD19$^+$ B lymphocytes and Ly6G$^+$ neutrophils (Figure 6B). Therefore, this study provided evidence of reduced inflammatory responses in the intestinal tract, which is in contrast to the significant inflammatory responses found in lung tissues. We have summarized the main similarity and differences between hACE2-KI and K18-hACE2 mouse models found in this study (Table 2).

4. Discussion

Although most people have received vaccinations for COVID-19, SARS-CoV-2 has been spreading at an alarming rate. Therefore, the development of small animal models that could show COVID-19 clinical manifestations similar to those in humans is essential for enhancing the understanding of candidate therapeutics. In the present study, we report a novel mouse model in which the CDS of endogenous ACE2 was precisely replaced with that of human ACE2 *in situ* using CRISPR/Cas9 technology. Driven by the endogenous mouse ACE2 promoter, the expression of hACE2 was assumed to be physiologically distributed in crosses tissues. Immunochemistry, real-time qPCR, and immunoblotting were performed to confirm this hypothesis, and it was found that hACE2-KI was highly expressed in the lungs, trachea, and intestines. Interestingly, analyzes of single-cell transcriptomic data in humans also revealed that ACE2 and TMPRSS2 were co-expressed in the intestines (Zhang et al., 2020). High hACE2 expression that is commonly observed in the brains of K18-hACE2 mice was not detected in this established model (Gan et al., 2021; Dong et al., 2022). In contrast, the robust hACE2 expression identified in the intestinal brush border was consistent with the finding that human intestinal tissues also show a high hACE2 expression (Hamming et al., 2004). It is worth mentioning that this tissue distribution pattern was not observed in other mouse models used for the SARS-CoV-2 research (Bao et al., 2020; Jiang et al., 2020; Asaka et al., 2021; Sefik et al., 2021, 2022; Winkler et al., 2022).

Furthermore, the susceptibility of this model to SARS-CoV-2 was evaluated at a dose of 1 × 10^5 PFU/mouse. Following infection, the mice exhibited a considerable reduction in body weight, which could serve as a good indicator for monitoring disease progression. In survival analysis, this mouse model showed a certain ratio of fatal cases. In comparison, no obvious clinical symptoms or mortality was observed in other hACE2 transgenic models that did not seem to recapitulate severe diseases. Although the K18-hACE2 model had a

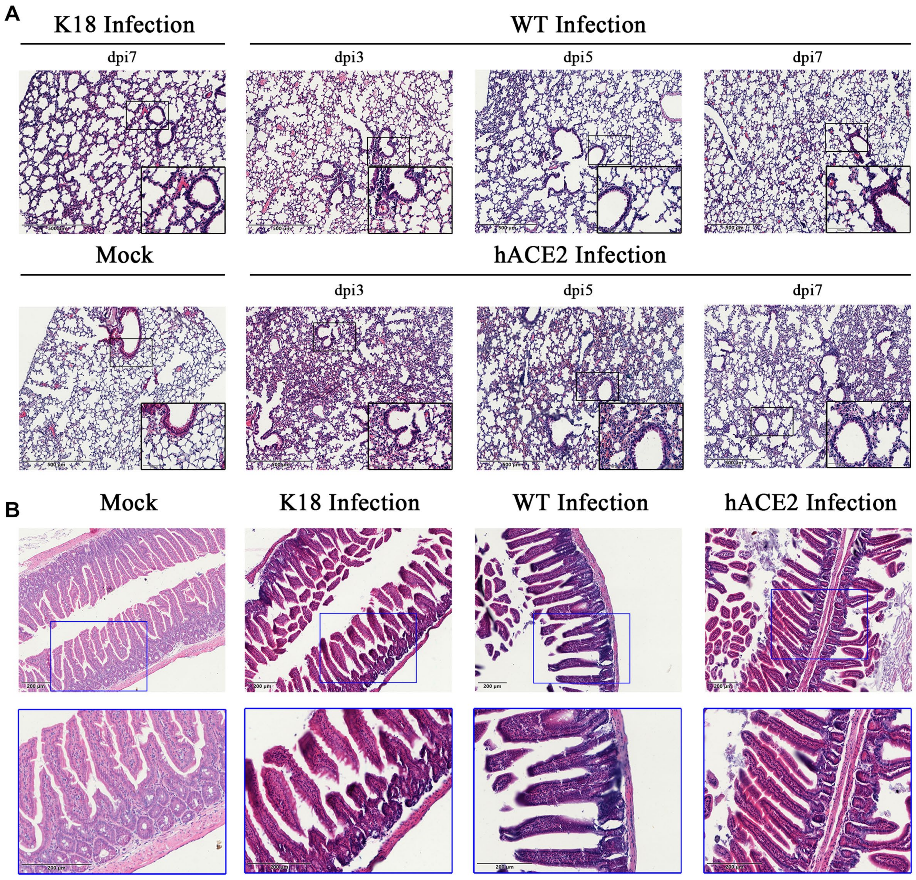


FIGURE 4
Histopathological changes in hACE2-KI mice infected with SARS-CoV-2. **(A)** Hematoxylin and eosin (H&E) staining of lung sections from hACE2-KI and WT mice after intranasal inoculation with 1x10^5 PFU per mouse at 3, 5, and 7 dpi. K18-hACE2 lung tissue collected at 7 dpi is shown. **(B)** H&E staining of intestine sections from hACE2-KI, WT, and K18-hACE2 mice after intranasal inoculation with 1x 10^5 PFU per mouse at 7 dpi. Images are representative of each study group.

higher mortality rate, most of these mice showed neuroinvasion and high titer of viruses (demonstrating viral replication) in the brain, which was assumed to be the primary cause of lethality (Jiang et al., 2020; Rathnasinghe et al., 2020; Zhang et al., 2021; Dong et al., 2022). This is in contrast with the findings that most COVID-19 patients died due to lung injury (Hu et al., 2021). Notably, evidence of enteric infection, which is seldom identified in other mouse models, was also found and this may be due to the physiological distribution of hACE2 expression in the intestines.

To further investigate the pathological damage, H&E staining of lung and intestinal tissues was performed. After SARS-CoV-2 infection, hACE2-expressing mice showed signs of pneumonia accompanied by immune cell infiltration. Immunofluorescence assays demonstrated that the extent of the inflammatory response in this model was comparable with that observed in K18-hACE2 mice. These results indicate that this mouse model successfully developed lung damage, which is a common histopathological change that manifests in COVID-19 patients (Bradley et al., 2020; D'Onofrio et al., 2022). In addition, damage in the intestines with limited immune cell infiltration was demonstrated. Compared with K18-hACE2 and WT mice, this model only exhibited a slightly higher proportion of $CD19^+$ B lymphocytes and $Ly6G^+$ neutrophils. Another study, which included clinical cohorts, also observed evidence of reduced inflammatory responses in the intestinal tract (Livanos et al., 2021). The attenuation of inflammation is consistent with data from autopsy studies conducted on COVID-19 patients (Skok et al., 2021).

Notably, recent clinical studies have reported that an increasing proportion of COVID-19 patients develop gastrointestinal symptoms (Chen et al., 2020; Jin et al., 2020; Redd et al., 2020; Blackett et al., 2022; Freedberg and Chang, 2022; Jin et al., 2022) and direct evidence of active SARS-CoV-2 replication has been found in the intestines (Qian et al., 2021; Zuo et al., 2021; Cuicchi et al., 2022). Intestinal infection can induce the infiltration of plasma cells and lymphocytes in the intestinal mucosa of COVID-19 patients (Xiao et al., 2020). Consistent with these reports, elevated levels of $CD19^+$ B lymphocytes and infiltration of $CD3^+$ T lymphocytes were observed. In addition, intestinal immune responses induced by enteric infections may regulate lung symptoms through the gut-lung axis (Yang et al., 2021). It is reasonable to assume that targeting the intestines may be a candidate therapeutic strategy for patients with COVID-19. However, there is no appropriate mouse model that can mimic the intestinal infection process with accurate physiological manifestations, and the

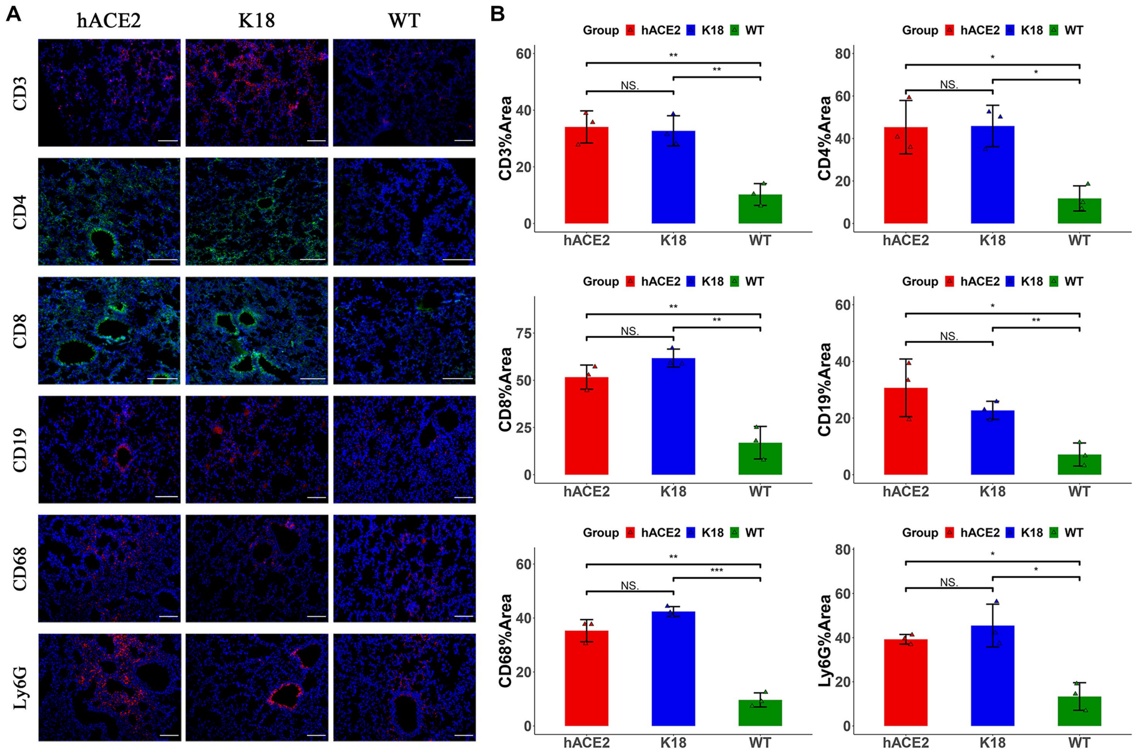


FIGURE 5
Characterization of immune responses in lungs of hACE2-KI mice infected with SARS-CoV-2. **(A)** Representative images of immunofluorescence staining analysis for the immune cells in lung tissues collected at 7 dpi. Cell markers: CD3, CD19, CD68, and Ly6G (red); CD4, CD8 (green). Scale bar: 100μm. **(B)** Semi-quantitative analysis of area fraction of immune cells with immunofluorescence staining (n=3 fields per group). Statistical analysis was performed using unpaired Student's t-test or Wilcoxon rank test. NS: not significant; *p<0.05; **p<0.01; ***p<0.001.

TABLE 2 Comparisons of similarity and difference between hACE2-KI and K18-hACE2 mouse model.

Characteristics		
	hACE2-KI model	K18-hACE2 model
Promotor of hACE2 expression	Mouse endogenous promoter	Human keratin 18 promoter
Expression of hACE2 in tissues	Intestine, lung, trachea and kidney	Brain, trachea, lung, and kidney (Dong et al., 2022)
Mortality after SARS-CoV-2 infection	30% within 10 dpi at 10^5 PFU	60–100% within 10 dpi at 2×10^1–10^5 PFU (Dong et al., 2022)
Viral infection pattern in tissues	Lung and intestine	Brain, trachea and lung (Dong et al., 2022)
Immune responses in lung	Both exhibited elevated immune responses in the lung tissues, but there were no remarkable differences in inflammatory cell counts	
Immune responses in intestine	Higher CD19+ B lymphocytes and Ly6G+ neutrophils in hACE2-KI model	

current *in vitro* models used for SARS-CoV-2 intestinal infection are human cell lines and organoids (Stanifer et al., 2020; Zhou et al., 2020). Consequently, it would be of great benefit to develop a COVID-19 mouse model that can be used to study intestinal infection.

Emerging evidence has demonstrated that the intestinal immune microenvironment can be regulated by many factors excluding gut immune cells, such as microbiome composition and microbial metabolites (Haak et al., 2018; Sencio et al., 2020). Alterations in gut microbiota have already been found not only in animal studies but also in clinical research (Gu et al., 2020; Zuo et al., 2020; Seibert et al., 2021; Yeoh et al., 2021; Blackett et al., 2022; Cuicchi et al., 2022). A limitation of this study is that the regulatory mechanism of intestinal infection during COVID-19 disease progression was not further investigated. Nevertheless, the established mouse model in this study showed similarities to COVID-19 patients in the tissue distribution of hACE2 expression and SARS-CoV-2 infection characteristics in both the lungs and intestines. Therefore, this model could be a useful tool for studying SARS-CoV-2 intestinal infection and developing preventive and therapeutic drugs for COVID-19 using different strategies.

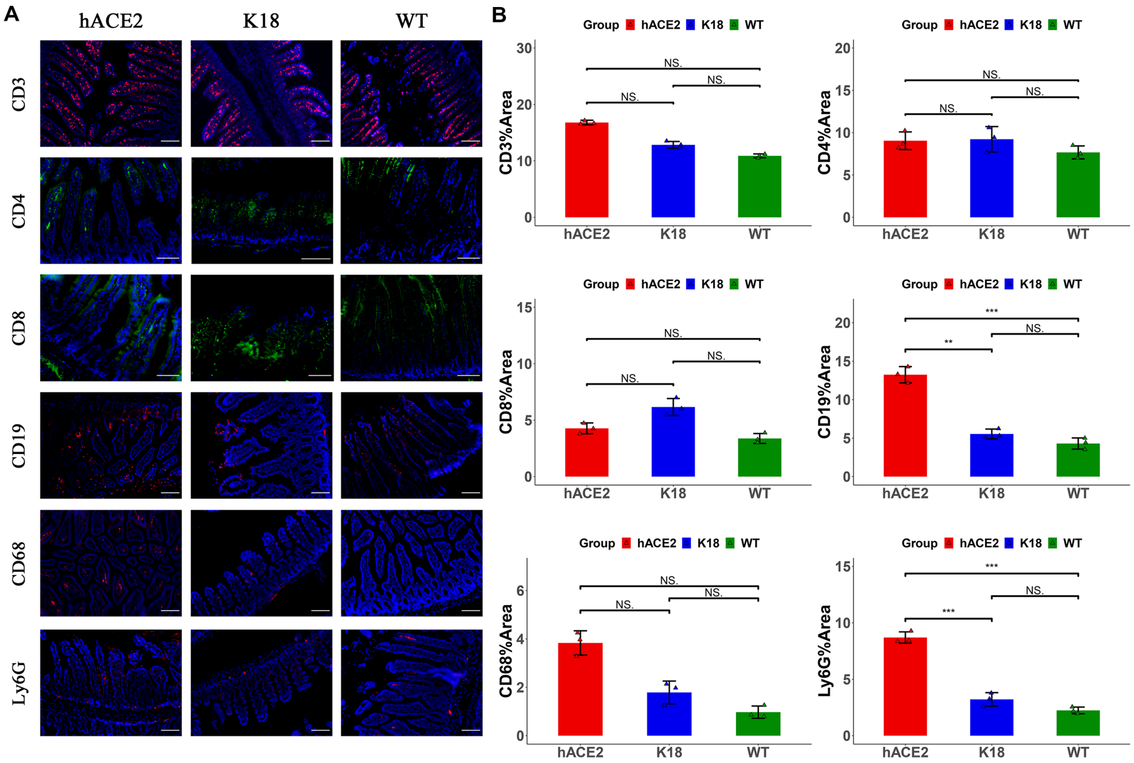


FIGURE 6
Characterization of immune responses in intestines of hACE2-KI mice infected with SARS-CoV-2. **(A)** Representative images of immunofluorescence staining analysis for the immune cells in intestinal tissues collected at 7 dpi. Cell markers: CD3, CD19, CD68, and Ly6G (red); CD4, CD8 (green). Scale bar: 100μm. **(B)** Semi-quantitative analysis of area fraction of immune cells with immunofluorescence staining (n=3 fields per group). Statistical analysis was performed using unpaired Student's t-test or Wilcoxon rank test. NS: not significant; *p<0.05; **p<0.01; ***p<0.001.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Ethics statement

All experimental procedures involving mice in this study were approved by the Laboratory Animal Welfare and Ethics Committee of the Third Military Medical University (AMUWE20201373).

Author contributions

JL conceived and planned the overall structure of the article. XZ, WS, YZ, and HG contributed to draft the manuscript and development of methodology. RW, PX, YKZ, MQ, and XD performed the experiments and prepared the figures and tables. JL, HW, and YG edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The project was supported by National Natural Science Foundation of China (Grant No. 81570497) and State Key Laboratory of Pathogen and Biosecurity (Academy of Military Medical Science, SKLPBS2106).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Asaka, M. N., Utsumi, D., Kamada, H., Nagata, S., Nakachi, Y., Yamaguchi, T., et al. (2021). Highly susceptible SARS-CoV-2 model in CAG promoter-driven hACE2-transgenic mice. *JCI Insight.* 6:e152529. doi: 10.1172/jci.insight.152529

Bao, L., Deng, W., Huang, B., Gao, H., Liu, J., Ren, L., et al. (2020). The pathogenicity of SARS-CoV-2 in hACE2 transgenic mice. *Nature* 583, 830–833. doi: 10.1038/s41586-020-2312-y

Blackett, J. W., Sun, Y., Purpura, L., Margolis, K. G., Elkind, M. S. V., O'Byrne, S., et al. (2022). Decreased gut microbiome tryptophan metabolism and serotonergic signaling in patients with persistent mental health and gastrointestinal symptoms after COVID-19. *Clin. Transl. Gastroenterol.* 13:e00524. doi: 10.14309/ctg.0000000000000524

Bradley, B. T., Maioli, H., Johnston, R., Chaudhry, I., Fink, S. L., Xu, H., et al. (2020). Histopathology and ultrastructural findings of fatal COVID-19 infections in Washington state: a case series. *Lancet* 396, 320–332. doi: 10.1016/S0140-6736(20)31305-2

Chen, A., Agarwal, A., Ravindran, N., To, CZhang, T., and Thuluvath, P. J. (2020). Are gastrointestinal symptoms specific for coronavirus 2019 infection? A prospective case-control study from the United States. *Gastroenterology* 159, 1161–1163.e2. doi: 10.1053/j.gastro.2020.05.036

Cuicchi, D., Gabrielli, L., Tardio, M. L., Rossini, G., D'Errico, A., Viale, P., et al. (2022). Virological and histological evaluation of intestinal samples in COVID-19 patients. *World J. Gastroenterol.* 28, 6282–6293. doi: 10.3748/wjg.v28.i44.6282

D'Onofrio, V., Keulen, L., Vandendriessche, A., Dubois, J., Cartuyvels, R., Vanden Abeele, M. E., et al. (2022). Studying the clinical, radiological, histological, microbiological, and immunological evolution during the different COVID-19 disease stages using minimal invasive autopsy. *Sci. Rep.* 12:1360. doi: 10.1038/s41598-022-05186-y

Dong, W., Mead, H., Tian, L., Park, J. G., Garcia, J. I., Jaramillo, S., et al. (2022). The K18-human ACE2 transgenic mouse model recapitulates non-severe and severe COVID-19 in response to an infectious dose of the SARS-CoV-2 virus. *J. Virol.* 96:e0096421. doi: 10.1128/JVI.00964-21

Empson, S., Rogers, A. J., and Wilson, J. G. (2022). COVID-19 acute respiratory distress syndrome: one pathogen, multiple phenotypes. *Crit. Care Clin.* 38, 505–519. doi: 10.1016/j.ccc.2022.02.001

Freedberg, D. E., and Chang, L. (2022). Gastrointestinal symptoms in COVID-19: the long and the short of it. *Curr. Opin. Gastroenterol.* 38, 555–561. doi: 10.1097/MOG.0000000000000876

Gan, E. S., Syenina, A., Linster, M., Ng, B., Zhang, S. L., Watanabe, S., et al. (2021). A mouse model of lethal respiratory dysfunction for SARS-CoV-2 infection. *Antivir. Res.* 193:105138. doi: 10.1016/j.antiviral.2021.105138

Gu, S., Chen, Y., Wu, Z., Chen, Y., Gao, H., Lv, L., et al. (2020). Alterations of the gut microbiota in patients with coronavirus disease 2019 or H1N1 influenza. *Clin. Infect. Dis.* 71, 2669–2678. doi: 10.1093/cid/ciaa709

Haak, B. W., Littmann, E. R., Chaubard, J. L., Pickard, A. J., Fontana, E., Adhi, F., et al. (2018). Impact of gut colonization with butyrate-producing microbiota on respiratory viral infection following Allo-HCT. *Blood* 131, 2978–2986. doi: 10.1182/blood-2018-01-828996

Hajifathalian, K., Krisko, T., Mehta, A., Kumar, S., Schwartz, R., Fortune, B., et al. (2020). Gastrointestinal and hepatic manifestations of 2019 novel coronavirus disease in a large cohort of infected patients from New York: clinical implications. *Gastroenterology* 159, 1137–1140.e2. doi: 10.1053/j.gastro.2020.05.010

Hamming, I., Timens, W., Bulthuis, M. L., Lely, A. T., Navis, G., and 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. doi: 10.1002/path.1570

Hassan, A. O., Case, J. B., Winkler, E. S., Thackray, L. B., Kafai, N. M., Bailey, A. L., et al. (2020). A SARS-CoV-2 infection model in mice demonstrates protection by neutralizing antibodies. *Cells* 182, 744–753.e4. doi: 10.1016/j.cell.2020.06.011

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cells* 181, 271–280.e8. doi: 10.1016/j.cell.2020.02.052

Hu, B., Guo, H., Zhou, P., and Shi, Z. L. (2021). Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* 19, 141–154. doi: 10.1038/s41579-020-00459-7

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan. *Lancet* 395, 497–506. doi: 10.1016/S0140-6736(20)30183-5

Jiang, R. D., Liu, M. Q., Chen, Y., Shan, C., Zhou, Y. W., Shen, X. R., et al. (2020). Pathogenesis of SARS-CoV-2 in transgenic mice expressing human angiotensin-converting enzyme 2. *Cells* 182, 50–58.e8. doi: 10.1016/j.cell.2020.05.027

Jin, X., Lian, J. S., Hu, J. H., Gao, J., Zheng, L., Zhang, Y. M., et al. (2020). Epidemiological, clinical and virological characteristics of 74 cases of coronavirus-infected disease 2019 (COVID-19) with gastrointestinal symptoms. *Gut* 69, 1002–1009. doi: 10.1136/gutjnl-2020-320926

Jin, S., Lu, X., and Xu, C. (2022). COVID-19 induces gastrointestinal symptoms and affects patients' prognosis. *J. Int. Med. Res.* 50:211295. doi: 10.1177/03000605221129543

Liang, W., Feng, Z., Rao, S., Xiao, C., Xue, X., Lin, Z., et al. (2020). Diarrhoea may be underestimated: a missing link in 2019 novel coronavirus. *Gut* 69, 1141–1143. doi: 10.1136/gutjnl-2020-320832

Lin, L., Jiang, X., Zhang, Z., Huang, S., Zhang, Z., Fang, Z., et al. (2020). Gastrointestinal symptoms of 95 cases with SARS-CoV-2 infection. *Gut* 69, 997–1001. doi: 10.1136/gutjnl-2020-321013

Livanos, A. E., Jha, D., Cossarini, F., Gonzalez-Reiche, A. S., Tokuyama, M., Aydillo, T., et al. (2021). Intestinal host response to SARS-CoV-2 infection and COVID-19 outcomes in patients with gastrointestinal symptoms. *Gastroenterology* 160, 2435–2450.e34. doi: 10.1053/j.gastro.2021.02.056

Muus, C., Luecken, M. D., Eraslan, G., Sikkema, L., Waghray, A., Heimberg, G., et al. (2021). Single-cell meta-analysis of SARS-CoV-2 entry genes across tissues and demographics. *Nat. Med.* 27, 546–559. doi: 10.1038/s41591-020-01227-z

Nobel, Y. R., Phipps, M., Zucker, J., Lebwohl, B., Wang, T. C., Sobieszczyk, M. E., et al. (2020). Gastrointestinal symptoms and coronavirus disease 2019: a case-control study from the United States. *Gastroenterology* 159, 373–375.e2. doi: 10.1053/j.gastro.2020.04.017

Qian, Q., Fan, L., Liu, W., Li, J., Yue, J., Wang, M., et al. (2021). Direct evidence of active SARS-CoV-2 replication in the intestine. *Clin. Infect. Dis.* 73, 361–366. doi: 10.1093/cid/ciaa925

Ramadori, G. P. (2022). SARS-CoV-2-infection (COVID-19): clinical course, viral acute respiratory distress syndrome (ARDS) and cause(s) of death. *Med. Sci.* 10:58. doi: 10.3390/medsci10040058

Rathnasinghe, R., Strohmeier, S., Amanat, F., Gillespie, V. L., Krammer, F., García-Sastre, A., et al. (2020). Comparison of transgenic and adenovirus hACE2 mouse models for SARS-CoV-2 infection. *Emerg. Microbes Infect.* 9, 2433–2445. doi: 10.1080/22221751.2020.1838955

Redd, W. D., Zhou, J. C., Hathorn, K. E., McCarty, T. R., Bazarbashi, A. N., Thompson, C. C., et al. (2020). Prevalence and characteristics of gastrointestinal symptoms in patients with severe acute respiratory syndrome coronavirus 2 infection in the United States: a multicenter cohort study. *Gastroenterology* 159, 765–767.e2. doi: 10.1053/j.gastro.2020.04.045

Sefik, E., Israelow, B., Mirza, H., Zhao, J., Qu, R., Kaffe, E., et al. (2022). A humanized mouse model of chronic COVID-19. *Nat. Biotechnol.* 40, 906–920. doi: 10.1038/s41587-021-01155-4

Sefik, E., Israelow, B., Zhao, J., Qu, R., Song, E., Mirza, H., et al. (2021). A humanized mouse model of chronic COVID-19 to evaluate disease mechanisms and treatment options. *Res Sq.* 2021:279341. doi: 10.21203/rs.3.rs-279341/v1

Seibert, B., Cáceres, C. J., Cardenas-Garcia, S., Carnaccini, S., Geiger, G., Rajao, D. S., et al. (2021). Mild and severe SARS-CoV-2 infection induces respiratory and intestinal microbiome changes in the K18-hACE2 transgenic mouse model. *Microbiol. Spectr.* 9:e0053621. doi: 10.1128/Spectrum.00536-21

Sencio, V., Barthelemy, A., Tavares, L. P., Machado, M. G., Soulard, D., Cuinat, C., et al. (2020). Gut dysbiosis during influenza contributes to pulmonary pneumococcal superinfection through altered short-chain fatty acid production. *Cell Rep.* 30, 2934–2947.e6. doi: 10.1016/j.celrep.2020.02.013

Skok, K., Vander, K., Setaffy, L., Kessler, H. H., Aberle, S., Bargfrieder, U., et al. (2021). COVID-19 autopsies: procedure, technical aspects and cause of fatal course. Experiences from a single-center. *Pathol. Res. Pract.* 217:153305. doi: 10.1016/j.prp.2020.153305

Stanifer, M. L., Kee, C., Cortese, M., Zumaran, C. M., Triana, S., Mukenhirn, M., et al. (2020). Critical role of type III interferon in controlling SARS-CoV-2 infection in human intestinal epithelial cells. *Cell Rep.* 32:107863. doi: 10.1016/j.celrep.2020.107863

Sun, S. H., Chen, Q., Gu, H. J., Yang, G., Wang, Y. X., Huang, X. Y., et al. (2020). A mouse model of SARS-CoV-2 infection and pathogenesis. *Cell Host Microbe* 28, 124–133.e4. doi: 10.1016/j.chom.2020.05.020

Trompette, A., Gollwitzer, E. S., Pattaroni, C., Lopez-Mejia, I. C., Riva, E., Pernot, J., et al. (2018). Dietary fiber confers protection against flu by shaping Ly6c(−) patrolling monocyte hematopoiesis and CD8(+) T cell metabolism. *Immunity* 48, 992–1005.e8. doi: 10.1016/j.immuni.2018.04.022

Wan, Y., Shang, J., Graham, R., Baric, R. S., and Li, F. (2020). Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *J. Virol.* 94:e00127. doi: 10.1128/JVI.00127-20

Wang, Y., and Perlman, S. (2022). COVID-19: inflammatory profile. *Annu. Rev. Med.* 73, 65–80. doi: 10.1146/annurev-med-042220-012417

Winkler, E. S., Chen, R. E., Alam, F., Yildiz, S., Case, J. B., Uccellini, M. B., et al. (2022). SARS-CoV-2 causes lung infection without severe disease in human ACE2 knock-in mice. *J. Virol.* 96:e0151121. doi: 10.1128/JVI.01511-21

Wölfel, R., Corman, V. M., Guggemos, W., Seilmaier, M., Zange, S., Müller, M. A., et al. (2020). Virological assessment of hospitalized patients with COVID-2019. *Nature* 581, 465–469. doi: 10.1038/s41586-020-2196-x

Xiao, F., Tang, M., Zheng, X., Liu, Y., Li, X., and Shan, H. (2020). Evidence for gastrointestinal infection of SARS-CoV-2. *Gastroenterology* 158, 1831–1833.e3. doi: 10.1053/j.gastro.2020.02.055

Yang, Y., Huang, W., Fan, Y., and Chen, G. Q. (2021). Gastrointestinal microenvironment and the gut-lung axis in the immune responses of severe COVID-19. *Front. Mol. Biosci.* 8:647508. doi: 10.3389/fmolb.2021.647508

Yeoh, Y. K., Zuo, T., Lui, G. C., Zhang, F., Liu, Q., Li, A. Y., et al. (2021). Gut microbiota composition reflects disease severity and dysfunctional immune responses in patients with COVID-19. *Gut* 70, 698–706. doi: 10.1136/gutjnl-2020-323020

Zhang, H., Kang, Z., Gong, H., Xu, D., Wang, J., Li, Z., et al. (2020). The digestive system is a potential route of 2019-nCov infection: a bioinformatics analysis based on single-cell transcriptomes. *bioRxiv* 2020:927806. doi: 10.1101/2020.01.30.927806

Zhang, L., Zhou, L., Bao, L., Liu, J., Zhu, H., Lv, Q., et al. (2021). SARS-CoV-2 crosses the blood-brain barrier accompanied with basement membrane disruption without tight junctions alteration. *Signal Transduct. Target. Ther.* 6:337. doi: 10.1038/s41392-021-00719-9

Zhou, J., Li, C., Liu, X., Chiu, M. C., Zhao, X., Wang, D., et al. (2020). Infection of bat and human intestinal organoids by SARS-CoV-2. *Nat. Med.* 26, 1077–1083. doi: 10.1038/s41591-020-0912-6

Zuo, T., Liu, Q., Zhang, F., Lui, G. C., Tso, E. Y., Yeoh, Y. K., et al. (2021). Depicting SARS-CoV-2 faecal viral activity in association with gut microbiota composition in patients with COVID-19. *Gut* 70, 276–284. doi: 10.1136/gutjnl-2020-322294

Zuo, T., Zhang, F., Lui, G. C. Y., Yeoh, Y. K., Li, A. Y. L., Zhan, H., et al. (2020). Alterations in gut microbiota of patients with COVID-19 during time of hospitalization. *Gastroenterology* 159, 944–955.e8. doi: 10.1053/j.gastro.2020.05.048

Frontiers in Microbiology

TYPE Original Research
PUBLISHED 14 July 2023
DOI 10.3389/fmicb.2023.1188380

OPEN ACCESS

EDITED BY
Shailendra Saxena,
King George's Medical University, India

REVIEWED BY
Ángel L. Álvarez,
University of Oviedo, Spain
Margarida Duarte,
National Institute for Agrarian and Veterinary Research (INIAV), Portugal
Lauro Velazquez-Salinas,
Agricultural Research Service (USDA),
United States

*CORRESPONDENCE
Jian Huang
✉ hjvet03@sina.cn

RECEIVED 17 March 2023
ACCEPTED 26 June 2023
PUBLISHED 14 July 2023

CITATION
Li Y, Du D, Zhou L, Su L, You C, Zhang H, Yu J, Xiao L and Huang J (2023) First report of GI.1aP-GI.2 recombinants of rabbit hemorrhagic disease virus in domestic rabbits in China.
Front. Microbiol. 14:1188380.
doi: 10.3389/fmicb.2023.1188380

First report of GI.1aP-GI.2 recombinants of rabbit hemorrhagic disease virus in domestic rabbits in China

Yan Li[1], Deyan Du[2], Long Zhou[1], Liyin Su[1], Chengcheng You[1], Huai Zhang[1], Jifeng Yu[3], Lu Xiao[3] and Jian Huang[1]*

[1]College of Animal Science and Veterinary Medicine, Southwest Minzu University, Chengdu, China, [2]Huapai Biological Group, Chengdu, China, [3]Sichuan Animal Science Academy, Sichuan Provincial Key Laboratory of Animal Breeding and Genetics, Chengdu, China

The rabbit hemorrhagic disease virus 2 (RHDV2 or GI.2) is a highly contagious agent leading to lethal disease in rabbits. It frequently recombines with other *Lagovirus* genus, generating epidemical variants with high pathogenicity. In this study, twenty-two liver samples tested positive for GI.2 VP60 gene, were collected in rabbit farms from several geographical regions in China. All GI.2 positive specimens were submitted for RT-PCR detection, nucleotide sequencing and phylogenetic analysis. In addition, suspected GI.2 recombinants were evaluated for virus virulence. The results showed that nine presumptive recombinants were identified by testing for RdRp-VP60 recombination. In these recombinants, four were selected to fully characterize the genome of novel GI.2 recombinant variants, which were described as GI.1aP-GI.2. The nucleotide sequence of these novel variants showed unique recombination pattern and phylogenetic features compared to currently prevalent GI.2 variants. Furthermore, this distinctive recombination of new variant SCNJ-2021 moderately enhanced the virulence of GI.2, even for rabbits vaccinated against parental GI.2. In conclusion, the novel GI.1aP-GI.2 recombinants were identified in rabbit industry in China for the first time, which expanded the knowledge on the phylodynamics and genomic diversity of GI.2 genotype. The rapid molecular evolution and varied pathogenicity of these virus recombinants highlight the urgent need for epidemiological surveillance and for future prevention of these neglected GI.2 variants.

KEYWORDS

recombination, GI.1aP-GI.2 variants, phylogenetic features, rabbit hemorrhagic disease virus 2, pathogenicity

1. Introduction

The rabbit hemorrhagic disease virus (RHDV) is a common and highly contagious agent that causes acute multi-organ hemorrhagic syndrome with high morbidity and mortality (Teifke et al., 2002; Neimanis et al., 2018a). RHDV is a positive-sense single-stranded RNA virus from the *Lagovirus* genus, *Caliciviridae* family (Lopes et al., 2017). Each particle contains a genome of about 7.4 kb and a subgenomic RNA (sgRNA) of about 2.5 kb. The genome encodes two open reading frames (ORFs) with slightly overlapping. ORF1 encodes a large polyprotein that is cleaved by a virus-encoded protease, generating seven non-structural proteins (NSPs) and the major structural protein (VP60), and ORF2 encodes a minor structural protein (VP10) (Le Pendu et al., 2017). Based on the phylogenetic classification and VP60 gene variability, the RHDVs were divided into GI.1 and GI.2 genotypes. The GI.1 genotype was further subdivided into classic RHDV (G1/GI.1b, G2/GI.1c and G3-G5/GI.1d) and antigenic variant RHDVa (G6/GI.1a) (Le Pendu et al., 2017).

Since the first outbreak in Jiangsu province, China, in 1984 (Liu et al., 1984), classic RHDVs that underwent constant evolution with cumulative genomic alterations presented varied etiologic and epidemiological features (Wang et al., 2012; Abrantes et al., 2020a). In the last twenty years, GI.1c and GI.1a genotypes co-circulated in China along with intergenotypic recombination during their widespread transmission (Hu et al., 2016, 2017). In 2010, a new RHDV variant was identified in France, named RHDVb or RHDV2 (GI.2), which showed distinctive genetic and antigenic characteristics compared to GI.1. Moreover, this virus exhibited low cross-protection with other *Lagoviruses* (Le Gall-Reculé et al., 2011). Subsequently, the highly pathogenic GI.2 damaged the rabbit industry heavily in Europe, Australia, Africa, and North America, which rapidly replaced GI.1 as the predominant genotype in the past decade (Dalton et al., 2012; Mahar et al., 2018; Neimanis et al., 2018b; Chehida et al., 2021; Aguayo-Adán et al., 2022). Owing to the rapid spread of RHDVs and the resulting economic and ecological losses, the emerging pathogenic GI.1a and GI.2 raised increased concern in recent years (Lopes et al., 2017; Pacioni et al., 2022). These circumstances may entail ongoing awareness of the genome diversity and the virulence alteration of GI.2 in consequence of its persistent transmission in rabbits.

In 2020, the GI.2 strain was identified in the Sichuan province in China (Hu et al., 2021), during an outbreak of the RHD, which was suspected to be the result of international import due to its high nucleotide homology with Netherlands isolates in 2016 (Qi et al., 2022). Recently, intergenotypic recombination between non-structural and structural genome segments derived from different genotypes was considered as the main mechanism of genetic evolution in Lagoviruses (Mahar et al., 2021). Hence this mechanism may also be a robust driver for GI.2 variants to expand host range and adaption. Hitherto, several recombination patterns have been confirmed for GI.2 variants, including intergenotypic recombination between pathogenic GI.1b and GI.2 (e.g., GI.1bP-GI.2), between non-pathogenic RCV and GI.2 (e.g., GI.4eP-GI.2, GI.4cP-GI.2, GI.3P-GI.2) (Lopes et al., 2017; Mahar et al., 2018; Silvério et al., 2018; Abrantes et al., 2020b). The GI.4eP-GI.2 and GI.4cP-GI.2 variants are progressively replacing the previous parental GI.2 in a relatively short period, strengthening the inference that genome substitution on the non-structural region may accelerate the evolutionary adaption of the virus (Mahar et al., 2021) and alter its virulence (Smertina et al., 2021).

To understand the dynamics of GI.2 in domestic rabbits since its invasion into mainland China, we characterized the genome of GI.2 variants of concern and confirmed their pathogenicity alteration in present study. Here, we first described recombination events between GI.1a and GI.2 on rabbit farms in China, which generated the novel GI.1aP-GI.2 variants. The results of this study emphasize the need to implement epidemiological surveillance of *Lagoviruses* to unravel their co-circulation and evolution, in order to adapt the prevention program.

2. Materials and methods

2.1. Sample collection and molecular detection

Twenty-two rabbit liver samples were collected from twelve rabbit farms affected by RHD in Sichuan, Shandong, Anhui, and Yunnan provinces, from May 2020 to November 2022. Total RNA was extracted from the liver samples using the RNAiso plus reagent (TaKaRa, China), then the reverse transcription was performed with the PrimeScript™ RT Reagent Kit (TaKaRa, China). All the samples were confirmed as GI.2 positive by a differential Taqman RT-PCR assay as previously described (Zhou et al., 2022).

Primers targeting the RdRp-VP60 region were designed using Primer Premier 6.0 software (PREMIER Biosoft, USA) to generate a 994 bp-long amplicon by RT-PCR, then the PCR products were sequenced using ABI 3730XL platform (Sangon Biotech Co., China) for further recombination analysis. Eight pairs of primers spanning the complete GI.2 genome were used to obtain the PCR products of five representative GI.2 strains. The PCR products were purified and inserted into the pMD19-T vector (TaKaRa, China), and at least three positive clones of each fragment were submitted for nucleic acid sequencing. Information on all primers and clinical samples were listed in Supplementary Table S1, S2.

2.2. Histopathology and transmission electron microscopy

The gross pathological findings in the dead rabbits were recorded, and subsequent necropsies were performed according to routine procedures. The liver tissue blocks were fixed in 4% paraformaldehyde for 24 h, then paraffin-embedded, sectioned at 4 μm, and stained with hematoxylin and eosin. The histopathology of the liver section was observed under the light microscope (Leica, Germany). For the transmission electron microscopy (TEM), virus particles were purified as described before (Hu et al., 2010) with a minor modification. The infected liver tissues were homogenized and quickly frozen and thawed to release the virus particles. The virus suspension was collected after centrifugation (10,000 g, 20 min) at 4°C. Then, the supernatant was treated with 6% (w/v) polyethylene glycol (PEG 6000) and 3% (w/v) NaCl overnight at 4°C. The precipitate was resuspended in PBS after low-speed centrifugation (4,450 g, 40 min, 4°C), and then combined with a mixture of butanol and isopentanol (24:1, v/v) and stirred for 5 min. The suspension was clarified by low-speed centrifugation (430 g, 40 min, 4°C). The aqueous phase was collected and centrifuged at 15,000 g for 40 min. The supernatant was sent to the Chengdu Lilai Biomedicine experiment center for virus particle detection under the TEM (JEOL, Japan).

2.3. Genome alignment and phylogenetic analysis

All sequences were retrieved from the GenBank database, including the representative genomic sequences of 61 *Lagoviruses* of different genotypes (Supplementary Table S3). The complete genome sequences of SCMS-2020 (GenBank accession: OQ570964), SCNJ-2021 (GenBank accession: OQ570963), SDRZ-2021 (GenBank accession: OQ570961), SCMS-2022 (GenBank accession: OQ570960), and AHFY-2022 (GenBank accession: OQ570962) were obtained by sequence assembly. The nucleotide and amino acids identity of the RHDV strains alignment were analyzed using the MegAlign program within DNASTAR 7.0 software (DNASTAR Inc., Madison, WI, USA). The phylogenetic analysis of complete genome sequences was performed using MEGA 10 with the maximum-likelihood approach based on NSPs fragments (nt positions 10-5304), VP60 fragments (nt positions 5305-7044), and complete genome using the GTR+ G+I model. The reliability of nodes was assessed by the bootstrap resampling procedure consisting of 1,000 replicates.

2.4. Recombination analysis

The Recombination Detection Program 4 (RDP4, v4.24) containing seven evaluation algorithms (RDP, Bootscan, GENECONV, MaxChi, Chimera, SiScan, and 3Seq) was used to confirm the putative recombination events and precise recombination breakpoints. Recombination events were deemed significant (value of $p \leq 1\times 10^{-6}$) when supported by at least five of the seven algorithms. SimPlot (v3.5.1, Baltimore, MD, USA) with a 200-bp window sliding along the genome (20-bp step size) was used to analyze the new variants. The recombinant *Lagoviruses* were defined using the nomenclature [RdRp genotype]P-[capsid genotype].

2.5. Hemagglutination and hemagglutination inhibition assays

The Hemagglutination (HA) and hemagglutination inhibition (HI) were performed as described previously (Mizoguchi et al., 2003; Song et al., 2017). For the HA, the liver tissue was homogenized on ice, then the supernatant was collected after centrifugation. Type B human red blood cells were washed in phosphate-buffered saline (PBS) and later centrifuged (280 g, 10 min) at room temperature. The RBC pellets were then resuspended and diluted in PBS (pH 7.2) to the final concentration of 1%. Then, 25 μL supernatant of liver homogenate was added into 96-well V-shaped bottom microtiter plates and two-fold serially diluted with equal volume of PBS (pH 7.2). Later, 25 μL of 1% human type B RBCs was added to each well and incubated at 25°C for 30–60 min. The HA titer was determined as the highest dilution that caused complete hemagglutination of RBCs.

For the HI, the collected sera was inactivated and pretreated with 25% kaolin (Macklin, China). Then, 25 μL of serum was added into 96-well V-shaped bottom microtiter plates and two-fold serially diluted with equal volume of PBS 25 μL of RHDV antigens (4 HAU) was added into each well and incubated at 25°C for 30–60 min. Subsequently, 25 μL of human type B RBCs were added into each well and settled at 25°C for 30–60 min. The highest dilution that caused complete inhibition was considered the hemagglutination inhibition titer. HI titer $\leq 2^3$ was considered as antibody negative.

2.6. Vaccine inoculation and lethal challenge with GI.1aP-GI.2 in experimental rabbits

Four-week-old (juvenile) and three-month-old (adult) New Zealand white rabbits were raised in the Experimental Animal Center of Huapai Biological Group Co., Ltd. (Chengdu, China). These rabbits were tested and shown to be seronegative to GI.1/GI.2 (HI titer$\leq 2^3$). After adaptive feeding for one week, 60 rabbits were randomly allocated into the unvaccinated group and the vaccinated group (receiving a single dose of inactivated bivalent RHDV vaccination consisting of inactivated GI.1a and GI.2 antigen). The inactivated bivalent RHDV vaccine was prepared as follows. First, the liver homogenates were prepared from naive rabbits that died of RHD with GI.1a or GI.2 challenge. After being inactivated by formaldehyde, the liver homogenates were mixed with a 1:1 ratio of each antigen (512 HAU of each genotype). The rabbits in the vaccinated group received a subcutaneous inoculation of 1-ml laboratory-produced bivalent vaccine. To determine the virus challenge dose, we implemented the calculation of median lethal dose (LD_{50}) by inoculating rabbits with a series of 10-fold dilutions of liver homogenates containing SCMS-2020 strain or SCNJ-2021 strain (5 rabbits per dilution). Then, the LD_{50} values were calculated by the Reed-Muench method (Supplementary Table S4). The LD_{50} values for SCMS-2020 strain and SCNJ-2021 strain were determined as $10^{-4.68}$ LD_{50} and $10^{-5.5}$ LD_{50} in 1-ml liver homogenates, separately. On the 14th day after vaccination, the rabbits in the vaccinated group (The antibody HI titers were between $2^7$and 2^9) were challenged with a dose of 10,000 LD_{50} for GI.2 (SCMS-2020 strain) or 10,000 LD_{50} for GI.1aP-GI.2 (SCNJ-2021 strain), respectively. The rabbits in the unvaccinated group were also challenged with a dose of 10,000 LD_{50} for each strain. The survival time and mortality were recorded within 96 h post-infection.

3. Results

3.1. Pathological findings and virus particle identification

The diseased rabbits succumbed to the RHD were examined at necropsy and evaluated histologically. The results showed that epistaxis was observed in approximate 6% of the rabbits undergoing acute and subacute RHD (Figure 1A). Enlarged, yellow-tan and mottled liver lobes were remarkably abnormal, and multifocal coagulopathy were also seen in lungs and other organs (Figure 1B) at necropsy. Significant histopathological lesions were confirmed in the liver of all animals. Evident cellular necrosis with hemorrhage appeared throughout the disarranged hepatic parenchyma, which was infiltrated by a large number of heterophils (Figure 1C). In addition, the evidence of virions in infected liver tissues was confirmed by the TEM. The visible icosahedral symmetry of the virus particles, approximately 30 nm in diameter with an inner shell, were consistent with GI.1/GI.2 (Figure 1D).

3.2. Identification of novel GI.1aP-GI.2 variants

Twenty-two rabbit liver samples tested positive for GI.2 VP60 gene, were collected in rabbit farms from Sichuan, Anhui, Shandong, and Yunnan provinces from 2020 to 2022 during RHD outbreaks (Supplementary Table S2) and none was positive for GI.1 VP60 gene. Then, we obtained the nucleotide sequences of the RdRp-VP60 junction derived from above GI.2 isolates. The recombinant analysis of the RdRp-VP60 junctions confirmed that nine of the 22 isolates (40.9%) were presumptive recombinant GI.1aP-GI.2 strains, while the others belonged to the parental GI.2 strains (Supplementary Figure S1; Supplementary Table S2). Eight overlapping fragments of each representative strain were obtained and sequenced

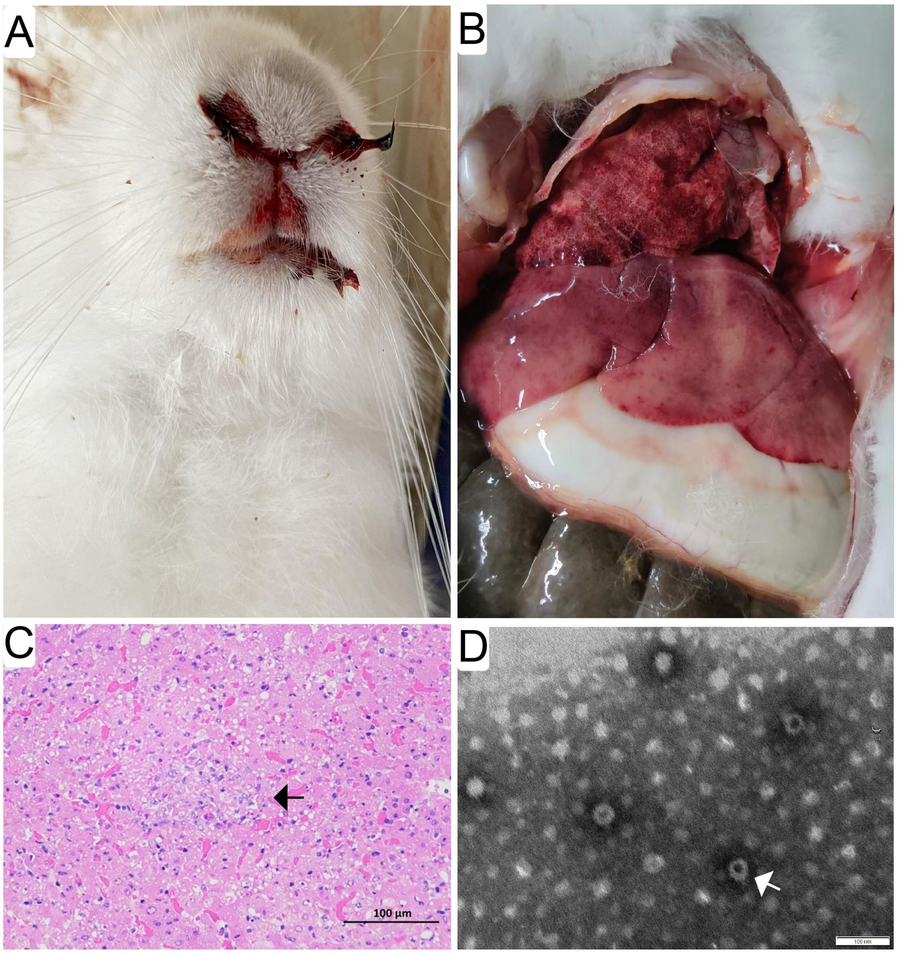


FIGURE 1
Gross pathological and histopathological findings in infected rabbits. Evident epistaxis was observed in approximate 6% of the infected rabbits undergoing acute and subacute RHD **(A)**. Enlarged, yellow-tan and mottled liver lobes were remarkable abnormal, and multifocal coagulopathy was also found in lungs and other organs **(B)** at necropsy. Significant histopathological lesions were confirmed in the liver and spleen of all animals. Evident cellular necrosis with hemorrhage appeared throughout the disarranged hepatic parenchyma, which was infiltrated by a large number of heterophils (white arrow) (bar=100 μm) **(C)**. The purified and negatively stained virus particles were shown (white arrow) under electron micrographs (bar=100 nm) **(D)**.

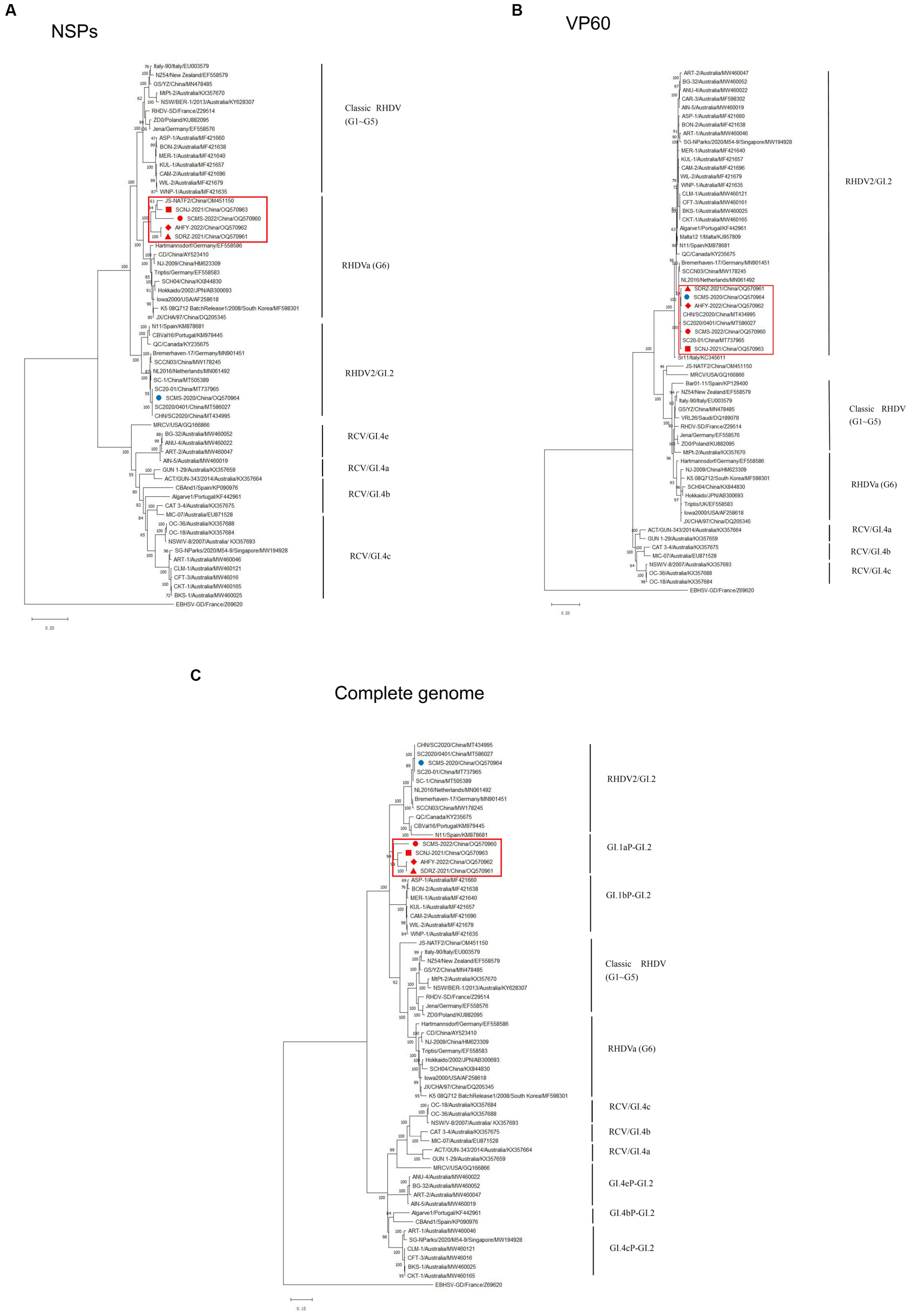
(Supplementary Figure S2). Nucleotide alignments of the consensus sequences confirmed that SCNJ-2021 (OQ570963), SDRZ-2021 (OQ570961), SCMS-2022 (OQ570960) and AHFY-2022 (OQ570962) were the recombinant strains belonging to the GI.1aP-GI.2 clade, whereas SCMS-2020 (OQ570964) was classified as a prototype of GI.2.

3.3. Phylogenetic features of the GI.aP-GI.2 variants

To unveil the genetic characterizations of these recombinant GI.2 variants, their genome sequences were comprehensively analyzed by bioinformatics tools. The ML phylogenetic trees based on the NSPs coding region (nt 10-5304, indicating the sequence used as reference of these positions), VP60 gene (nt 5305-7044, indicating the sequence used as reference of these positions), and complete genome were constructed separately. The genetic analysis based on the NSP genes revealed that SCNJ-2021, SDRZ-2021, SCMS-2022, and AHFY-2022 had 86.4-89.1% nucleotide identity and 96.0-97.8% amino acid identity with reference RHDVa strains (i.e., Triptis, Iowa2000, JX/CHA/97) and grouped into a new branch with a recently reported virus strain (JS-NATF2/China/OM451150) identified in *Oryctolagus cuniculus* (Figure 2A). However, the phylogenetic profiles based on the VP60 gene showed that these four recombinant strains closely clustered with other reference GI.2 strains from China (i.e., CHN/SC2020, SC2020/0401, SC20-01 and SC-1), and the nucleotide identity was up to 98.6-99.0% (Figure 2B). Significantly, the four variants (SCNJ-2021, SDRZ-2021, AHFY-2022, and SCMS-2022) branched into a monophyletic group showing 84.7-88.6% nucleotide identity with other representative strains based on the complete genome analysis (Figure 2C; Supplementary Tables S3, S6).

3.4. Recombination events of the GI.1aP-GI.2 variants

To further verify this novel recombination pattern, the recombination events of these four variants were analyzed by


A
NSPs
Italy-90/Italy/EU003579
NZ54/New Zealand/EF558579
GS/YZ/China/MN478485
MtPt-2/Australia/KX357670
NSW/BER-1/2013/Australia/KY628307
RHDV-SD/France/Z29514
ZD0/Poland/KU882095
Jena/Germany/EF558576
ASP-1/Australia/MF421660
BON-2/Australia/MF421638
MER-1/Australia/MF421640
KUL-1/Australia/MF421657
CAM-2/Australia/MF421696
WIL-2/Australia/MF421679
WNP-1/Australia/MF421635
JS-NATF2/China/OM451150
SCNJ-2021/China/OQ570963
SCMS-2022/China/OQ570960
AHFY-2022/China/OQ570962
SDRZ-2021/China/OQ570961
Hartmannsdorf/Germany/EF558586
CD/China/AY523410
NJ-2009/China/HM623309
Triptis/Germany/EF558583
SCH04/China/KX844830
Hokkaido/2002/JPN/AB300693
Iowa2000/USA/AF258618
K5 08Q712 BatchRelease1/2008/South Korea/MF598301
JX/CHA/97/China/DQ205345
N11/Spain/KM878681
CBVal16/Portugal/KM979445
QC/Canada/KY235675
Bremerhaven-17/Germany/MN901451
SCCN03/China/MW178245
NL2016/Netherlands/MN061492
SC-1/China/MT505389
SC20-01/China/MT737965
SCMS-2020/China/OQ570964
SC2020/0401/China/MT586027
CHN/SC2020/China/MT434995
MRCV/USA/GQ166866
BG-32/Australia/MW460052
ANU-4/Australia/MW460022
ART-2/Australia/MW460047
AIN-5/Australia/MW460019
GUN 1-29/Australia/KX357659
ACT/GUN-343/2014/Australia/KX357664
CBAnd1/Spain/KP090976
Algarve1/Portugal/KF442961
CAT 3-4/Australia/KX357675
MIC-07/Australia/EU871528
OC-36/Australia/KX357688
OC-18/Australia/KX357684
NSW/V-8/2007/Australia/ KX357693
SG-NParks/2020/M54-9/Singapore/MW194928
ART-1/Australia/MW460046
CLM-1/Australia/MW460121
CFT-3/Australia/MW46016
CKT-1/Australia/MW460165
BKS-1/Australia/MW460025
EBHSV-GD/France/Z69620
0.20
Classic RHDV (G1~G5)
RHDVa (G6)
RHDV2/GI.2
RCV/GI.4e
RCV/GI.4a
RCV/GI.4b
RCV/GI.4c
B
VP60
ART-2/Australia/MW460047
BG-32/Australia/MW460052
ANU-4/Australia/MW460022
CAR-3/Australia/MF598302
AIN-5/Australia/MW460019
ASP-1/Australia/MF421660
BON-2/Australia/MF421638
ART-1/Australia/MW460046
SG-NParks/2020/M54-9/Singapore/MW194928
MER-1/Australia/MF421640
KUL-1/Australia/MF421657
CAM-2/Australia/MF421696
WIL-2/Australia/MF421679
WNP-1/Australia/MF421635
CLM-1/Australia/MW460121
CFT-3/Australia/MW460161
BKS-1/Australia/MW460025
CKT-1/Australia/MW460165
Algarve1/Portugal/KF442961
Malta12 1/Malta/KJ957809
N11/Spain/KM878681
QC/Canada/KY235675
Bremerhaven-17/Germany/MN901451
SCCN03/China/MW178245
NL2016/Netherlands/MN061492
SDRZ-2021/China/OQ570961
SCMS-2020/China/OQ570964
AHFY-2022/China/OQ570962
CHN/SC2020/China/MT434995
SC2020/0401/China/MT586027
SCMS-2022/China/OQ570960
SC20-01/China/MT737965
SCNJ-2021/China/OQ570963
Sr11/Italy/KC345611
JS-NATF2/China/OM451150
MRCV/USA/GQ166866
Bar01-11/Spain/KP129400
NZ54/New Zealand/EF558579
Italy-90/Italy/EU003579
GS/YZ/China/MN478485
VRL26/Saudi/DQ189078
RHDV-SD/France/Z29514
Jena/Germany/EF558576
ZD0/Poland/KU882095
MtPt-2/Australia/KX357670
Hartmannsdorf/Germany/EF558586
NJ-2009/China/HM623309
K5 08Q712/South Korea/MF598301
SCH04/China/KX844830
Hokkaido/JPN/AB300693
Triptis/UK/EF558583
Iowa2000/USA/AF258618
JX/CHA/97/China/DQ205345
ACT/GUN-343/2014/Australia/KX357664
GUN 1-29/Australia/KX357659
CAT 3-4/Australia/KX357675
MIC-07/Australia/EU871528
NSW/V-8/2007/Australia/KX357693
OC-36/Australia/KX357688
OC-18/Australia/KX357684
EBHSV-GD/France/Z69620
0.20
RHDV2/GI.2
Classic RHDV (G1~G5)
RHDVa (G6)
RCV/GI.4a
RCV/GI.4b
RCV/GI.4c
C
Complete genome
CHN/SC2020/China/MT434995
SC2020/0401/China/MT586027
SCMS-2020/China/OQ570964
SC20-01/China/MT737965
SC-1/China/MT505389
NL2016/Netherlands/MN061492
Bremerhaven-17/Germany/MN901451
SCCN03/China/MW178245
QC/Canada/KY235675
CBVal16/Portugal/KM979445
N11/Spain/KM878681
SCMS-2022/China/OQ570960
SCNJ-2021/China/OQ570963
AHFY-2022/China/OQ570962
SDRZ-2021/China/OQ570961
ASP-1/Australia/MF421660
BON-2/Australia/MF421638
MER-1/Australia/MF421640
KUL-1/Australia/MF421657
CAM-2/Australia/MF421696
WIL-2/Australia/MF421679
WNP-1/Australia/MF421635
JS-NATF2/China/OM451150
Italy-90/Italy/EU003579
NZ54/New Zealand/EF558579
GS/YZ/China/MN478485
MtPt-2/Australia/KX357670
NSW/BER-1/2013/Australia/KY628307
RHDV-SD/France/Z29514
Jena/Germany/EF558576
ZD0/Poland/KU882095
Hartmannsdorf/Germany/EF558586
CD/China/AY523410
NJ-2009/China/HM623309
Triptis/Germany/EF558583
Hokkaido/2002/JPN/AB300693
SCH04/China/KX844830
Iowa2000/USA/AF258618
JX/CHA/97/China/DQ205345
K5 08Q712 BatchRelease1/2008/South Korea/MF598301
OC-18/Australia/KX357684
OC-36/Australia/KX357688
NSW/V-8/2007/Australia/ KX357693
CAT 3-4/Australia/KX357675
MIC-07/Australia/EU871528
ACT/GUN-343/2014/Australia/KX357664
GUN 1-29/Australia/KX357659
MRCV/USA/GQ166866
ANU-4/Australia/MW460022
BG-32/Australia/MW460052
ART-2/Australia/MW460047
AIN-5/Australia/MW460019
Algarve1/Portugal/KF442961
CBAnd1/Spain/KP090976
ART-1/Australia/MW460046
SG-NParks/2020/M54-9/Singapore/MW194928
CLM-1/Australia/MW460121
CFT-3/Australia/MW46016
BKS-1/Australia/MW460025
CKT-1/Australia/MW460165
EBHSV-GD/France/Z69620
0.10
RHDV2/GI.2
GI.1aP-GI.2
GI.1bP-GI.2
Classic RHDV (G1~G5)
RHDVa (G6)
RCV/GI.4c
RCV/GI.4b
RCV/GI.4a
GI.4eP-GI.2
GI.4bP-GI.2
GI.4cP-GI.2

(Continued)

FIGURE 2 Continued
Phylogenetic trees based on NSPs **(A)**, VP60 **(B)** and full-length genome sequence **(C)** of five isolates with 56 RHDV and RCV representative strains available in GenBank. Major genetic groups (genogroups) are indicated and the five isolates in this study are labeled. The phylogenetic tree was constructed by using the MLmethod (1,000 bootstrap) in MEGA 10. Numbers along branches are bootstrap values. Scale bar indicates nucleotide substitute per site.

Recombination Detection Program 4 and Simplot software. At least five methods were used to confirm the recombination of the complete genome of SCNJ-2021, SDRZ-2021, AHFY-2022, and SCMS-2022 variants by the analysis (p-values of $\leq 1 \times 10^{-6}$) (Supplementary Table S7). Similarity plot analysis confirmed the recombination breakpoints along the genomic RdRp-VP60 junction (nt 5240, nt 5274, or nt 5304) (Figure 3). Determined by the RDP4 software, the most likely parental variants for the four strains were pathogenic GI.1a donating the non-structural genome segment (Genbank accession EF558583) and pathogenic GI.2 donating the structural genome segment (Genbank accession MT586027).

3.5. Cross-protection between GI.2 and GI.1aP-GI.2 variant

In order to demonstrate whether the vaccine prepared from the prototype GI.2 strain conferred protection against the recombinant GI.2 variant, the challenge study was carried out. As expected, all unvaccinated juvenile and adult rabbits died of the challenge with the GI.2 or GI.1aP-GI.2 variant from 24 h to 96 h post-infection. However, the vaccination with GI.2 (SCMS-2020) provided complete protection against parental GI.2 infection and incomplete cross-protection against GI.1aP-GI.2 (SCNJ-2021) infection. The protection was slightly lower in vaccinated juvenile rabbits than in adult rabbits for G1.1aP-GI.2 (SCNJ-2021) variant (Table 1), and all rabbits succumbed to the virus infection experienced subacute or acute disease course. The survival time for rabbits challenged with GI.2 (SCMS-2020) or GI.1aP-GI.2 (SCNJ-2021) in both unvaccinated and vaccinated groups were not significantly different (Supplementary Table S5).

4. Discussion

4.1. Nucleotide sequencing provides insight for newly emerging GI.1aP-GI.2 variants

Since the first notice of GI.2 strains in 2010 (Le Gall-Reculé et al., 2011), the ongoing recombination events among GI.2 and other genotypes of *Lagoviruses* generated several variants that emerged as prevalent strains with wide distribution in the world. For a long time, GI.1a once had been the predominant variant in China until the GI.2 outbreak in 2020. When we investigated the presence of GI.1 and GI.2 within the scope of routine disease monitoring due to the typical symptoms and necropsy findings in dead rabbits, new GI.2 variants from suspect liver specimens were identified and further investigations were conducted to determine the genetic diversity of GI.2 genotypes. The RdRp-VP60 junction is considered as a robust recombination hotspot, therefore, nucleotide sequencing for this region is a rapid detection approach for recombination analysis. The highly intergenotypic recombination frequency (40.9%) of GI.1a and GI.2 indicates that these novel variants are likely to be the predominant strains in the following years. Although this study may underestimate the current incidence of these predominant variants in the rabbit industry in China due to the lack of submission data. The results strongly support the rapid emergence of the new epidemic variants (GI.1aP-GI.2) since the outbreak of GI.2 in China. As far as we know, this recombination pattern has never been described before.

4.2. Genetic recombination inferred evolutionary adaption of GI.1aP-GI.2 variants

The NSP encoding genes of *Lagoviruses* determine the potential of virus replication and immune evasion (Urakova et al., 2015; Zhu et al., 2022), which is a complex process promoting virus evolution with frequent nucleotide variation (Silvério et al., 2018; Mahar et al., 2021). In this study, the recombination analysis reveals that the GI.1aP-GI.2 variants branches into a single clade, which possess similar molecular origins of the parental GI.1a and GI.2, indicating that this new recombination event may occur in China's RHDV strains after the outbreak of GI.2 in 2020 (Chen et al., 2022). Furthermore, the high nucleotide within these four variants also demonstrates that these novel recombinant strains have close geographical relationship. Meanwhile, the position of recombination breakpoints in these four variants is flexible, which shapes the intragenotypic diversity of GI.2 recombination pattern under the evolutionary selection. Nevertheless, the mechanism of the recombination between GI.1a and GI.2 is not fully elucidated.

It has been widely reported that exposure to prototype GI.1 and GI.2 infection in rabbits might promote robust recombination between these two genotypes in a large-scale temporal and spatial context (Silvério et al., 2018; Abrantes et al., 2020a; Al-Ebshahy et al., 2022). Similar results in rabbits and hares co-infected with RHDV2 (GI.2) and the European brown hare syndrome virus (EBHSV GII.1) were also confirmed (Le Gall-Reculé et al., 2017). Hence, intergenotypic recombination between GI.1a and GI.2 may indicate a new exaptation of GI.2 counterparts in response to co-evolutionary interaction between host and virus (Schwensow et al., 2014). This predominant recombination pattern in our findings reveals that GI.1aP-GI.2 variants possibly adapt certain population expansion strategies, which may be attributed to RdRp speed and fidelity, to gain their evolutionary advantage and persistence among GI.2 variants (Mahar et al., 2021). Additionally, the possibly antibody-mediated selection pressure to these two viruses may also favor GI.1aP-GI.2 variants to replace the parental GI.2 and to acquire strengthening adaptability (Hall et al., 2021; O'Connor et al., 2022; Patel et al., 2022). However, lack of direct evidence that host immunity contributes to GI.1aP-GI.2 recombination warrants further investigation.

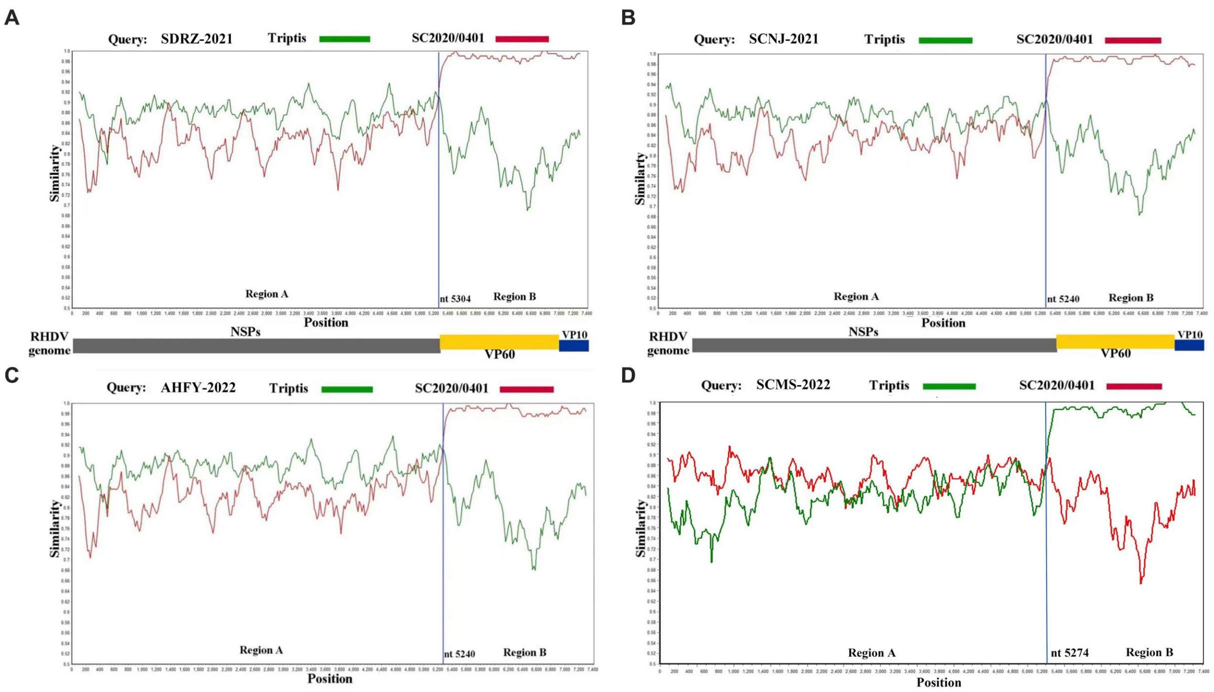


FIGURE 3
Recombination analysis of SDRZ-2021 **(A)**, SCNJ-2021 **(B)**, AHFY-2022 **(C)**, and SCMS-2022 **(D)** strains. Analysis was conducted using a sliding window of 200-bp window and a 20-bp step. The y-axis indicates the percentage similarity between the query sequence and the reference sequences. **(A)** Genome scale similarity comparisons of SDRZ-2021 (query) with Triptis (green) and SC2020/0401 (red); **(B)** genome scale similarity comparisons of SCNJ-2021 (query) with Triptis (green) and SC2020/0401 (red); **(C)** genome scale similarity comparisons of AHFY-2022 (query) with Triptis (green) and SC2020/0401 (red); **(D)** genome scale similarity comparisons of SCMS-2022 (query) with Triptis (green) and SC2020/0401 (red). The recombination breakpoints are marked at the bottom with nucleotide sites and viral genome structure referenced to SC2020/0401.

TABLE 1 Challenge results using prototype GI.2 and GI.1aP-GI.2 strain.

Challenge virus	Infection dose	Unvaccinated group[a]		Vaccinated group[a,b]	
		Juvenile rabbits	Adult rabbits	Juvenile rabbits	Adult rabbits
		Survival/Total (Survival rate)	Survival/Total (Survival rate)	Survival/Total (Survival rate)	Survival/Total (Survival rate)
SCMS-2020 (GI.2)	10,000 LD_{50}	0/6 (0%)	0/6 (0%)	6/6 (100%)	6/6 (100%)
SCNJ-2021 (GI.1aP-GI.2)	10,000 LD_{50}	0/6 (0%)	0/6 (0%)	4/6 (66.6%)	5/6 (83.3%)
Negative control	—	3/3 (100%)	3/3 (100%)	3/3 (100%)	3/3 (100%)

[a] GI.1/GI.2 antigens and antibodies are negative at the beginning of the trial.
[b] Vaccinated with bivalent vaccines consisting of inactivated GI.1and GI.2 antigen.

4.3. Novel recombination might contribute to altering the virulence of GI.1aP-GI.2 variants

The impact of genetic recombination on the virulence of GI.2 variants has been investigated in previous studies (Calvete et al., 2018; Müller et al., 2021). The occurrences of broad-spectrum cross-protection among GI.1bP-GI.2, GI.4eP-GI.2, and GI.4cP-GI.2 variants, irrespectively of recombination patterns and challenge dosages, were confirmed (Calvete et al., 2018; Müller et al., 2021; O'Connor et al., 2022). Nevertheless in our study, vaccination with parental GI.2 confers incomplete cross-protection against challenge by GI.1aP-GI.2 variants in vaccinated rabbits. This suggests that the new variant can escape the host's adaptive immunity yielded by parental GI.2 and evolve to disequilibrate the host-virus interaction (Mahar et al., 2016; Lopes et al., 2018).

Although there is no statistical difference among the survival time for rabbits in each group, the increased mortality and lower LD_{50} value for the GI.1aP-GI.2 variant still support our hypothesis that the distinctive recombination mechanisms may favor GI.1aP-GI.2 variant to acquire moderately enhanced pathogenicity. Although a large amount of nucleotide and amino acid variation emerged along the genome of recombinants, no consistent substitution sites within the non-structural or structural genes were clearly confirmed to be associated with the virulence alteration for the recombinant GI.2 variants (data not shown), which also warrants further biological study and bioinformatic research to elucidate these genetic variations. According to the biocontrol management of wild *Lagomorphs* in several countries, early exposure to non-pathogenic RCV or pathogenic GI.1 conferred partial protection to GI.2 challenge (Patel et al., 2022; Taggart et al., 2022). However, rapid emergence of recombinant variants among co-circulation strains made the wild *Lagomorphs* biocontrol program less successful than before, which also implys that undesirable genetic recombination may disable the vaccine-induced protection conferred by parent RHDV strains. As well known, VP60 protein is considered as a major antigen and virulence determinant of GI.2 (Miao et al., 2019; Liu et al., 2022). Interestingly, in a survey study, the emergence of GI.2 VP60-based recombinants were more likely to be predominant strains among circulating intergenotypic variants (Mahar et al., 2021), which suggests the competitive advantage of the GI.2 recombinants over other genotypes due to frequent amino acid substitions along the NSPs, but not their VP60 protein. In addition, the molecular disparity between GI.1aP-GI.2 and other GI.2 recombinants also implies that the NSPs may not only be relevant to the virus fitness (Mahar et al., 2021), but also be associated with the virulence alteration to a certain extent.

These findings underpin the importance of genetic variability for the rapid spread of the GI.1aP-GI.2 strains in the rabbitries in China under environmental pressure and also implicate the potential of these variants to manipulate the host's immunity.

5. Conclusion

To the best of our knowledge, the new recombinant GI.1aP-GI.2 was identified in domestic rabbits in China for the first time, which expanded the knowledge on the phylodynamics and genomic diversity of GI.2 genotypes. The rapid molecular evolution and varied pathogenicity of these virus recombinants highlight the urgent need for epidemiological surveillance and for future prevention of neglected GI.2 variants.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Ethics statement

The animal study was reviewed and approved by the Committee on the Ethics of Animal Experimental Center of Huapai Biological Group (permit number: 2022HPAES018).

Author contributions

JH and YL study design. DD, YL, LS, CY, and HZ experiment implement. LZ, JY, and JH data analysis and manuscript revision. YL and JH manuscript writing. All authors have approved the manuscript.

Funding

This research is supported by the Southwest Minzu University Research Startup Funds (Grant No. RQD2021098), Natural Science Foundation of Sichuan Province (Grant No. 2022NSFSC0081), and the Public Welfare Scientific Research Institutes Basic Research Projects (Grant No. SASA202302).

Acknowledgments

The authors acknowledge Huapai Bio-engineering Group Co., Ltd (Chengdu, China) for providing the rabbits for this study, and acknowledge Zhang, R.H. (College of Animal Science and Veterinary Medicine, Southwest Minzu University, Chengdu City, China.) for language advice.

Conflict of interest

DD was employed by Huapai Bio-engineering Group Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1188380/full#supplementary-material

References

Abrantes, J., Droillard, C., Lopes, A. M., Lemaitre, E., Lucas, P., Blanchard, Y., et al. (2020a). Recombination at the emergence of the pathogenic rabbit haemorrhagic disease virus Lagovirus europaeus/GI.2. *Sci. Rep.* 10:14502. doi: 10.1038/s41598-020-71303-4

Abrantes, J., Lopes, A. M., Lemaitre, E., Ahola, H., Banihashem, F., Clément Droillard, C., et al. (2020b). Retrospective analysis shows that most RHDV GI.1 strains circulating since the late 1990s in France and Sweden were recombinant GI.3P-GI.1d strains. *Genes (Basel)* 11:910. doi: 10.3390/genes11080910

Aguayo-Adán, J. A., Rouco, C., Delibes-Mateos, M., and Santoro, S. (2022). Lack of evidence for differences in the spread of classic (Lagovirus europaeus/GI.1) and novel (Lagovirus europaeus/GI.2) rabbit haemorrhagic disease viruses in Europe and North Africa. *Vet. Rec.* 190:e1067. doi: 10.1002/vetr.1067

Al-Ebshahy, E., Abas, O., and Abo-ElKhair, M. (2022). Co-circulation of GI.1 and GI.2 genotypes of rabbit hemorrhagic disease virus in Egypt. *Virus Dis.* 33, 422–428. doi: 10.1007/s13337-022-00791-x

Calvete, C., Mendoza, M., Alcaraz, A., Sarto, M. P., Jiménez-de-Bagüéss, M. P., Calvo, A. J., et al. (2018). Rabbit haemorrhagic disease: cross-protection and comparative pathogenicity of GI.2/RHDV2/b and GI.1b/RHDV lagoviruses in a challenge trial. *Vet. Microbiol.* 219, 87–95. doi: 10.1016/j.vetmic.2018.04.018

Chehida, F. B., Lopes, A. M., Côrte-Real, J. V., Sghaier, S., Aouini, R., Messadi, L., et al. (2021). Multiple introductions of rabbit hemorrhagic disease virus Lagovirus europaeus/GI.2 in Africa. *Biology* 10:883. doi: 10.3390/biology10090883

Chen, W., Tu, T., Luo, Y., Yang, Z. X., Yao, X. P., Wu, X. L., et al. (2022). Detection of a new emerging strain of rabbit haemorrhagic disease virus 2 (GI.2) in China. *J. Vet. Res.* 66, 289–295. doi: 10.2478/jvetres-2022-0048

Dalton, K. P., Nicieza, I., Balseiro, A., Muguerza, M. A., Rosell, J. M., Casais, R., et al. (2012). Variant rabbit hemorrhagic disease virus in young rabbits. *Spain. Emerg. Infect. Dis.* 18:2009. doi: 10.3201/eid1812.120341

Hall, R. B., King, T. G., O'Connor, T. W., Read, A. J., Vrankovic, S., Piper, M., et al. (2021). Passive immunisation against RHDV2 induces protection against disease but not infection. *Vaccines* 9:1197. doi: 10.3390/vaccines9101197

Hu, B., Fan, Z. Y., Wang, F., Song, Y. H., Wei, H. J., Liu, X., et al. (2016). A new variant of rabbit hemorrhagic disease virus G2-like strain isolated in China. *Virus Res.* 215, 20–24. doi: 10.1016/j.virusres.2016.01.018

Hu, Z. J., Tian, X. J., Zhai, Y. J., Xu, W., Zheng, D., and Sun, F. (2010). Cryo-electron microscopy reconstructions of two types of wild rabbit hemorrhagic disease viruses characterized the structural features of Lagovirus. *Protein Cell* 1, 48–58. doi: 10.1007/s13238-010-0007-0

Hu, B., Wang, F., Fan, Z. Y., Song, Y. H., Abrantes, J., Zuo, Y. Y., et al. (2017). Recombination between G2 and G6 strains of rabbit hemorrhagic disease virus (RHDV) in China. *Arch. Virol.* 162, 269–272. doi: 10.1007/s00705-016-3082-6

Hu, B., Wei, H. J., Fan, Z. Y., Song, Y. H., Chen, M. M., Qiu, R. L., et al. (2021). Emergence of rabbit haemorrhagic disease virus 2 in China in 2020. *Vet. Med. Sci.* 7, 236–239. doi: 10.1002/vms3.332

Le Gall-Reculé, G., Lemaitre, E., Bertagnoli, S., Hubert, C., Top, S., Decors, A., et al. (2017). Large-scale lagovirus disease outbreaks in European brown hares (*Lepus europaeus*) in France caused by RHDV2 strains spatially shared with rabbits (*Oryctolagus cuniculus*). *Vet. Res.* 48:70. doi: 10.1186/s13567-017-0473-y

Le Gall-Reculé, G., Zwingelstein, F., Boucher, S., Le Normand, B., Plassiart, G., Portejoie, Y., et al. (2011). Detection of a new variant of rabbit haemorrhagic disease virus in France. *Vet. Rec.* 168, 137–138. doi: 10.1136/vr.d697

Le Pendu, J., Abrantes, J., Bertagnoli, S., Guitton, J. S., Le Gall-Reculé, G., Lopes, A. M., et al. (2017). Proposal for a unified classification system and nomenclature of lagoviruses. *J. Gen. Virol.* 98, 1658–1666. doi: 10.1099/jgv.0.000840

Liu, C. J., Lin, M., Hu, H. Y., Liu, X. L., Bian, Y. C., and Deng, S. Z. (2022). Rabbit hemorrhagic disease virus VP60 protein expressed in recombinant swinepox virus self-assembles into virus-like particles with strong immunogenicity in rabbits. *Front. Microbiol.* 13:960374. doi: 10.3389/ fmicb. 2022. 960 374

Liu, S. J., Xue, H. P., Pu, B. Q., and Qian, N. H. (1984). A new viral disease in rabbit (in Chinese). *Anim. Husb. Vet. Med.* 16, 253–255.

Lopes, A. M., Breiman, A., Lora, M., Moullac-Vaidye, B. L., Galanina, O., and Nyström, K. (2018). Host-specific glycans are correlated with susceptibility to infection by Lagoviruses, but not with their virulence. *J. Virol.* 92, e01759–e01717. doi: 10.1128/JVI.01759-17

Lopes, A. M., Silvério, D., Magalhães, M. J., Areal, H., Alves, P. C., Esteves, P. J., et al. (2017). Characterization of old RHDV strains by complete genome sequencing identifies a novel genetic group. *Sci. Rep.* 7:13599. doi: 10.1038/s41598-017-13902-2

Mahar, J. E., Hall, R. N., Peacock, D., Kovaliski, J., Piper, M., Mourant, R., et al. (2018). Rabbit hemorrhagic disease virus 2 (RHDV2; GI.2). Is replacing endemic strains of RHDV in the Australian landscape within 18 months of its arrival. *J. Virol.* 92, e01374–e01317. doi: 10.1128/JVI.01374-17

Mahar, J. E., Jencke, M., Huang, N., Smertina, E., Holmes, E. C., Strive, T., et al. (2021). Frequent intergenotypic recombination between the non-structural and structural genes is a major driver of epidemiological fitness in caliciviruses. *Virus. Evol.* 7:veab080. doi: 10.1093/ve/veab080

Mahar, J. E., Nicholson, L., Eden, J. S., Duchêne, S., Kerr, P. J., Duckworth, J., et al. (2016). Benign rabbit caliciviruses exhibit evolutionary dynamics similar to those of their virulent relatives. *J. Virol.* 90, 9317–9329. doi: 10.1128/JVI.01212-16

Miao, Q., Qi, R., Veldkamp, L., Ijzer, J., Kik, M. L., Zhu, J., et al. (2019). Immunogenicity in rabbits of virus-like particles from a contemporary rabbit haemorrhagic disease virus type 2 (GI.2/RHDV2/b) isolated in the Netherlands. *Viruses* 11:553. doi: 10.3390/v11060553

Mizoguchi, T., Ken-ichi, I., and Sakurai, M. (2003). Hemagglutination and antigenic comparison of rabbit hemorrhagic disease virus. *J. Vet. Med. Sci.* 65, 95–97. doi: 10.1292/jvms.65.95

Müller, C., Hrynkiewicz, R., Bębnowska, D., Maldonado, J., Massimiliano Baratelli, M., Köllner, B., et al. (2021). Immunity against Lagovirus europaeus and the impact of the immunological studies on vaccination. *Vaccines* 9:255. doi: 10.3390/vaccines9030255

Neimanis, A. S., Ahola, H., Zohari, S., Pettersson, U. L., Bröjer, C., Capucci, L., et al. (2018b). Arrival of rabbit haemorrhagic disease virus 2 to northern Europe: emergence and outbreaks in wild and domestic rabbits (*Oryctolagus cuniculus*) in Sweden. *Transbound. Emerg. Dis.* 65, 213–220. doi: 10.1111/tbed.12650

Neimanis, A. S., Pettersson, U. L., Huang, N., Gavier-Widén, D., and Strive, T. (2018a). Elucidation of the pathology and tissue distribution of Lagovirus europaeus GI.2/RHDV2 (rabbit haemorrhagic disease virus 2) in young and adult rabbits (*Oryctolagus cuniculus*). *Vet. Res.* 49:46. doi: 10.1186/s13567-018-0540-z

O'Connor, T. F., Read, A. J., Hall, R. N., Strive, T., and Kirkland, P. D. (2022). Immunological cross-protection between different rabbit hemorrhagic disease viruses-implications for rabbit biocontrol and vaccine development. *Vaccines* 10:666. doi: 10.3390/vaccines10050666

Pacioni, C., Hall, R. N., Strive, T., Ramsey, D. S. L., Gill, M. S., and Vaughan, T. G. (2022). Comparative epidemiology of rabbit haemorrhagic disease virus strains from viral sequence data. *Viruses* 15:21. doi: 10.3390/v15010021

Patel, K. P., Strive, T., Hall, R. N., Mutze, G., Page, B., Korcz, M., et al. (2022). Cross-protection, infection and case fatality rates in wild European rabbits experimentally challenged with different rabbit haemorrhagic disease viruses. *Transbound. Emerg. Dis.* 69, e1959–e1971. doi: 10.1111/tbed.14530

Qi, R. B., Meng, C. C., Zhu, J., Li, H., Miao, Q. H., Tang, J. Y., et al. (2022). The outbreak of rabbit hemorrhagic virus type 2 in the interior of China may be related to imported semen. *Virol. Sin.* 37, 623–626. doi: 10.1016/j.virs.2022.04.003

Schwensow, N. I., Cooke, B., Kovaliski, J., Sinclair, R., Peacock, D., Fickel, J., et al. (2014). Rabbit haemorrhagic disease: virus persistence and adaptation in Australia. *Evol. Appl.* 7, 1056–1067. doi: 10.1111/eva.12195

Silvério, D., Lopes, A. M., Melo-Ferreira, J., Magalhães, M. J., Monterroso, P., Serronha, A., et al. (2018). Insights into the evolution of the new variant rabbit haemorrhagic disease virus (GI.2) and the identification of novel recombinant strains. *Transbound. Emerg. Dis.* 65, 983–992. doi: 10.1111/tbed.12830

Smertina, E., Hall, R. N., Urakova, N., Strive, T., and Frese, M. (2021). Calicivirus non-structural proteins: potential functions in replication and host cell manipulation. *Front. Microbiol.* 12:2021. doi: 10.3389/fmicb.2021.712710.eCollection

Song, Y. H., Fan, Z. Y., Zuo, Y. Y., Wei, H. J., Hu, B., Chen, M. M., et al. (2017). Binding of rabbit hemorrhagic disease virus-like particles to host histo-blood group antigens is blocked by antisera from experimentally vaccinated rabbits. *Arch. Virol.* 162, 3425–3430. doi: 10.1007/s00705-017-3509-8

Taggart, P. L., Hall, R. N., Cox, T. E., Kovaliski, J., McLeod, S. R., and Strive, T. (2022). Changes in virus transmission dynamics following the emergence of RHDV2 shed light on its competitive advantage over previously circulating variants. *Transbound. Emerg. Dis.* 69, 1118–1130. doi: 10.1111/tbed.14071

Teifke, J. P., Reimann, I., and Schirrmeier, H. (2002). Subacute liver necrosis after experimental infection with rabbit haemorrhagic disease virus (RHDV). *J. Comp. Pathol.* 126, 231–234. doi: 10.1053/jcpa.2001.0534

Urakova, N., Frese, M., Hall, R. N., Liu, J., Matthaei, M., and Strive, T. (2015). Expression and partial characterisation of rabbit haemorrhagic disease virus non-structural proteins. *Virology* 484, 69–79. doi: 10.1016/j.virol.2015.05.004

Wang, X. L., Hao, H. F., Qiu, L., Dang, R. Y., Du, E. Q., Zhang, S. X., et al. (2012). Phylogenetic analysis of rabbit hemorrhagic disease virus in China and the antigenic variation of new strains. *Arch. Virol.* 157, 1523–1530. doi: 10.1007/s00705-012-1340-9

Zhou, J., Ma, Y. J., Wang, M., Zhang, Y., Chen, B., Chen, D. S., et al. (2022). Establishment of a duplex TaqMan RT-PCR for the differential detection of RHDV GI.1 and GI.2. *J Virol. Methods* 304:114526. doi: 10.1016/j.jviromet.2022.114526

Zhu, J., Miao, Q. H., Guo, H. Y., Tang, A. X., Dong, D. D., Tang, J. Y., et al. (2022). Nucleolin interacts with the rabbit hemorrhagic disease virus replicase RdRp, nonstructural proteins p16 and p23, playing a role in virus replication. *Virol. Sin.* 37, 48–59. doi: 10.1016/j.virs.2022.01.004

frontiers | Frontiers in Microbiology

TYPE Review
PUBLISHED 25 July 2023
DOI 10.3389/fmicb.2023.1188249

OPEN ACCESS

EDITED BY
Naveen Kumar,
ICAR-National Institute of High Security Animal Diseases (ICAR-NIHSAD), India

REVIEWED BY
Christopher Gerner,
University of Vienna, Austria
Kavyashree Manjunath,
Institute for Stem Cell Science and Regenerative Medicine (inStem), India

*CORRESPONDENCE
Xuhua Xia
✉ xxia@uottawa.ca

RECEIVED 17 March 2023
ACCEPTED 10 July 2023
PUBLISHED 25 July 2023

CITATION
Xia X (2023) Identification of host receptors for viral entry and beyond: a perspective from the spike of SARS-CoV-2.
Front. Microbiol. 14:1188249.
doi: 10.3389/fmicb.2023.1188249

Identification of host receptors for viral entry and beyond: a perspective from the spike of SARS-CoV-2

Xuhua Xia[1,2]*

[1]Department of Biology, University of Ottawa, Ottawa, ON, Canada, [2]Ottawa Institute of Systems Biology, University of Ottawa, Ottawa, ON, Canada

Identification of the interaction between the host membrane receptor and viral receptor-binding domain (RBD) represents a crucial step for understanding viral pathophysiology and for developing drugs against pathogenic viruses. While all membrane receptors and carbohydrate chains could potentially be used as receptors for viruses, prioritized searches focus typically on membrane receptors that are known to have been used by the relatives of the pathogenic virus, e.g., ACE2 used as a receptor for SARS-CoV is a prioritized candidate receptor for SARS-CoV-2. An ideal receptor protein from a viral perspective is one that is highly expressed in epithelial cell surface of mammalian respiratory or digestive tracts, strongly conserved in evolution so many mammalian species can serve as potential hosts, and functionally important so that its expression cannot be readily downregulated by the host in response to the infection. Experimental confirmation of host receptors includes (1) infection studies with cell cultures/tissues/organs with or without candidate receptor expression, (2) experimental determination of protein structure of the complex between the putative viral RDB and the candidate host receptor, and (3) experiments with mutant candidate receptor or homologues of the candidate receptor in other species. Successful identification of the host receptor opens the door for mechanism-based development of candidate drugs and vaccines and facilitates the inference of what other animal species are vulnerable to the viral pathogen. I illustrate these approaches with research on identification of the receptor and co-factors for SARS-CoV-2.

KEYWORDS

receptor-binding protein, host receptor, protein structure, phylogenetics, gene expression, drug target

1. Introduction

Many processes are involved in viral infection, including attachment to host cells, entry of host cells, evasion of host defense mechanisms, viral genome replication, transcription and translation within host cells, viral packaging, lysing host cells and initiating a new infection cycle. Among these processes, attaching to and entering the host cell are often the limiting step requiring the viral pathogen to evolve specific adaptation to the host. Once inside the host cell, the cytoplasmic environment for viral genome replication, transcription and translation are similar across diverse mammalian species.

The need and urgency of identifying the host receptor used by viruses are highlighted by the COVID-19 pandemic. What is the host receptor for the spike protein of SARS-CoV-2 (SARS-2S)? Which part of the SARS-2S binds to which part of the host receptor? What are the amino acid residues that interact between the host receptor and the viral receptor-binding protein? Can drugs be developed to block the binding of SARS-2S to the receptor? Will the drug interfere with the normal function of the receptor and cause a strong side effect? What are the normal functions of the host receptor? Are people who express less of this receptor protein less vulnerable to COVID-19 infection? What other mammalian species have similar host receptors that render them vulnerable to SARS-CoV-2 infection? How well can we predict species vulnerability based on receptor protein similarity in sequence and in structure? Answers to these questions contribute to a good understanding of pathophysiology and epidemiology and provide a foundation for drug development. I illustrate the multi-omics approaches to address these questions related to host receptors.

2. Identification of host receptor and cofactors

Cell membranes are composed of many single-pass transmembrane receptors. Some of them can be internalized into cells upon ligand binding, and could serve as candidate receptors mediating viral attachment and cell entry. However, instead of screening all of them as candidate receptors, existing biological knowledge can speed up the search of host receptors or cofactors that support or enhance viral attachment and cell entry.

2.1. Identification of ACE2 as the host receptor

The first genomic sequence of SARS-CoV-2 was obtained on January 5, 2020, and made public on Jan. 11, 2020 (Wu et al., 2020). Previously, ACE2 was found to be the host receptor of SARS-S (Li et al., 2003; Kuba et al., 2005). The receptor-binding domain of SARS-S alone can bind to ACE2, leading to its internalization together with the host ACE2 (Wang et al., 2008). Because of the similarity in sequence and domain organization between SARS-S and SARS-2S (Zhou P. et al., 2020; Xia, 2021), it is natural to infer that SARS-2S may use the same host receptor ACE2 as SARS-S (Zhou P. et al., 2020).

ACE2 is a typical single-pass transmembrane receptor with a hydrophobic signal peptide of 17 aa and a single hydrophobic transmembrane domain (Figure 1A). ACE2 is a metallopeptidase with a 5-aa HEMGH zinc-binding motif (Figure 1A). Several proteases including TMPRSS2 (transmembrane serine protease 2), ADAM-17 (a disintegrin and metalloprotease 17, also known as TACE) and HAT (Human airway trypsin-like protease) can cleave ACE2 at the segment rich in lysine (K) and arginine (R) close to the transmembrane domain (Figure 1A) to shed enzymatically active soluble ACE2 (sACE2). This KR-rich segment is hydrophilic and consequently disordered, and is missing in the ACE2 structure (1R42, Figure 1B; Towler et al., 2004). What is particularly interesting is that such cleavage of ACE2 by proteases, especially by ADAM-17 (Haga et al., 2008, 2010; Scheller et al., 2011), is activated in SARS-CoV infection. One naturally would think that such cleavage might be a protective response by the host cells, i.e., if membrane-bound ACE2 mediates viral entry, then cleaving them off membrane would decrease infection. Surprisingly, the generation of sACE2 enhances infection (Haga et al., 2008, 2010). This shows the complexity in pathogen-host interactions that I will discuss in more detail later.

Ever since the first characterization of ACE2 (Donoghue et al., 2000; Tipnis et al., 2000), gene expression of ACE2 has been found high in kidney, heart, testis, colon and small intestine, but low in lungs (Hikmet et al., 2020; Li et al., 2020; Figure 1C). This low expression of ACE2 in lungs has motivated the search for alternative receptors and cofactors, until it was found that ACE2 is highly expressed in type II pneumocytes in lungs (Hamming et al., 2004; To et al., 2004; To and Lo, 2004; Mossel et al., 2008; Xu et al., 2020; Zhao et al., 2020), so the low expression in lungs is due to the mixture of these type II pneumocytes with other types of lung cells that express little ACE2. I should add that the "high" expression of ACE2 in type II pneumocytes is relative to other types of lung cells. The ACE2 expression in type II pneumocytes is still negligibly low relative to other cell types such as those in the digestive system (e.g., enterocytes) or connected to the digestive system (e.g., cholangiocytes), or cells in kidney (e.g., proximal tubular epithelial cells) or in testes (e.g., Sertoli cells), according to data in The Human Protein Atlas (Uhlén et al., 2015).

Interestingly, ACE2 was found to be expressed in oral tissues, especially in tongue (Xu et al., 2020), although the expression is generally low compared to that in the digestive tract. The expression of ACE2 in tongue indicates the potential of destruction of tongue cells upon COVID-19 infection. Whether this might be linked to the loss of taste, a common symptom of COVID-19 infection, has not been explored.

Other candidate receptors that have been used by various coronaviruses include amino peptidase N (APN) and dipeptidyl peptidase 4 (DPP4). However, only cells expressing ACE2 are susceptible to SARS-CoV-2 infection. The presence/absence of APN or DPP4 is irrelevant to SARS-CoV-2 infection (Zhou P. et al., 2020). While ACE2 binds to SARS-S and SARS-2S, DPP4 does not (Wang et al., 2020). In particular, ACE2 from mice which is substantially diverged from human ACE2 does not support viral entry (Zhou P. et al., 2020). However, transgenic mice expressing human ACE2 are vulnerable to SARS-CoV-2 and can develop COVID-19 symptoms (Bao et al., 2020), suggesting that ACE2 is a sufficient receptor for SARS-CoV-2 attachment and cell entry.

Further corroboration of the interaction between the viral RBD and the host ACE2 comes from microscopy methods and structural characterization. Microscopy methods such as confocal fluorescence microscopy can visualize the binding of coronavirus spike proteins to GFP-tagged ACE2 (Wang et al., 2020). Structural studies have characterized not only the structure of SARS-2S monomer and trimmer (Hoffmann et al., 2020; Walls et al., 2020; Wrapp et al., 2020; Yan et al., 2020), but also the SARS-2S trimer and the ACE2 in complex (Gui et al., 2017; Wang et al., 2020; Zhou T. et al., 2020; Xu et al., 2021). What remains to be elucidated is the mechanistic aspects of how the ACE2-binding triggers the transformation of the SARS-2S trimer from the prefusion state to the postfusion state.

These structural studies also provide a list of amino acids in physical contact with each other from the two interacting partners (Lu et al., 2015; Adhikari and Ching, 2020; Adhikari et al., 2020). The sharing of the interacting amino acids in ACE2 were subsequently

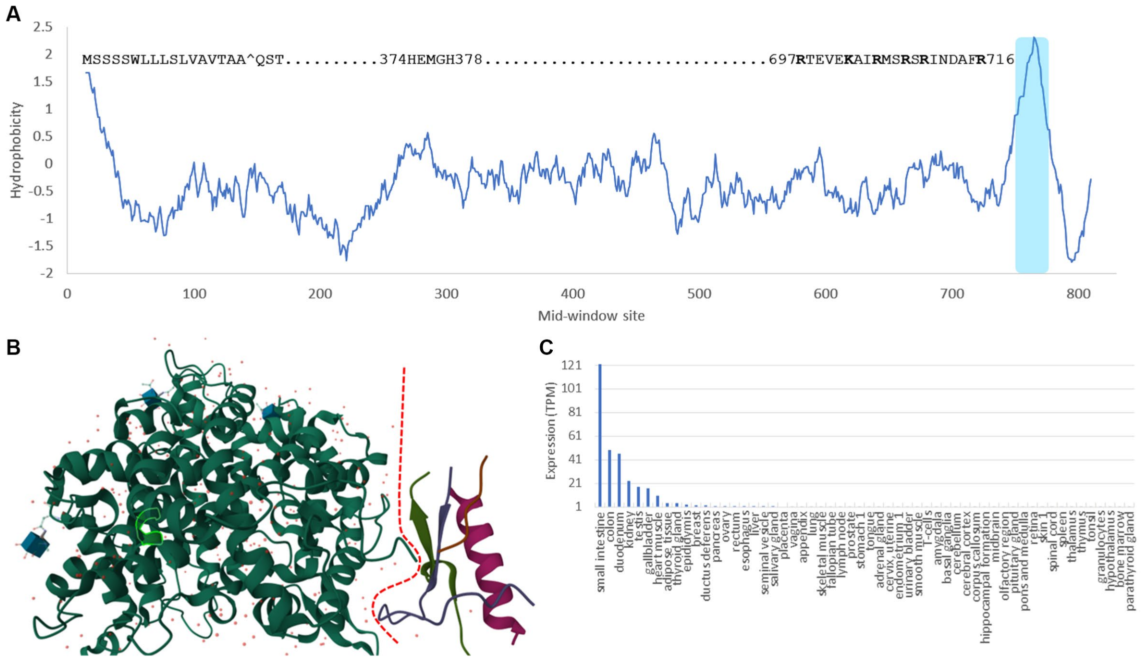


FIGURE 1
Domains, structure and tissue-specific expression of human ACE2. **(A)** Hydrophobicity plot generated from DAMBE (Xia, 2018b) based on hydrophobicity values in Kyte and Doolittle (1982) along a sliding window of 40 amino acids. The 17-aa signal peptide at the N-terminus, the 5-aa zinc-binding motif, the segment rich in lysine and arginine (KR-rich segment) serving as cleavage sites for TMPRSS2 and HAT proteases (Heurich et al., 2014), and the shaded hydrophobic transmembrane domain are indicated. **(B)** Structure of ACE2 (1R42) (Towler et al., 2004) with the dashed red line separating the extracellular domain on the left and the intracellular domain on the right. The KR-rich segment is hydrophilic and therefore disordered. It is missing in the structure, so is the transmembrane domain. The 5-aa zinc-binding motif is highlighted within the green-line enclosure. **(C)** The tissue-specific expression data is extracted from The Human Protein Atlas (Uhlén et al., 2015).

used to predict what other mammalian species have an ACE2 that can serve as a host receptor for SARS-CoV-2 infection (Shi et al., 2020; Kruglikov et al., 2021; Wei et al., 2021), which I discuss later. Such information also facilitates the identification of key residues that contribute to the host tropism of SARS-CoV-2. For example, SARS-CoV-2 cannot infect mice because of differences in five key residues in ACE2 between mouse and human. Replacing these residues created a mouse model susceptible to SARS-CoV-2 infection (Adams et al., 2021).

Protein structures also shed light on interactions between the receptor and the viral spike proteins. For example, the structure (7KNB) of human ACE2 in complex with SARS-2S trimer (Zhou T. et al., 2020) shows four segments in ACE2 (19–39, 323–330, 352–357, 385–390) and two segments in SARS-2S (443–458, 472–506) to be in close physical proximity. The amino acids in the four ACE2 segments jointly have an isoelectric point (pI) of 4.38, and those in the two SARS-2S segments jointly have a pI of 9.40. Thus, at neutral pH, the former is negatively charged, but the latter is positively charged. The two therefore would have favorable electrostatic interactions facilitating their binding to each other. This result makes sense of a previous mutation experiment (Adams et al., 2021) to convert the mouse ACE2 that cannot bind to SARS-2S to one that can, based on the sequence difference between human and mouse ACE2. The two mutation constructs (hmACE2.3 and hmACE2.4) that introduced negatively charged amino acid residues present in human ACE2 into mouse ACE2 (N30D in hmACE2.3, and A329E in hmACE2.4) can functionally interact with SARS-2S just as well as human ACE2. The other two constructs (hmACE2.1 and hmACE2.2) introduced mutations to increase hydrophilicity (e.g., H353K in hmACE2.1 and N31K in hmACE2.2), which also improved the interaction of mouse ACE2 with SARS-2S, albeit to a smaller degree than hmACE2.3 and hmACE2.4.

It is important to keep in mind the difference between SARS-CoV and SARS-CoV-2 in their use of ACE2 for cell entry (Xia, 2021). First, there are two documented alternative pathways of cell entry for coronaviruses after receptor binding: (1) cell entry by membrane fusion when the spike trimer is cleaved at the polybasic furin site, and (2) cell entry by clathrin-mediated endocytosis (Inoue et al., 2007) and the endosome-cathepsin pathway (Matsuyama et al., 2005, 2010). SARS-CoV-2, with the cleaved furin site, uses mainly pathway 1, whereas SARS-CoV uses pathway 2. The inhibition of the clathrin-mediated endocytosis dramatically reduces cell entry by SARS-CoV (Inoue et al., 2007). SARS-CoV cannot use pathway 1 because of the lack of the polybasic furin site. However, a polybasic furin site experimentally introduced into SARS-CoV at the same location as in SARS-CoV-2 created a much more infectious SARS-CoV (Belouzard et al., 2009) with syncytium formation characteristic of SARS-CoV-2 infection. Similarly, SARS-CoV-2 lacking the polybasic furin site are less infective with little syncytium formation (Peacock et al., 2021). Second, SARS-CoV-2 infection is frequently associated with syncytia formation (Daly et al., 2020; Hoffmann et al., 2020; Li et al., 2021) which is rarely reported with SARS-CoV infection. This syncytia formation implies that, once SARS-CoV-2 has entered a cell, it can infect neighboring cells without using ACE2. Thus, high ACE2 abundance in young people than old people (Plaas et al., 2021; Bastolla et al., 2022) renders them more susceptible to SARS-CoV than old people. However, the reduced dependence of SARS-CoV-2 on membrane ACE2 allows SARS-CoV-2 to infect old people who are immunologically weak (Montecino-Rodriguez et al., 2013), even though they do not express a high level of ACE2.

2.2. Other candidate receptors and cofactors

The identification of ACE2 as the host receptor does not imply that it is the only host receptor. Several viruses are known to use multiple receptors and co-factors. For example, Dengue virus uses both human mannose-binding receptor (MR) and DC-SIGN on macrophages as primary receptors (Lo et al., 2016), and HIV-1 uses both CD4 as a primary receptor and a cellular coreceptor (Wilen et al., 2012). Are there other receptors or cofactors that facilitate SARS-CoV-2 attachment and cell entry? Existing evidence points to a hypothesis that has not yet been fully explored. SARS-CoV-2 can bind to both membrane-bound ACE2 or soluble sACE2 which could then bind to membrane proteins such as neuropilin-1 (Cantuti-Castelvetri et al., 2020; Daly et al., 2020), integrins (Nader et al., 2021; Nader and Kerrigan, 2022), or other membrane proteins to anchor SARS-CoV-2 to host cell membrane.

If SARS-S and SARS-2S use ACE2 as the only host receptor to mediate attachment and cell entry, then SARS-CoV and SARS-CoV-2 should infect the same tissue. However, the two viral lineages differ in their tissue tropism, with SARS-CoV more likely infecting lower respiratory tract than SARS-CoV-2. Two mutations in SARS-2S relative to SARS-S have been hypothesized to contribute to differences in cell tropism between SARS-CoV and SARS-CoV-2, and to involve alternative receptors or cofactors.

2.2.1. Polybasic furin site and NRP1 (neuropilin-1)

One conspicuous difference between SARS-S and SARS-2S is the presence of a polybasic furin site, RRAR^S, in the latter but not in the former (Andersen et al., 2020; Coutard et al., 2020; Hoffmann et al., 2020). Two lines of evidence suggests that this furin site is responsible for the difference in tissue tropism between SARS-CoV and SARS-CoV-2. First, a previous study demonstrated that inserting a polybasic furin site at the S1 and S2 boundary in SARS-S resulted in dramatic changes in cell tropism (Belouzard et al., 2009). Second, the spike protein trimer in SARS-CoV-2 virions is already cleaved at this furin site to prime the fusion between the viral and host membranes (Hoffmann et al., 2020; Xia, 2021), and the C-terminus of the cleaved S1 is accessible for interacting with other proteins (Walls et al., 2020; Wrapp et al., 2020). One may therefore infer that host membrane proteins with structural affinity to the cleaved end of SARS-CoV-2 could serve either as an alternative host receptor or an enhancer of viral infection.

NRP1 is a single-pass membrane protein which is obvious from a hydrophobicity plot (Figure 2) as it has just a single hydrophobic transmembrane domain. The hydrophobic stretch of 21 amino acids at the N-terminus is the signal peptide (Figure 2). The a1 and a2 domain are also known as the CUB domains. The b1-b2 domains bind to the furin-cleaved C-terminus of the S1 subunit of SARS-CoV-2 (Daly et al., 2020). The 23-aa segment near the C-terminus of NRP1 is the single-pass transmembrane domain that divides the NRP1 protein into the long extracellular domain and the short 43-aa cytoplasmic domain.

NRP1 is a receptor for other glycoproteins such as VEGF-A and SEMA3A (Plein et al., 2014). Its b1-b2 domain binds specifically to furin-cleaved substrates that has an R/KXXR/K motif at the C terminus where X is any amino acid (Teesalu et al., 2009; Plein et al., 2014). Experimentally determined NRP1 structure shows the negatively charged D320 in NRP-1 interacting electrostatically with the positively charged R/K residue at the C-terminus of the ligand (Guo and Vander Kooi, 2015), and a ligand with the terminal R/K removed may serve as an NRP1 inhibitor. The furin-cleaved S1 subunit of SARS-2S, with the C-terminal RRAR conforming to the R/KXXR/K motif, binds directly to NRP1 (Daly et al., 2020). The following three experimental studies demonstrated NRP1 to be a cofactor that enhances ACE2-mediated viral attachment and cell entry, although it does not serve a sufficient host receptor for SARS-CoV-2 independent of ACE2 (Cantuti-Castelvetri et al., 2020; Daly et al., 2020). First, blocking the binding between the b1-b2 domain and the C-terminus of the viral S1 subunit significantly reduces viral internalization (Cantuti-Castelvetri et al., 2020). Second, removing the RRAR at the C-terminus of the S1 subunit of SARS-2S decreases the binding of S1 to NRP1, and knocking out NRP1 decreases SARS-CoV-2 infection of Hela cells expressing ACE2 (Daly et al., 2020). Third, x-ray crystallography and biochemical approaches revealed that NRP1 enhances internalization of SARS-CoV-2 and syncytia formation (Daly et al., 2020) that has been observed previously to enhance SARS-CoV-2 propagation from cell to cell (Hoffmann et al., 2020).

NRP1 is highly expressed in the olfactory epithelium, and the SARS-CoV-2 infection appears to be correlated with NRP1 expression (Cantuti-Castelvetri et al., 2020). This could explain why SARS-CoV-2 infects predominantly the upper respiratory tracts, in contrast to SARS-CoV that infects lower respiratory tracts and lungs. However, NRP1 is also abundantly expressed in almost all pulmonary tissues (Cantuti-Castelvetri et al., 2020) including lungs (Figure 3), so more studies are needed to check if patients with SARS-CoV-2 infection of lungs also have higher expression of NRP1 in lungs than those without lung infections.

2.2.2. The K403R mutation in SARS-2S, the resulting RGD motif and integrins

Another mutation in SARS-2S relative to SARS-S is K403R (Figure 4) which creates an RGD motif known to be a general integrin-binding motif (Takada et al., 2007). This motif is shared between SARS-CoV-2 and its close relatives isolated from pangolins. The homologous motif in SARS-CoV is KGD (Figure 4). Thus, both SARS-2S and the spike protein from pangolin-isolated SARSr (where r stands for coronaviruses closely related to SARS) are expected to bind to integrins, especially the major endothelial cell integrin, αVβ3 (Nader et al., 2021; Nader and Kerrigan, 2022).

Integrins are membrane receptors existing as αβ heterodimers (Figure 5). Like NRP1, both α and β subunit of integrins are single-pass membrane proteins with a single transmembrane domain, illustrated with αV and β3 subunits (Figure 5). Human genomes encode at least 18 α subunits and eight β subunit. The RGD-recognizing integrins include α5β1, αVβ1, αVβ3, αVβ5,

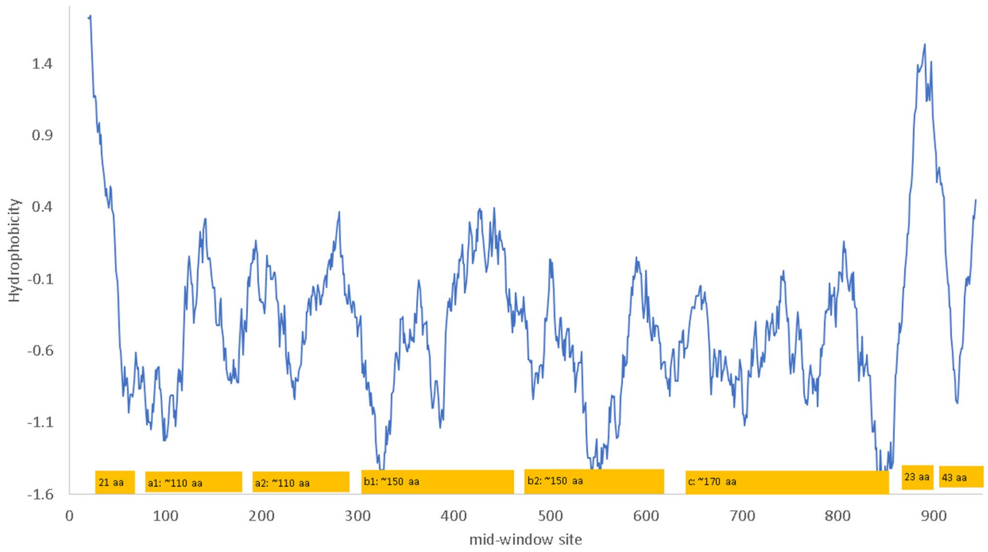


FIGURE 2
Hydrophobicity plot and domain structure of human neuropilin-1 (NRP1) along a sliding window of 40 amino acids (aa). The b1-b2 domains bind to the furin-cleaved C-terminus of the S1 subunit of SARS-CoV-2. The signal peptide includes the first 21 aa. The 23 aa near the C-terminus of NRP1 constitute the single-pass transmembrane domain that divides the NRP1 protein into the long extracellular domain and the short 43-aa cytoplasmic domain. The hydrophobicity plot was generated from DAMBE (Xia, 2018b) based on hydrophobicity values in Kyte and Doolittle (1982). The domains are not drawn exactly to scale. The numbering of amino acids on the horizontal axis follows the neuropilin-1 isoform X1 annotated on human chromosome 10 (NC_000010).

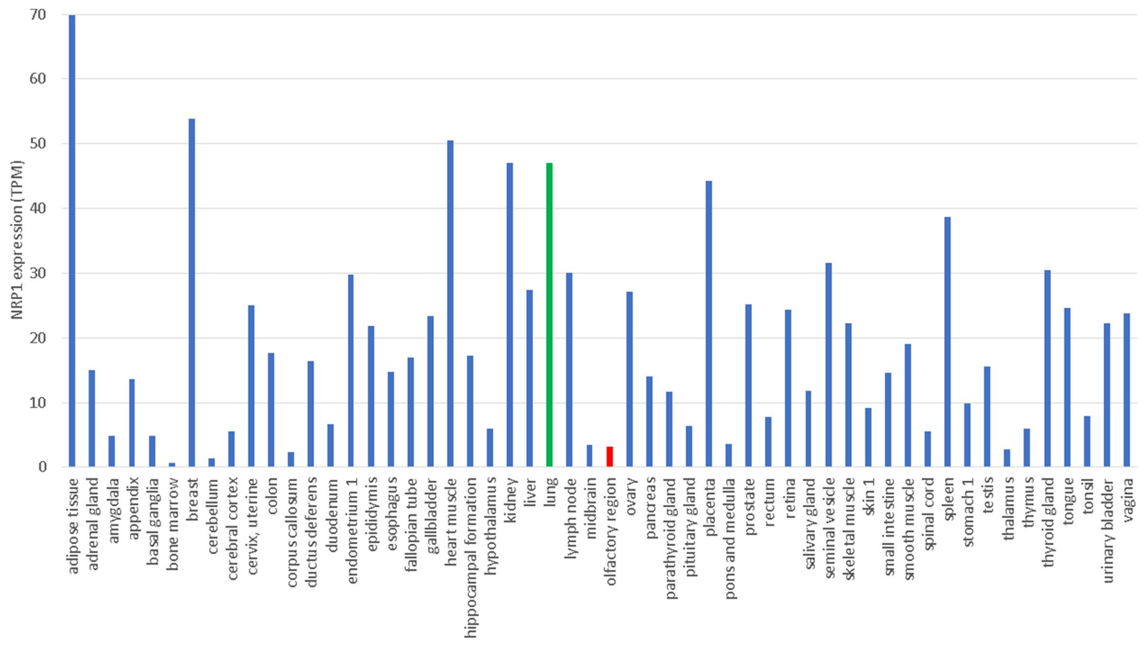


FIGURE 3
Tissue-specific expression of *NRP1* from The Human Protein Atlas (Uhlén et al., 2015), in unit of TPM (transcripts per million). Gene expression in lung and olfactory regions is colored green and red, respectively.

αVβ6, αVβ8, and αIIbβ3 (Takada et al., 2007). Most integrins are localized to specific tissues but the major endothelial cell integrin, αVβ3, is widely distributed in endothelium (Takada et al., 2007). The αV subunit has multiple β partners to form heterodimers, but β3 subunit form heterodimers mainly with αV. For this reason, αV can be highly expressed in tissues without β3 because αV has other β partners, but β3 is expressed mainly in tissues with αV (Figure 5).

Three interesting findings were derived from *in-silico* molecular simulation (Nader et al., 2021). Firstly, both RGD in SARS-2S and KGD in SARS-S (Figure 5) are located in a long flexible loop (PDB ID 6M0J for SARS-2S and 5XLR for SARS-S) free to interact with

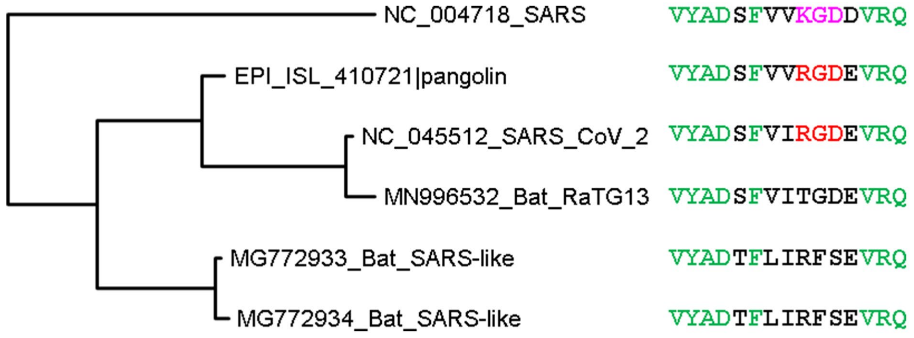


FIGURE 4
Phylogenetic tree of spike protein sequences from SARS-CoV-2 and close relatives. OTU names are in the form of accession (GenBank or GISAID) followed by viral strain designation. The protein sequences were aligned by MAFFT (Katoh and Toh, 2008) with the accurate but slow L-INS-i option. The unrooted phylogenetic tree was reconstructed with PhyML (Guindon and Gascuel, 2003), with the empirical LG substitution matrix and optimization of topology, branch lengths and rates. Identical sites are colored green. The RGD motif, colored red, is shared between SARS-CoV-2 and a close relative isolated from pangolin, and differ from the KGD motif in SARS-CoV by a conservative K403R mutation. The RGD motif is the binding target of the major endothelial cell integrin, αVβ3.

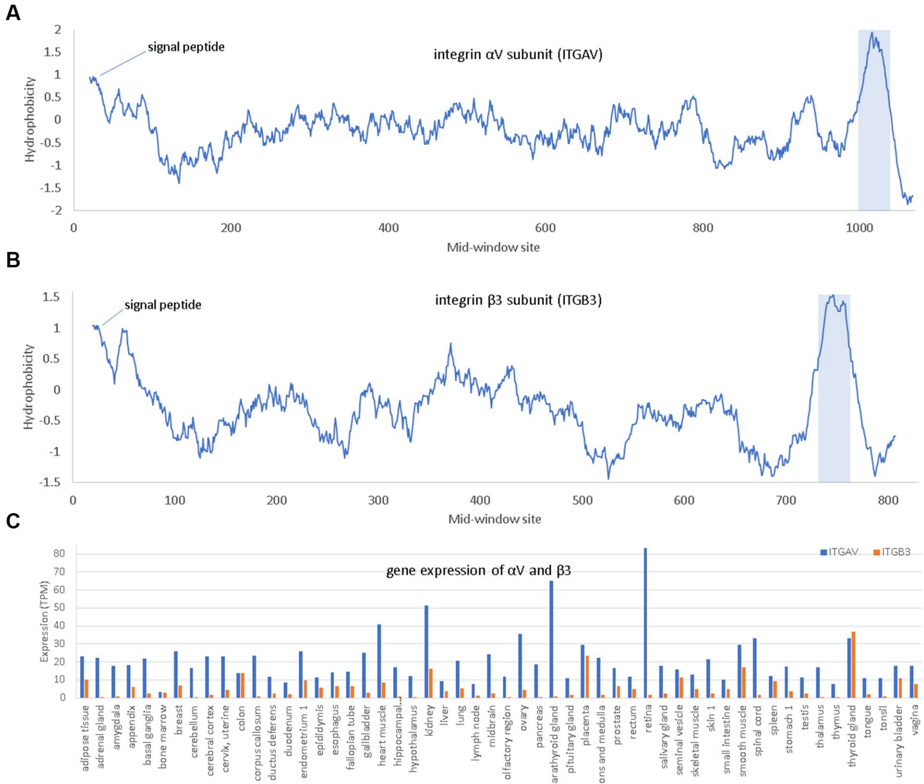


FIGURE 5
Hydrophobicity plot and gene expression for the two subunits (αV and β3) of the major endothelial integrin αVβ3. **(A,B)** Hydrophobicity plot for αV and β3, respectively, with gene names *ITGAV* and *ITGB3*, respectively, along a 40-aa window. **(C)** Gene expression for αV and β3 in different tissues extracted from The Human Protein Atlas (Uhlén et al., 2015), in unit of TPM (transcripts per million). At the N-terminus is the hydrophobic signal peptide. The hydrophobic transmembrane domain is shaded.

other proteins. However, R is larger in volume than K (124 vs. 119), so RGD in SARS-2S is more solvent exposed than KGD in SARS-S. Secondly, the RGD motif fits nicely into the ligand-binding pocket of the host αVβ3. Thirdly, the RGD motif is located about 32 aa upstream of the receptor-binding domain (RBD) for ACE2, and the interaction between the RGD motif in SARS-2S and the host αVβ3 appears physically independent of the interaction between the viral RBD and the host ACE2. Therefore, the host αVβ3 could serve as an additional receptor for SARS-2S independent of ACE2.

Three lines of experimental evidence support the hypothesis that αVβ3 may serve as an alternative receptor (Nader et al., 2021). First, SARS-2S bounds strongly to αVβ3 *in vitro*. Second, SARS-CoV-2 binds strongly to endothelial cells (which could be due to binding of

SARS-2S to ACE2, αVβ3, or any other potential receptors). Third, the binding between SARS-CoV-2 and endothelial cells can be inhibited by Cilengitide (a specific αVβ3 antagonist), which suggests that the binding between SARS-CoV-2 and endothelial cells is mediated by αVβ3. However, the consequence of the binding between SARS-CoV-2 and endothelial cells is not clear. The binding could lead to cell entry of SARS-CoV-2, which would qualify αVβ3 as an alternative receptor. The binding could also interfere with the normal function of αVβ3 which participates in many cellular processes including angiogenesis, cell adhesion and migration, and signaling (Takada et al., 2007), leading to loss of vascular barrier integrity and consequently enhance SARS-CoV-2 infection and increase the severity of COVID-19 (Nader and Kerrigan, 2022).

One may argue that the expression of αVβ3 mainly in endothelial cells would limit its availability for SARS-CoV-2 infection, i.e., SARS-CoV-2 would need to first infect epithelial cells and then traverse to endothelial cells to access αVβ3 as a host receptor. In contrast, ACE2 is expressed in both epithelial and endothelial cells (Hamming et al., 2004). However, the epithelial cells and the endothelial cells are separated by only a very thin basement membrane in lungs. There are also integrins that are expressed in epithelial cells. For example, α5β1, which also binds to the RGD motif, is expressed in a variety of cells including epithelial cells in digestive tract (Sheppard, 1996). Human α5β1 was also implicated in SARS-CoV-2 infection (Robles et al., 2022), and inhibition of human α5β1 by its inhibitor ATN-161 has been shown to reduce viral load in k18-hACE2 transgenic mice infected with SARS-CoV-2 (Amruta et al., 2021; Beddingfield et al., 2021).

How important the RGD motif is in binding to αVβ3 or α5β1 in terms of sequence context could be investigated by either mutating the KGD motif in SARS-S to RGD or investigating the binding properties of the RGD-containing spike protein of the pangolin-derived SARSr. The latter can be done by *in-silico* protein docking and dynamic modelling. Whether the RGD motif binds to integrins as hypothesized above remains controversial (Zech et al., 2021; Othman et al., 2022). Structural modelling of molecular dynamics suggests that the RGD motif does not bind to integrin (Othman et al., 2022).

The R residue in the RGD motif is experimentally shown to enhance binding of the virus to human cells and subsequent viral entry into the cell (Zech et al., 2021). In the bat-derived virus RaTG31, the closest relative of SARS-CoV-2, the site homologous to R403 is T403 (Figure 4). The spike protein of RaTG13 is weak in binding to human ACE2 relative to SARS-2S (Li et al., 2021). Replacing T403 in RaTG13 by R403 enhances viral receptor binding and viral entry into human cells (Zech et al., 2021). Similarly, replacing R403 in SARS-2S by T403 reduces the viral binding and cell entry (Zech et al., 2021). However, R403 was interpreted to enhance the binding between SARS-2S and ACE2, especially between positively charged R403 in SARS-2S and negatively charged E37 in ACE2, but not between SARS-2S and integrin (Zech et al., 2021). This interpretation, based only on structural modelling (Zech et al., 2021), is probably tenuous. The experimentally determined structure (7KNB) of human ACE2 in complex with the SARS-2S trimer (Zhou T. et al., 2020) shows that R403 in SARS-2S and E37 in ACE2 are not close physically (Figure 6A). Of the three R403 residues, one in each of the SARS-2S monomers, the closest distance between E37 and R403 is 11.41 Å apart (Figure 6A). I should mention that there are many different formulations of inter-residue distances. The first (and the simplest) is the distance between the alpha-carbon in one residue and the alpha-carbon in the other residue. This tends to be the most stable across different experimentally determined structures, and is the distance in Figure 6A. The second is to first calculate the centroid for each amino acid, and then calculate the distance between the two centroids. The third is to compute the centroid of the interacting functional groups, e.g., the amino group in Lys and the carboxyl group in glutamate, and then compute the distance between the two centroids.

Instead of interaction between R403 and E37, the structure (7KNB) shows electrostatic interaction between positively charged R403 side chain and the negatively charged D405 side chain (Figure 6B), the two being 4.15 Å apart. The force of attraction between oppositely charged side chains decreases with d^2 (where d is the distance between the interacting partners), so the electrostatic interaction between R403 and D405 within SARS-2S should be much stronger than that between R403 in SARS-2S and E37 in human ACE2. The structural relationship among residues appears consistent across different structural experiments. For example, when human ACE2 is in complex with a SARS-2S monomer instead of a trimer, the distance between R403 in SARS-2S and E37 in ACE2 is 11.15 Å (negligibly smaller than the previous 11.41 Å). Thus, the structure does not suggest a strong interaction between the RGD motif and ACE2, so the RGD motif is free to interact with others, including integrins.

However, there could be an indirect interaction between SARS-2S and integrin through soluble ACE2 (sACE2) as follows. Membrane proteins ADAM-17 and TMPRSS2 cleave the extracellular domain of ACE2 generating sACE2 (Donoghue et al., 2000; Kuba et al., 2010; Scheller et al., 2011; Heurich et al., 2014). sACE2, which features its own RGD motif at sites 204–206, can bind to integrins either in an RGD-dependent or an RGD-independent manner (Clarke et al., 2012). SARS-2S could first bind to sACE2 and then brought close to integrin through sACE2-integrin binding. This is consistent with the observation that shedding of ACE2 results in increased uptake of SARS-CoV virions into host cells (Haga et al., 2008, 2010; Heurich et al., 2014).

There has been insufficient exploration of the functional consequence of the K403R change. Lysine acetylation occurs in both nucleus and cytoplasm (Sadoul et al., 2011; Mu et al., 2020) and removes the positive charge of the lysine residue. Because the lysine in the KGD motif in SARS-S is located in a long flexible loop, it could be acetylated and lose its potential to interact electrostatically with a negative amino acid residue. In contrast, R403 will always be positively charged under normal cellular or tissue pH. It is consequently important to know if K403 in SARS-S is acetylated during virion assembly.

2.2.3. Other candidate receptors and cofactors requiring further empirical confirmation

It has also been suggested that kidney injury molecule-1 (KIM1) may serve as an alternative host receptor for SARS-S and SARS-2S (Yang et al., 2021). However, the evidence is not strong, and the argument that ACE2 alone cannot explain the kidney impairment associated with COVID-19 infection is weak. ACE2 is more highly expressed in kidney than in lung based on tissue-specific expression of protein-coding genes (Fagerberg et al., 2014; Uhlén et al., 2015), as well as on ACE2 activity assays in diabetic mice (Wysocki et al., 2006), which seems sufficient to explain the susceptibility of kidneys to COVID-19 infection without any need to invoke alternative receptors.

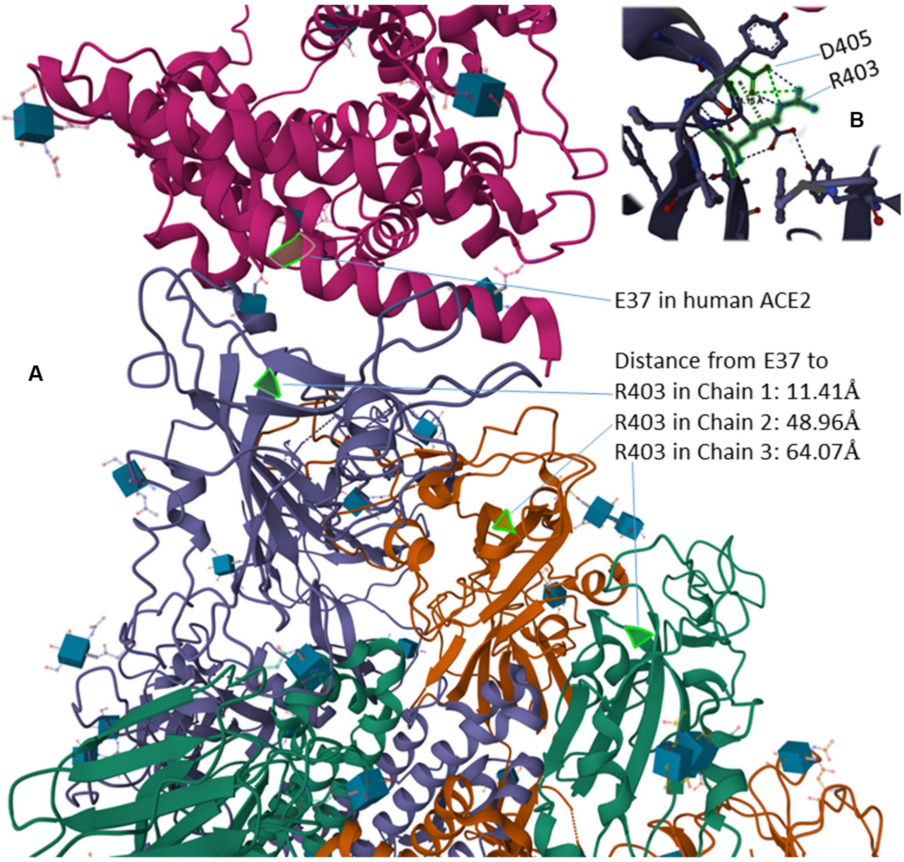


FIGURE 6
Protein structure of human ACE2 in complex with the SARS-2S trimer (PDB accession 7KNB) (Zhou T. et al., 2020). **(A)** The distance from the negatively charged E37 in ACE2 to the three positively charged R403 residues, one in in each of the three SARS-2S monomers. **(B)** R403 and D405 within SARS-2S, which are 4.15 Å apart, interact with each other electrostatically.

Kidney impairment associated with COVID-19 can be explained by the impairment of ACE2 function in degrading Ang II (Figure 7A). ACE2 protects kidneys from unchecked RAS responses including hypertension, inflammation and tissue damage (Kuba et al., 2010; Soler et al., 2013). COVID-19 infection in kidney destroys kidney cells expressing ACE2 and exposes the kidneys to unchecked RAS responses causing kidney impairment.

Two other membrane proteins, AXL and CD147, may deserve attention. AXL is a putative cell receptor for Zika virus (Nowakowski et al., 2016), and implicated in mediating cell entry via the endosome pathway by SRAR-CoV-2 (Bohan et al., 2021). CD147 is a membrane glycoprotein known to be involved in infection by eukaryotic, prokaryotic and viral pathogens (Fenizia et al., 2021), and may also bind to SARS-2S and mediate the cell entry of SARS-CoV-2 through endocytosis (Brodowski et al., 2022; Cavezzi et al., 2022; Kalejaiye et al., 2022), although infection mediated by CD147 is most likely secondary because CD147 is highly expressed in neural tissues but not in respiratory tract (Qiao et al., 2020). Both genes appear weakly expressed in lungs based on tissue-specific gene expression data in The Human Protein Atlas (Uhlén et al., 2015). However, as I mentioned before, ACE2 is highly expressed in type II pneumocytes in lungs (Hamming et al., 2004; To et al., 2004; To and Lo, 2004; Mossel et al., 2008; Xu et al., 2020; Zhao et al., 2020), so the low tissue-specific expression of these two genes may not exclude the possibility of high expression in certain types of cells.

Another cofactor proposed to bind SARS-2S and facilitate SARS-CoV-2 cell entry is sialic acid-containing glycolipids (Nguyen et al., 2022). Depletion of these glycolipids decreases SARS-CoV-2 infection. However, the depletion of these glycolipids could have multiple consequences. It may impair membrane integrity and render epithelial cells more exposed. For example, mucins are important membrane component, and the loss of mucins enhances SARS-CoV-2 infection (Biering et al., 2022). Also, loss of membrane integrity may cause shedding of ACE2 and candidate cofactors such as NRP1 and consequently generate outcomes that are difficult to interpret.

The hypothesis of sialic acid-containing glycolipids or sialylated glycans as a receptor for SARS-2S is vague because many membrane proteins are sialylated glycoproteins, including ACE2 and CD147. However, ACE2 glycan processing has little effect on SARS-CoV-2 recognition (Allen et al., 2021). One may therefore infer that it is features other than sialylated glycans that is important in mediating SARS-CoV-2 infection.

3. Host receptors and cofactors as drug targets?

ACE2 has been studied as a drug target ever since it was identified as the host receptor for SARS-CoV. The rationale seems straightforward. Given that ACE2 is a gate to let SARS-CoV-2 into the

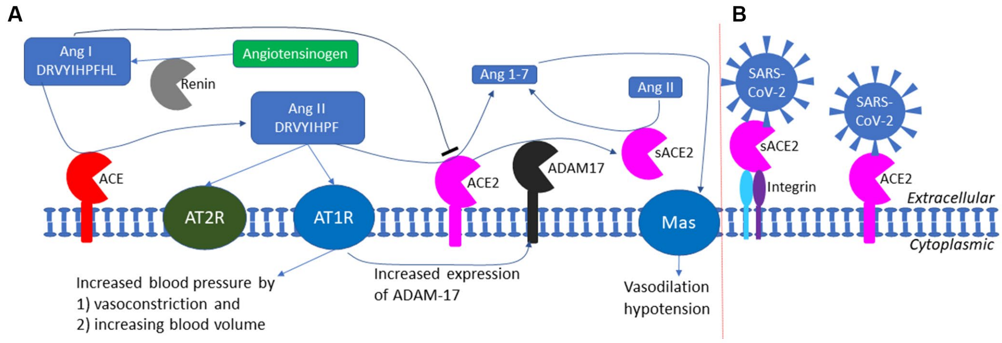


FIGURE 7
The RAS (renin-angiotensin system) and the exploitation of ACE2 by SARS-CoV-2 as a cell receptor. **(A)** The RAS system in maintaining blood pressure homeostasis through negative feedback. Low blood pressure triggers the release of renin which cleaves angiotensinogen (colored green) to produce Ang I; ACE cleaves Ang I to generate Ang II which binds to AT1R receptor to increase blood pressure; ACE2 degrade Ang II to prevent prolonged hypertension. ADAM17 cleaves the extracellular domain of ACE2 to generate soluble ACE2 (sACE2) which is also enzymatically active in degrading Ang II. **(B)** SARS-CoV-2 could anchor itself to the cell membrane by binding either to ACE2 or to sACE2 through other membrane proteins such as integrin (not to scale).

cell, the gate should be blocked. Two questions need to be answered. First, does ACE2 abundance really increase the risk to COVID-19 (where "risk" is a term combining the vulnerability to COVID-19 and the severity of COVI-19 symptoms)? Second, how should ACE2 be targeted to reduce the risk to COVID-19 without interfering with the essential function of ACE2?

3.1. Does ACE2 abundance increase with the risk to COVID-19?

There are no direct experiments on ACE2 abundance and the risk to COVID-19. Consequently, an indirect approach has been used to address the question. COVID-19 symptoms are more severe in old-age group (OG) than the young or middle age group (YG). ACE2 abundance was obtained from different age groups to establish the relationship between ACE2 abundance and age. If OG expresses more ACE2 than YG, then ACE2 abundance is a likely contributor to the severity of COVID-19 in OG.

In a well-planned comparative study among groups of different ages and ACE2 expression in a hospital cohort (Plaas et al., 2021), ACE2 expression is higher in YG than in OG. Similarly, careful and structured meta-analysis also supports higher ACE2 in YG than in OG (Bastolla et al., 2022). However, large-scale compilation of data sometimes leads to contradictory results (Zheng, 2022). Some of the discrepancy could be explained by pooling unbalanced data. For example, if ACE2 expression levels in YG and OG in region 1 are $ACE2_{YG.r1} = 20$ and $ACE2_{OG.r1} = 10$, respectively, but $ACE2_{YG.r2} = 40$ and $ACE2_{OG.r2} = 30$ in region 2 (where the subscript r stands for region). If sample size is $n_{YG.r1} = 200$ and $n_{OG.r1} = 20$, but $n_{YG.r2} = 50$ and $n_{OG.r2} = 500$, then the weighted mean of ACE2 expression for YG and OG, pooled over the two regions, would become

$$\boldsymbol{Mean_{YG}} = \frac{\boldsymbol{ACE2_{YG.r1} \times n_{YG.r1} + ACE2_{YG.r2} \times n_{YG.r2}}}{\boldsymbol{n_{YG.r1} + n_{YG.r2}}} = \frac{\mathbf{20 \times 200 + 40 \times 50}}{\mathbf{200 + 50}} = \mathbf{24}$$

$$\boldsymbol{Mean_{OG}} = \frac{\boldsymbol{ACE2_{OG.r1} \times n_{OG.r1} + ACE2_{OG.r2} \times n_{OG.r2}}}{\boldsymbol{n_{OG.r1} + n_{OG.r2}}} = \frac{\mathbf{10 \times 20 + 30 \times 500}}{\mathbf{20 + 500}} \approx \mathbf{29.23}$$

These two mean values would mislead us to conclude that ACE2 expression is higher in OG than in YG. This Simpson paradox, typically illustrated with the data from surgery on kidney stone data (Xia, 2018a), is often forgotten in large-scale data compilations.

The observation that OG has lower ACE2 abundance than YG (Plaas et al., 2021; Bastolla et al., 2022) seems incompatible with the observation that OG suffers more from COVID-19 than YG. There are two explanations. First, although ACE2 is higher in YG than in OG, the level of ACE2 in OG is still sufficient for initiating SARS-CoV-2 infection. Second, as I mentioned before, SARS-CoV-2 infection is associated with syncytia formation. This means that, once SARS-CoV-2 infected a cell, the spread of SARS-CoV-2 from this infected cell to neighboring uninfected cell may not need ACE2. In contrast to SARS-CoV-2, SARS-CoV infection does not form syncytia, so infection of new cells requires ACE2. Because ACE2 is more abundant in YG than in OG, people in YG tend to have higher risk to SARS-CoV than those in OG, which is consistent with SARS epidemiological data. I should emphasize that previous studies quantifying ACE2 expression does not take into consideration the sACE2 (the soluble portion of ACE2), so one should be cautious in interpreting ACE2 abundance and COVID-19 risk in different age groups.

3.2. Targeting ACE2 to reduce the COVID-19 risk without impacting ACE2 function

ACE2 has multiple functions (Fyhrquist and Saijonmaa, 2008; Kuba et al., 2010), but its most well-documented function is to buffer the RAS (renin-angiotensin system) effect for blood pressure homeostasis (Figure 7A). Human liver produces the 485-aa angiotensinogen which, after cleaving the 33-aa N-terminal signal peptide (Kumar et al., 2011), is released as the 452-aa mature circulating angiotensinogen (Figure 7A). When blood pressure falls, renin released from kidney cells converts angiotensinogen to Ang I, with cleavage between 10 L and 11 V (Yan et al., 2019). Ang I is in turn converted by ACE, a peptidyl dipeptidase, to Ang II (Figure 7A). Ang II interacts with the two receptors, but mainly through receptor AT1R (Figure 7A), to increase the blood pressure by (1) increasing the blood volume and (2) shrinking the blood vessel (vasoconstriction). This RAS function, if unchecked, would lead to hypertension, inflammation, tissue damage, heart failure, and other cardiovascular abnormalities (Kuba et al., 2010). Carboxypeptidase ACE2 takes short oligopeptides such as peptide hormones and cleaves efficiently at the Pro^X junction (where X is a hydrophobic amino acid at the C-terminus) (Donoghue et al., 2000; Tipnis et al., 2000; Dales et al., 2002). This reduction in Ang II, together with the binding of the resulting Ang1-7 to MAS receptors (Figure 7A), buffers the RAS effect to maintain blood pressure homeostasis. People with low levels of ACE2 tend to have high level of Ang II and hypertension, and need to be treated with ACE inhibitors so that Ang I is not converted to Ang II (Imai et al., 2005; Kuba et al., 2010). Alternatively, one may use drugs such as griseofulvin (a known vasodilator) which decreases blood pressure (Rubin, 1963; Aldinger, 1968). A recent study suggests that the griseofulvin effect may be mediated by its binding to ACE2 (Aris et al., 2022), i.e., griseofulvin may be an ACE2 enhancer.

Oligopeptides with His^X at the C-terminus can also serve as substrates for ACE2, although the cleavage is not as efficient as Pro^X (Dales et al., 2002). ACE2 can therefore cleave the terminal leucine in Ang I (Figure 7A). However, Ang I also inhibits ACE2 activity (Dales et al., 2002; Figure 7A), which is essential for the accumulation of Ang II. High levels of Ang I indicates weak activity of ACE and weak RAS effect, so ACE2 should be at low activity as well. When Ang I is converted to Ang II, the inhibitory effect of Ang I on ACE2 is removed, and the active ACE2 clears Ang II to prevent hypertension.

Ignoring the sACE2 activity in degrading Ang II may lead to misunderstanding of the negative feedback regulation of the RAS system. For example, an increase in Ang II level was associated with a decrease in myocardial ACE2 protein level (Patel et al., 2014). If one takes the decreased myocardial ACE2 protein level as decreased ACE2 activity, then one would conclude that an increase in Ang II, instead of increasing the ACE2 activity to degrade Ang II, actually decreases the ACE2 activity. This would imply a prolonged high concentration of Ang II because such a high Ang II concentration would seem to decrease ACE2 that degrade Ang II, so an increase in Ang II would lead to further increase in Ang II. However, the observed decrease in myocardial ACE2 may not imply decreased ACE2 activity because such decrease in myocardial ACE2 may be associated with an increase in sACE2 (Figure 7A). Because sACE2 is also enzymatically active in degrading Ang II (Kuba et al., 2010; Patel et al., 2014), the total ACE2 activity may not be decreased even though the myocardial ACE2 level is decreased. One needs to measure total ACE2 activity, including both the membrane-bound ACE2 and sACE2, in converting Ang II to Ang1-7 (Figure 7A).

Given the essential function of ACE2, simply downregulating ACE2 expression to reduce SARS-CoV-2 infection (Brevini et al., 2023) may incur the side effect of insufficient ACE2 activity. However, low ACE2 activity could be compensated by inhibitors of ACE such as MLN4760 (Dales et al., 2002) or angiotensin-receptor blocker (ARB) (Kuba et al., 2010; Bosso et al., 2020) or vasodilators such as griseofulvin (Aris et al., 2022). The binding site between ACE2 and the RBD of SARS-S and SARS-2S (Gui et al., 2017; Wang et al., 2020; Zhou T. et al., 2020; Xu et al., 2021) does not cover the zinc-binding metallopeptidase domain of ACE2 (Kuba et al., 2010). Therefore, it is theoretically possible to develop a drug that would interfere with the binding between the host ACE2 and the viral RBD without affecting ACE2's function in converting Ang II to Ang 1–7. Many drug-screening studies check only binding affinity between a candidate drug and human ACE2 (Mathew et al., 2021; Aris et al., 2022). A reasonable drug candidate should bind to the site of interaction between SARS-2S and ACE2, but does not bind to the zinc-binding catalytic site of ACE2.

The same principle of reducing infection without impacting function should be applied not only to host receptors such as ACE2, but also other cofactors such as NRP1 (Cantuti-Castelvetri et al., 2020; Daly et al., 2020) because a proper level of NRP1 protein is essential for cardiovascular and neuronal development (Guo and Vander Kooi, 2015). Overexpression of the gene (Kawasaki et al., 1999), or knock-out of the gene (Kitsukawa et al., 1995) are both lethal in mice.

The function implication of sACE2 remains elusive. The cleavage of ACE2 by ADAM-17 (Haga et al., 2008, 2010; Scheller et al., 2011) is activated in SARS-CoV infection, generating sACE2 (Figure 7A). This could be either a host-mediated protection response or a virus-mediated response to colonize cells not expressing ACE2. If membrane-bound ACE2 mediates viral entry, then cleaving them off membrane would protect the ACE2-expressing cell from infection. However, this hypothesis of host-mediated protection response is contradicted by the observation that the generation of sACE2 enhances infection (Haga et al., 2008, 2010). It is possible that SARS-CoV-2 can bind to both membrane-bound ACE2 and sACE2 which could then bind to membrane proteins such as neuropilin-1 (Cantuti-Castelvetri et al., 2020; Daly et al., 2020), integrins (Nader et al., 2021; Nader and Kerrigan, 2022), or other membrane proteins to anchor SARS-CoV-2 to host cell membrane that do not have ACE2 (Figure 7B). SARS-2S could first bind to sACE2 which then binds to membrane integrin (Figure 7B). This is consistent with the observation that the shedding of ACE2 results in increased uptake of SARS-CoV virions into host cells, and therefore supports the alternative hypothesis of virus-mediated response to colonize host cells that do not express ACE2.

4. Predicting mammalian species susceptible to SARS-CoV-2

Many studies have used similarity in ACE2 sequences and sharing of interacting amino acids between ACE2 and the viral RBD to predict vulnerability of other mammalian species to COVID-19 (Damas et al., 2020; Shi et al., 2020; Kruglikov et al., 2021; Wei et al., 2021). The general rationale is that the ACE2 of a mammalian species highly similar to the ACE2 of susceptible species (e.g., human) would serve as a host receptor for SARS-CoV-2 and predispose the species to SARS-CoV-2 infection. SARS-2S can use ACE2 in many species for cell entry (Hossain et al., 2020; Shi et al., 2020; Zhai et al., 2020; Li et al., 2021), including all tested primate species, pangolins, and several carnivorous species.

Two different approaches have been used for the prediction. The first and the simplest index of vulnerability is based on phylogenetic analysis of aligned ACE2 sequences (Figure 8). Species with a short root-to-tip distance have relatively conserved ACE2, and these species, colored in red (Figure 8), tend to be susceptible to SARS-CoV-2 infection. Within rodents, the golden hamster (*Mesocricetus auratus*) can be infected by both SARS-CoV and SARS-CoV-2, and its ACE2 is closer to the putative root than mouse and rat that are not vulnerable unless humanized with human ACE2. Within Chiroptera, *Rhinolophus affinis* is more vulnerable than other bat species (Li et al., 2021) and its ACE2 is closer to the putative root than others (Figure 8). Within primates, human ACE2 is closer to the putative root than other primates, and humans appear to be more vulnerable to COVID-19 infection than other primates.

An ideal receptor protein from a viral perspective is one that is (1) highly expressed in epithelial cell surface of mammalian respiratory or digestive tracts, (2) functionally important so that its expression cannot be readily downregulated by the host in response to the infection, and (3) strongly conserved in evolution so many mammalian species can serve as potential hosts. We have shown previously that ACE2 meets the first two criteria. Figure 8 shows that ACE2 also meets the last criterion.

The second approach for predicting species vulnerability incorporates information from protein structures. The characterization of the structure of SARS-2S (Hoffmann et al., 2020; Walls et al., 2020; Wrapp et al., 2020; Yan et al., 2020), especially those with the SARS-2S trimer in complex with ACE2 (Gui et al., 2017; Wang et al., 2020; Zhou T. et al., 2020; Xu et al., 2021), provides a list of amino acids in physical contact between ACE2 and SARS-2S (Lu et al., 2015; Adhikari et al., 2020; Adhikari and Ching, 2020). The sharing of the interacting amino acids in ACE2 have been used to predict which mammalian species have an ACE2 that can serve as a host receptor for SARS-CoV-2 infection (Shi et al., 2020; Zhai et al., 2020; Kruglikov et al., 2021; Wei et al., 2021).

One stretch of five amino acids in human ACE2 (hACE2), 353KGDFR357 (Figure 9), is particularly worth of highlighting because (1) the amino acid composition in the 5-aa motif implies that it is highly hydrophilic and should stay on the surface of the protein, (2) it is in close physical contact with the 500TNGVGY505 segment in SARS-2S (Figure 9), based on the structure 6M0J (Lan et al., 2020), and (3) it is highly conserved, except 354G, across representative species in Carnivora, Artiodactyla and Chiroptera (Wei et al., 2021). A highly conserved hACE2 binding motif means that SARS-2S can not only infect all hACE2, but also ACE2 in a variety of mammalian species. This would generate an unusually large array of potential host species and a consequently large viral reservoir in nature. In contrast, if the interacting motif in hACE2 binding to SARS-2S were highly variable even among different human populations, then SARS-CoV-2 would only be able to infect humans or just a specific genetically homogeneous human population.

Among rodents, the mouse and rat ACE2 that cannot serve as a SARS-2S receptor has 353H. One may infer that an H353K mutation would change the mouse ACE2 to a SARS-2S receptor. Such a mutation has been carried out both in nature and by virologists. The golden hamster (*Mesocricetus auratus*) has 353 K and is susceptible to SARS-CoV-2 infection. Experimental introduction of a H353K mutation into mouse ACE2, i.e., hmACE2.1 in Adams et al. (2021), substantially improved the function of mouse ACE2 as a SARS-2S

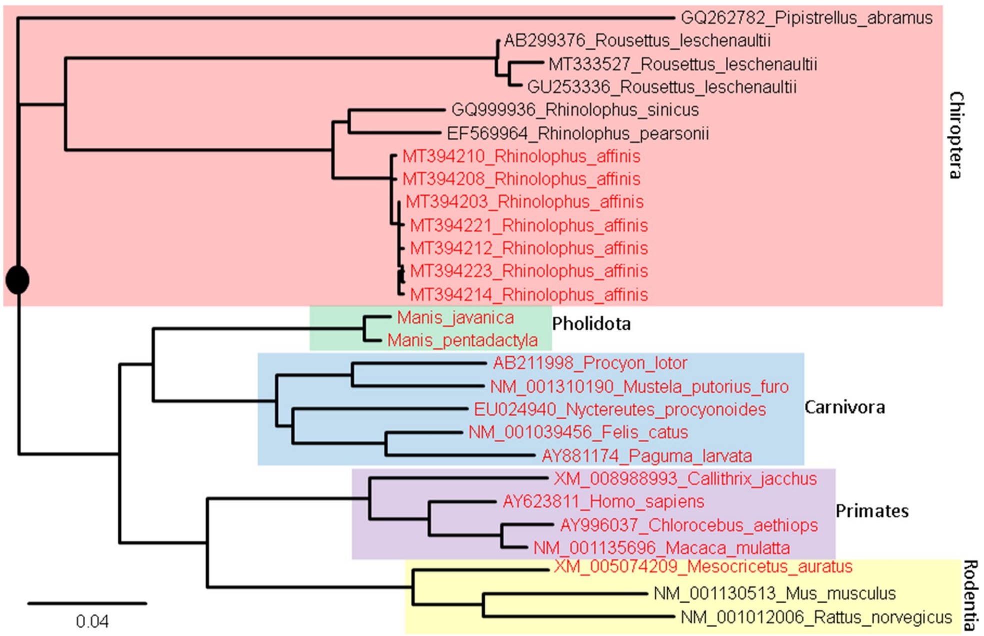


FIGURE 8
ACE2 Phylogeny of representative mammalian species in Chiroptera (bats), Pholidota (pangolins), Carnivora (felids and canids), Primates, and Rodentia. The protein sequences were aligned by MAFFT (Katoh and Toh, 2008) with the accurate but slow L-INS-i option. The unrooted phylogenetic tree was reconstructed with PhyML (Guindon and Gascuel, 2003), with the empirical LG substitution matrix and optimization of topology, branch lengths and rates. The reconstructed tree is unrooted but is rooted by mid-point. The species in red have been empirically shown to be vulnerable to SARS-CoV-2 infection. They are closer to (have fewer substitutions in ACE2 from) the putative common ancestor (indicated by a black solid circle) than species not vulnerable to SARS-CoV-2 infection.

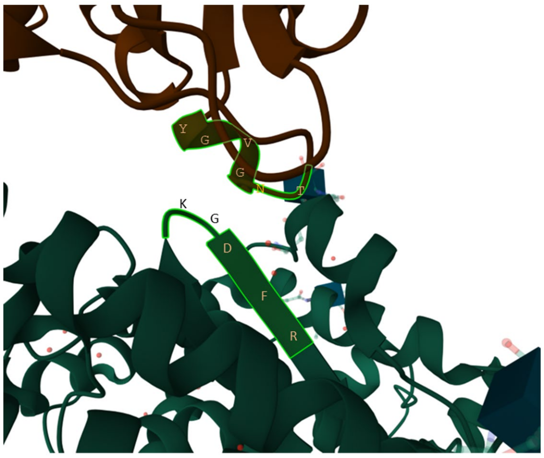


FIGURE 9
The 353KGDFR357 segment in human ACE2 in close contact with the 500TNGVGY505 segment in SARS-2S, based on structure 6M0J (Lan et al., 2020).

receptor. There should be more coevolutionary studies between hosts and pathogens.

Prediction of species vulnerability based on ACE2 alone is confounded by many factors. First, successful viral infection involves multiple steps including cell attachment, cell entry, evasion of host immune systems, viral genome replication, transcription, translation, packaging of virions, and cell lysis and viral release. Having a suitable ACE2 receptor represents just one of these steps. For example, pig ACE2 appears to serve as a good receptor for SARS-2S (Li et al., 2021), but SARS-CoV-2 does not infect pigs. Second, an ACE2 in an animal highly similar to human ACE2 may express little ACE2 in respiratory tract. For example, in contrast to humans and other primates, dogs express relatively little ACE2 in respiratory tract but high ACE2 in digestive tract (Naqvi et al., 2019; Zhai et al., 2020), so it is not surprising to find positive rectal swabs but not in pharyngeal swabs in experimental dogs a few days after the inoculation with SARS-CoV-2 (Shi et al., 2020). Thus, a prediction that SARS-CoV-2 would cause respiratory diseases in dogs because dogs have an ACE2 similar to human ACE2 is not quite true because of the low expression of ACE2 in the respiratory tract of dogs.

In summary, host receptor identification and related studies require a multidisciplinary approach involving diverse types of data and integrative data analyses. This review may contribute to the design of training programs for future virologists.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Funding

This research was funded by a Discovery Grant from the Natural Science and Engineering Research Council (NSERC, RGPIN/2018–03878) of Canada. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Acknowledgments

The author would like to thank D. Gray, Y. Wei, J. Xia and Z. Xie for discussion and comments.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Adams, L. E., Dinnon, K. H. 3rd, Hou, Y. J., Sheahan, T. P., Heise, M. T., and Baric, R. S. (2021). Critical ACE2 determinants of SARS-CoV-2 and group 2B coronavirus infection and replication. *MBio* 12:e03149-20. doi: 10.1128/mBio.03149-20

Adhikari, P., and Ching, W. Y. (2020). Amino acid interacting network in the receptor-binding domain of SARS-CoV-2 spike protein. *RSC Adv.* 10, 39831–39841. doi: 10.1039/D0RA08222H

Adhikari, P., Li, N., Shin, M., Steinmetz, N. F., Twarock, R., Podgornik, R., et al. (2020). Intra- and intermolecular atomic-scale interactions in the receptor binding domain of SARS-CoV-2 spike protein: implication for ACE2 receptor binding. *Physic. Chem. Chem. Physic.* 22, 18272–18283. doi: 10.1039/D0CP03145C

Aldinger, E. E. (1968). Cardiovascular effects of griseofulvin. *Circ. Res.* 22, 589–593. doi: 10.1161/01.RES.22.5.589

Allen, J. D., Watanabe, Y., Chawla, H., Newby, M. L., and Crispin, M. (2021). Subtle influence of ACE2 glycan processing on SARS-CoV-2 recognition. *J. Mol. Biol.* 433:166762. doi: 10.1016/j.jmb.2020.166762

Amruta, N., Engler-Chiurazzi, E. B., Murray-Brown, I. C., Gressett, T. E., Biose, I. J., Chastain, W. H., et al. (2021). In vivo protection from SARS-CoV-2 infection by ATN-161 in k18-hACE2 transgenic mice. *Life Sci.* 284:119881. doi: 10.1016/j.lfs.2021.119881

Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., and Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nat. Med.* 26, 450–452. doi: 10.1038/s41591-020-0820-9

Aris, P., Wei, Y., Mohamadzadeh, M., and Xia, X. (2022). Griseofulvin: an updated overview of old and current knowledge. *Molecules (Basel, Switzerland).* 27:7034. doi: 10.3390/molecules27207034

Bao, L., Deng, W., Huang, B., Gao, H., Liu, J., Ren, L., et al. (2020). The pathogenicity of SARS-CoV-2 in hACE2 transgenic mice. *Nature* 583, 830–833. doi: 10.1038/s41586-020-2312-y

Bastolla, U., Chambers, P., Abia, D., Garcia-Bermejo, M.-L., and Fresno, M. (2022). Is COVID-19 severity associated with ACE2 degradation? *Front. Drug Discov.* 1:789710. doi: 10.3389/fddsv.2021.789710

Beddingfield, B. J., Iwanaga, N., Chapagain, P. P., Zheng, W., Roy, C. J., Hu, T. Y., et al. (2021). The integrin binding peptide, ATN-161, as a novel therapy for SARS-CoV-2 infection. *JACC Basic Transl Sci.* 6, 1–8. doi: 10.1016/j.jacbts.2020.10.003

Belouzard, S., Chu, V. C., and Whittaker, G. R. (2009). Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc. Natl. Acad. Sci. U. S. A.* 106, 5871–5876. doi: 10.1073/pnas.0809524106

Biering, S. B., Sarnik, S. A., Wang, E., Zengel, J. R., Leist, S. R., Schäfer, A., et al. (2022). Genome-wide bidirectional CRISPR screens identify mucins as host factors modulating SARS-CoV-2 infection. *Nat. Genet.* 54, 1078–1089. doi: 10.1038/s41588-022-01131-x

Bohan, D., van Ert, H., Ruggio, N., Rogers, K. J., Badreddine, M., Aguilar Briseño, J. A., et al. (2021). Phosphatidylserine receptors enhance SARS-CoV-2 infection. *PLoS Pathog.* 17:e1009743. doi: 10.1371/journal.ppat.1009743

Bosso, M., Thanaraj, T. A., Abu-Farha, M., Alanbaei, M., Abubaker, J., and Al-Mulla, F. (2020). The two faces of ACE2: the role of ACE2 receptor and its polymorphisms in hypertension and COVID-19. *Mol Ther Methods Clin Dev.* 18, 321–327. doi: 10.1016/j.omtm.2020.06.017

Brevini, T., Maes, M., Webb, G. J., John, B. V., Fuchs, C. D., Buescher, G., et al. (2023). FXR inhibition may protect from SARS-CoV-2 infection by reducing ACE2. *Nature*. 615, 134–142. doi: 10.1038/s41586-022-05594-0

Brodowski, M., Pierpaoli, M., Janik, M., Kowalski, M., Ficek, M., Slepski, P., et al. (2022). Enhanced susceptibility of SARS-CoV-2 spike RBD protein assay targeted by cellular receptors ACE2 and CD147: multivariate data analysis of multisine impedimetric response. *Sens Actuat. B Chem.* 370:132427. doi: 10.1016/j.snb.2022.132427

Cantuti-Castelvetri, L., Ojha, R., Pedro, L. D., Djannatian, M., Franz, J., Kuivanen, S., et al. (2020). Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science* 370, 856–860. doi: 10.1126/science.abd2985

Cavezzi, A., Menicagli, R., Troiani, E., and Corrao, S. (2022). COVID-19, cation dysmetabolism, sialic acid, CD147, ACE2, viroporins, hepcidin and ferroptosis: a possible unifying hypothesis. *F1000Research.* 11:102. doi: 10.12688/f1000research.108667.2

Clarke, N. E., Fisher, M. J., Porter, K. E., Lambert, D. W., and Turner, A. J. (2012). Angiotensin converting enzyme (ACE) and ACE2 bind integrins and ACE2 regulates integrin signalling. *PLoS One* 7:e34747. doi: 10.1371/journal.pone.0034747

Coutard, B., Valle, C., de Lamballerie, X., Canard, B., Seidah, N. G., and Decroly, E. (2020). The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antivir. Res.* 176:104742. doi: 10.1016/j.antiviral.2020.104742

Dales, N. A., Gould, A. E., Brown, J. A., Calderwood, E. F., Guan, B., Minor, C. A., et al. (2002). Substrate-based design of the first class of angiotensin-converting enzyme-related carboxypeptidase (ACE2) inhibitors. *J. Am. Chem. Soc.* 124, 11852–11853. doi: 10.1021/ja0277226

Daly, J. L., Simonetti, B., Klein, K., Chen, K. E., Williamson, M. K., Antón-Plágaro, C., et al. (2020). Neuropilin-1 is a host factor for SARS-CoV-2 infection. *Science* 370, 861–865. doi: 10.1126/science.abd3072

Damas, J., Hughes, G. M., Keough, K. C., Painter, C. A., Persky, N. S., Corbo, M., et al. (2020). Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. *Proc. Natl. Acad. Sci.* 117, 22311–22322. doi: 10.1073/pnas.2010146117

Donoghue, M., Hsieh, F., Baronas, E., Godbout, K., Gosselin, M., Stagliano, N., et al. (2000). A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9. *Circ. Res.* 87, E1–E9. doi: 10.1161/01.RES.87.5.e1

Fagerberg, L., Hallström, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., et al. (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molec Cell. Proteom.* 13, 397–406. doi: 10.1074/mcp.M113.035600

Fenizia, C., Galbiati, S., Vanetti, C., Vago, R., Clerici, M., Tacchetti, C., et al. (2021). SARS-CoV-2 entry: at the crossroads of CD147 and ACE2. *Cells* 10:1434. doi: 10.3390/cells10061434

Fyhrquist, F., and Saijonmaa, O. (2008). Renin-angiotensin system revisited. *J. Intern. Med.* 264, 224–236. doi: 10.1111/j.1365-2796.2008.01981.x

Gui, M., Song, W., Zhou, H., Xu, J., Chen, S., Xiang, Y., et al. (2017). Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell Res.* 27, 119–129. doi: 10.1038/cr.2016.152

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704.

Guo, H. F., and Vander Kooi, C. W. (2015). Neuropilin functions as an essential cell surface receptor. *J. Biol. Chem.* 290, 29120–29126. doi: 10.1074/jbc.R115.687327

Haga, S., Nagata, N., Okamura, T., Yamamoto, N., Sata, T., Yamamoto, N., et al. (2010). TACE antagonists blocking ACE2 shedding caused by the spike protein of SARS-CoV are candidate antiviral compounds. *Antivir. Res.* 85, 551–555. doi: 10.1016/j.antiviral.2009.12.001

Haga, S., Yamamoto, N., Nakai-Murakami, C., Osawa, Y., Tokunaga, K., Sata, T., et al. (2008). Modulation of TNF-alpha-converting enzyme by the spike protein of SARS-CoV and ACE2 induces TNF-alpha production and facilitates viral entry. *Proc. Natl. Acad. Sci. U. S. A.* 105, 7809–7814. doi: 10.1073/pnas.0711241105

Hamming, I., Timens, W., Bulthuis, M. L., Lely, A. T., Navis, G., and 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. doi: 10.1002/path.1570

Heurich, A., Hofmann-Winkler, H., Gierer, S., Liepold, T., Jahn, O., and Pöhlmann, S. (2014). TMPRSS2 and ADAM17 cleave ACE2 differentially and only proteolysis by TMPRSS2 augments entry driven by the severe acute respiratory syndrome coronavirus spike protein. *J. Virol.* 88, 1293–1307. doi: 10.1128/JVI.02202-13

Hikmet, F., Méar, L., Edvinsson, Å., Micke, P., Uhlén, M., and Lindskog, C. (2020). The protein expression profile of ACE2 in human tissues. *Mol. Syst. Biol.* 16:e9610. doi: 10.15252/msb.20209610

Hoffmann, M., Kleine-Weber, H., and Pöhlmann, S. (2020). A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Mol. Cell* 78, 779–784.e5. doi: 10.1016/j.molcel.2020.04.022

Hossain, M. G., Javed, A., Akter, S., and Saha, S. (2020). SARS-CoV-2 host diversity: an update of natural infections and experimental evidence. *J. Microbiol Immunol. Infect* 54, 175–181. doi: 10.1016/j.jmii.2020.06.006

Imai, Y., Kuba, K., Rao, S., Huan, Y., Guo, F., Guan, B., et al. (2005). Angiotensin-converting enzyme 2 protects from severe acute lung failure. *Nature* 436, 112–116. doi: 10.1038/nature03712

Inoue, Y., Tanaka, N., Tanaka, Y., Inoue, S., Morita, K., Zhuang, M., et al. (2007). Clathrin-dependent entry of severe acute respiratory syndrome coronavirus into target cells expressing ACE2 with the cytoplasmic tail deleted. *J. Virol.* 81, 8722–8729. doi: 10.1128/JVI.00253-07

Kalejaiye, T. D., Bhattacharya, R., Burt, M. A., Travieso, T., Okafor, A. E., Mou, X., et al. (2022). SARS-CoV-2 employ BSG/CD147 and ACE2 receptors to directly infect human induced pluripotent stem cell-derived kidney podocytes. *Front. Cell Develop. Biol.* 10:855340. doi: 10.3389/fcell.2022.855340

Katoh, K., and Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform*. 9, 286–298.

Kawasaki, T., Kitsukawa, T., Bekku, Y., Matsuda, Y., Sanbo, M., Yagi, T., et al. (1999). A requirement for neuropilin-1 in embryonic vessel formation. *Development* 126, 4895–4902. doi: 10.1242/dev.126.21.4895

Kitsukawa, T., Shimono, A., Kawakami, A., Kondoh, H., and Fujisawa, H. (1995). Overexpression of a membrane protein, neuropilin, in chimeric mice causes anomalies in the cardiovascular system, nervous system and limbs. *Development* 121, 4309–4318. doi: 10.1242/dev.121.12.4309

Kruglikov, A., Rakesh, M., Wei, Y., and Xia, X. (2021). Applications of protein secondary structure algorithms in SARS-CoV-2 research. *J. Proteome Res.* 20, 1457–1463. doi: 10.1021/acs.jproteome.0c00734

Kuba, K., Imai, Y., Ohto-Nakanishi, T., and Penninger, J. M. (2010). Trilogy of ACE2: a peptidase in the renin-angiotensin system, a SARS receptor, and a partner for amino acid transporters. *Pharmacol. Ther.* 128, 119–128. doi: 10.1016/j.pharmthera.2010.06.003

Kuba, K., Imai, Y., Rao, S., Gao, H., Guo, F., Guan, B., et al. (2005). A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus–induced lung injury. *Nat. Med.* 11, 875–879. doi: 10.1038/nm1267

Kumar, R., Baker, K. M., and Pan, J. (2011). "Chapter 9 - activation of the renin-angiotensin system in heart failure" in *Heart failure: a companion to Braunwald's heart disease (second edition)* ed. D. L. Mann (Philadelphia: W.B. Saunders), 134–151.

Kyte, J., and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105–132. doi: 10.1016/0022-2836(82)90515-0

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., et al. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581, 215–220. doi: 10.1038/s41586-020-2180-5

Li, P., Guo, R., Liu, Y., Zhang, Y., Hu, J., Ou, X., et al. (2021). The *Rhinolophus affinis* bat ACE2 and multiple animal orthologs are functional receptors for bat coronavirus RaTG13 and SARS-CoV-2. *Sci. Bull. (Beijing).* 66, 1215–1227. doi: 10.1016/j.scib.2021.01.011

Li, M.-Y., Li, L., Zhang, Y., and Wang, X.-S. (2020). Expression of the SARS-CoV-2 cell receptor gene ACE2 in a wide variety of human tissues. *Infect. Dis. Poverty*. 9:45. doi: 10.1186/s40249-020-00662-x

Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., et al. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 426, 450–454. doi: 10.1038/nature02145

Lo, Y.-L., Liou, G.-G., Lyu, J.-H., Hsiao, M., Hsu, T.-L., and Wong, C.-H. (2016). Dengue virus infection is through a cooperative interaction between a mannose receptor and CLEC5A on macrophage as a multivalent hetero-complex. *PLoS One*. 11:e0166474. doi: 10.1371/journal.pone.0166474

Lu, G., Wang, Q., and Gao, G. F. (2015). Bat-to-human: spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond. *Trends Microbiol.* 23, 468–478. doi: 10.1016/j.tim.2015.06.003

Mathew, S. M., Benslimane, F., Althani, A. A., and Yassine, H. M. (2021). Identification of potential natural inhibitors of the receptor-binding domain of the SARS-CoV-2 spike protein using a computational docking approach. *Qatar Med. J.* 2021:12. doi: 10.5339/qmj.2021.12

Matsuyama, S., Nagata, N., Shirato, K., Kawase, M., Takeda, M., and Taguchi, F. (2010). Efficient activation of the severe acute respiratory syndrome coronavirus spike protein by the transmembrane protease TMPRSS2. *J. Virol.* 84, 12658–12664. doi: 10.1128/JVI.01542-10

Matsuyama, S., Ujike, M., Morikawa, S., Tashiro, M., and Taguchi, F. (2005). Protease-mediated enhancement of severe acute respiratory syndrome coronavirus infection. *Proc. Natl. Acad. Sci. U. S. A.* 102, 12543–12547. doi: 10.1073/pnas.0503203102

Montecino-Rodriguez, E., Berent-Maoz, B., and Dorshkind, K. (2013). Causes, consequences, and reversal of immune system aging. *J. Clin. Invest.* 123, 958–965. doi: 10.1172/JCI64096

Mossel, E. C., Wang, J., Jeffers, S., Edeen, K. E., Wang, S., Cosgrove, G. P., et al. (2008). SARS-CoV replicates in primary human alveolar type II cell cultures but not in type I-like cells. *Virology* 372, 127–135. doi: 10.1016/j.virol.2007.09.045

Mu, A., Latario, C. J., Pickrell, L. E., and Higgs, H. N. (2020). Lysine acetylation of cytoskeletal proteins: emergence of an actin code. *J. Cell Biol.* 219:e202006151. doi: 10.1083/jcb.202006151

Nader, D., Fletcher, N., Curley, G. F., and Kerrigan, S. W. (2021). SARS-CoV-2 uses major endothelial integrin αvβ3 to cause vascular dysregulation in-vitro during COVID-19. *PLoS One* 16:e0253347. doi: 10.1371/journal.pone.0253347

Nader, D., and Kerrigan, S. W. (2022). Molecular cross-talk between Integrins and Cadherins leads to a loss of vascular barrier integrity during SARS-CoV-2 infection. *Viruses* 14:891. doi: 10.3390/v14050891

Naqvi, S., Godfrey, A. K., Hughes, J. F., Goodheart, M. L., Mitchell, R. N., and Page, D. C. (2019). Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. *Science* 365:eaaw7317. doi: 10.1126/science.aaw7317

Nguyen, L., McCord, K. A., Bui, D. T., Bouwman, K. M., Kitova, E. N., Elaish, M., et al. (2022). Sialic acid-containing glycolipids mediate binding and viral entry of SARS-CoV-2. *Nat. Chem. Biol.* 18, 81–90. doi: 10.1038/s41589-021-00924-1

Nowakowski, T. J., Pollen, A. A., Di Lullo, E., Sandoval-Espinosa, C., Bershteyn, M., and Kriegstein, A. R. (2016). Expression analysis highlights AXL as a candidate Zika virus entry receptor in neural stem cells. *Cell Stem Cell* 18, 591–596. doi: 10.1016/j.stem.2016.03.012

Othman, H., Messaoud, H. B., Khamessi, O., Ben-Mabrouk, H., Ghedira, K., Bharuthram, A., et al. (2022). SARS-CoV-2 spike protein unlikely to bind to Integrins via the Arg-Gly-asp (RGD) motif of the receptor binding domain: evidence from structural analysis and microscale accelerated molecular dynamics. *Front. Mol. Biosci.* 9:834857. doi: 10.3389/fmolb.2022.834857

Patel, V. B., Clarke, N., Wang, Z., Fan, D., Parajuli, N., Basu, R., et al. (2014). Angiotensin II induced proteolytic cleavage of myocardial ACE2 is mediated by TACE/ADAM-17: a positive feedback mechanism in the RAS. *J. Mol. Cell. Cardiol.* 66, 167–176. doi: 10.1016/j.yjmcc.2013.11.017

Peacock, T. P., Goldhill, D. H., Zhou, J., Baillon, L., Frise, R., Swann, O. C., 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. doi: 10.1038/s41564-021-00908-w

Plaas, M., Seppa, K., Gaur, N., Kasenõmm, P., and Plaas, M. (2021). Age- and airway disease related gene expression patterns of key SARS-CoV-2 entry factors in human nasal epithelia. *Virology* 561, 65–68. doi: 10.1016/j.virol.2021.05.012

Plein, A., Fantin, A., and Ruhrberg, C. (2014). Neuropilin regulation of angiogenesis, arteriogenesis, and vascular permeability. *Microcirculation* 21, 315–323. doi: 10.1111/micc.12124

Qiao, J., Li, W., Bao, J., Peng, Q., Wen, D., Wang, J., et al. (2020). The expression of SARS-CoV-2 receptor ACE2 and CD147, and protease TMPRSS2 in human and mouse brain cells and mouse brain tissues. *Biochem. Biophys. Res. Commun.* 533, 867–871. doi: 10.1016/j.bbrc.2020.09.042

Robles, J. P., Zamora, M., Adan-Castro, E., Siqueiros-Marquez, L., Martinez de la Escalera, G., and Clapp, C. (2022). The spike protein of SARS-CoV-2 induces endothelial inflammation through integrin α5β1 and NF-κB signaling. *J. Biol. Chem.* 298:101695. doi: 10.1016/j.jbc.2022.101695

Rubin, A. A. (1963). Coronary vascular effects of griseofulvin. *JAMA* 185, 971–972. doi: 10.1001/jama.1963.03060120081033

Sadoul, K., Wang, J., Diagouraga, B., and Khochbin, S. (2011). The tale of protein lysine acetylation in the cytoplasm. *J. Biomed. Biotechnol.* 2011, 1–15. doi: 10.1155/2011/970382

Scheller, J., Chalaris, A., Garbers, C., and Rose-John, S. (2011). ADAM17: a molecular switch to control inflammation and tissue regeneration. *Trends Immunol.* 32, 380–387. doi: 10.1016/j.it.2011.05.005

Sheppard, D. (1996). Epithelial integrins. *BioEssays* 18, 655–660. doi: 10.1002/bies.950180809

Shi, J., Wen, Z., Zhong, G., Yang, H., Wang, C., Huang, B., et al. (2020). Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. *Science* 368, 1016–1020. doi: 10.1126/science.abb7015

Soler, M. J., Wysocki, J., and Batlle, D. (2013). ACE2 alterations in kidney disease. *Nephrol. Dial. Transplant.* 28, 2687–2697. doi: 10.1093/ndt/gft320

Takada, Y., Ye, X., and Simon, S. (2007). The integrins. *Genome Biol.* 8:215. doi: 10.1186/gb-2007-8-5-215

Teesalu, T., Sugahara, K. N., Kotamraju, V. R., and Ruoslahti, E. (2009). C-end rule peptides mediate neuropilin-1-dependent cell, vascular, and tissue penetration. *Proc. Natl. Acad. Sci. U. S. A.* 106, 16157–16162. doi: 10.1073/pnas.0908201106

Tipnis, S. R., Hooper, N. M., Hyde, R., Karran, E., Christie, G., and Turner, A. J. (2000). A human homolog of angiotensin-converting enzyme: cloning and functional expression as a captopril-insensitive carboxypeptidase*. *J. Biol. Chem.* 275, 33238–33243. doi: 10.1074/jbc.M002615200

To, K. F., and Lo, A. W. (2004). Exploring the pathogenesis of severe acute respiratory syndrome (SARS): the tissue distribution of the coronavirus (SARS-CoV) and its putative receptor, angiotensin-converting enzyme 2 (ACE2). *J. Pathol.* 203, 740–743. doi: 10.1002/path.1597

To, K. F., Tong, J. H., Chan, P. K., Au, F. W., Chim, S. S., Chan, K. C., et al. (2004). Tissue and cellular tropism of the coronavirus associated with severe acute respiratory syndrome: an in-situ hybridization study of fatal cases. *J. Pathol.* 202, 157–163. doi: 10.1002/path.1510

Towler, P., Staker, B., Prasad, S. G., Menon, S., Tang, J., Parsons, T., et al. (2004). ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis. *J. Biol. Chem.* 279, 17996–18007. doi: 10.1074/jbc.M311191200

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. *Science* 347:1260419. doi: 10.1126/science.1260419

Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cells* 181, 281–292. doi: 10.1016/j.cell.2020.02.058

Wang, S., Guo, F., Liu, K., Wang, H., Rao, S., Yang, P., et al. (2008). Endocytosis of the receptor-binding domain of SARS-CoV spike protein together with virus receptor ACE2. *Virus Res.* 136, 8–15. doi: 10.1016/j.virusres.2008.03.004

Wang, Q., Zhang, Y., Wu, L., Niu, S., Song, C., Zhang, Z., et al. (2020). Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cells* 181, 894–904.e9. doi: 10.1016/j.cell.2020.03.045

Wei, Y., Aris, P., Farookhi, H., and Xia, X. (2021). Predicting mammalian species at risk of being infected by SARS-CoV-2 from an ACE2 perspective. *Sci. Rep.* 11:1702. doi: 10.1038/s41598-41020-80573-x

Wilen, C. B., Tilton, J. C., and Doms, R. W. (2012). HIV: cell binding and entry. *Cold Spring Harb. Perspect. Med.* 2:a006866. doi: 10.1101/cshperspect.a006866

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., et al. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367, 1260–1263. doi: 10.1126/science.abb2507

Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269. doi: 10.1038/s41586-020-2008-3

Wysocki, J., Ye, M., Soler, M. J., Gurley, S. B., Xiao, H. D., Bernstein, K. E., et al. (2006). ACE and ACE2 activity in diabetic mice. *Diabetes* 55, 2132–2139. doi: 10.2337/db06-0033

Xia, X. (2018a). "Bioinformatics and translation elongation" in *Bioinformatics and the cell: Modern computational approaches in genomics, proteomics and transcriptomics*, ed. X. Xia (Switzerland: Springer, Cham), 197–238.

Xia, X. (2018b). DAMBE7: new and improved tools for data analysis in molecular biology and evolution. *Mol. Biol. Evol.* 35, 1550–1552. doi: 10.1093/molbev/msy073

Xia, X. (2021). Domains and functions of spike protein in SARS-COV-2 in the context of vaccine design. *Viruses* 13:109. doi: 10.3390/v13010109

Xu, C., Wang, Y., Liu, C., Zhang, C., Han, W., Hong, X., et al. (2021). Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. *Science Advances* 7:abe5575. doi: 10.1126/sciadv.abe5575

Xu, H., Zhong, L., Deng, J., Peng, J., Dan, H., Zeng, X., et al. (2020). High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. *Int. J. Oral Sci.* 12:8. doi: 10.1038/s41368-020-0074-x

Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., and Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 367, 1444–1448. doi: 10.1126/science.abb2762

Yan, Y., Zhou, A., Carrell, R. W., and Read, R. J. (2019). Structural basis for the specificity of renin-mediated angiotensinogen cleavage. *J. Biol. Chem.* 294, 2353–2364. doi: 10.1074/jbc.RA118.006608

Yang, C., Zhang, Y., Zeng, X., Chen, H., Chen, Y., Yang, D., et al. (2021). Kidney injury molecule-1 is a potential receptor for SARS-CoV-2. *J. Mol. Cell Biol.* 13, 185–196. doi: 10.1093/jmcb/mjab003

Zech, F., Schniertshauer, D., Jung, C., Herrmann, A., Cordsmeier, A., Xie, Q., et al. (2021). Spike residue 403 affects binding of coronavirus spikes to human ACE2. *Nat. Commun.* 12:6855. doi: 10.1038/s41467-021-27180-0

Zhai, X., Sun, J., Yan, Z., Zhang, J., Zhao, J., Zhao, Z., et al. (2020). Comparison of severe acute respiratory syndrome coronavirus 2 spike protein binding to ACE2 receptors from human, pets, farm animals, and putative intermediate hosts. *J. Virol.* 94:e00831-20. doi: 10.1128/JVI.00831-20

Zhao, Y., Zhao, Z., Wang, Y., Zhou, Y., Ma, Y., and Zuo, W. (2020). Single-cell RNA expression profiling of ACE2, the receptor of SARS-CoV-2. *Am. J. Respir. Crit. Care Med.* 202, 756–759. doi: 10.1164/rccm.202001-0179LE

Zheng, M. (2022). ACE2 and COVID-19 susceptibility and severity. *Aging Dis.* 13, 360–372. doi: 10.14336/AD.2021.0805

Zhou, T., Tsybovsky, Y., Gorman, J., Rapp, M., Cerutti, G., Chuang, G. Y., et al. (2020). Cryo-EM structures of SARS-CoV-2 spike without and with ACE2 reveal a pH-dependent switch to mediate endosomal positioning of receptor-binding domains. *Cell Host Microbe* 28, 867–879.e5. doi: 10.1016/j.chom.2020.11.004

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273. doi: 10.1038/s41586-020-2012-7

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 26 July 2023
DOI 10.3389/fmicb.2023.1165839

Identification of two novel papillomaviruses in belugas

Youyou Li[1†], Meifang Xiao[2†], Yun Zhang[1†], Zihan Li[1†], Shijie Bai[3], Haoxiang Su[4], Ruoyan Peng[1], Gaoyu Wang[1], Xiaoyuan Hu[1], Xinran Song[5], Xin Li[6], Chuanning Tang[1], Gang Lu[1], Feifei Yin[1,2]*, Peijun Zhang[3]* and Jiang Du[1,4]*

[1]Hainan Medical University-The University of Hong Kong Joint Laboratory of Tropical Infectious Diseases, Key Laboratory of Tropical Translational Medicine of Ministry of Education, Hainan Medical University, Haikou, China, [2]Department of Clinical Laboratory, Center for Laboratory Medicine, Hainan Women and Children's Medical Center, Haikou, China, [3]Marine Mammal and Marine Bioacoustics Laboratory, Laboratory of Marine Viruses and Molecular Biology, Institute of Deep-sea Science and Engineering, Chinese Academy of Sciences, Sanya, China, [4]National Health Commission, Key Laboratory of Systems Biology of Pathogens, Institute of Pathogen Biology, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [5]Dalian Sun Asia Tourism Holding Co., Ltd., Dalian, China, [6]Qingdao Polar Haichang Ocean Park, Qingdao, China

Introduction: Papillomaviruses (PVs) can cause hyperplasia in the skin and mucous membranes of humans, mammals, and non-mammalian animals, and are a significant risk factor for cervical and genital cancers.

Methods: Using next-generation sequencing (NGS), we identified two novel strains of papillomavirus, PV-HMU-1 and PV-HMU-2, in swabs taken from belugas (*Delphinapterus leucas*) at Polar Ocean Parks in Qingdao and Dalian.

Results: We amplified the complete genomes of both strains and screened ten belugas and one false killer whale (*Pseudorca crassidens*) for the late gene (L1) to determine the infection rate. In Qingdao, 50% of the two sampled belugas were infected with PV-HMU-1, while the false killer whale was negative. In Dalian, 71% of the eight sampled belugas were infected with PV-HMU-2. In their L1 genes, PV-HMU-1 and PV-HMU-2 showed 64.99 and 68.12% amino acid identity, respectively, with other members of *Papillomaviridae*. Phylogenetic analysis of combinatorial amino acid sequences revealed that PV-HMU-1 and PV-HMU-2 clustered with other known dolphin PVs but formed distinct branches. PVs carried by belugas were proposed as novel species under *Firstpapillomavirinae*.

Conclusion: The discovery of these two novel PVs enhances our understanding of the genetic diversity of papillomaviruses and their impact on the beluga population.

KEYWORDS

diversity, novel PVs, evolution, beluga whale, next-generation sequencing

EDITED BY
Mohamad S. Hakim,
Gadjah Mada University, Indonesia

REVIEWED BY
Marta Canuti,
University of Milan, Italy
Shu Shen,
Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE
Jiang Du
✉ dujiangemail@163.com
Peijun Zhang
✉ pjzhang@idsse.ac.cn
Feifei Yin
✉ yinfeifeiff@163.com

†These authors have contributed equally to this work

RECEIVED 14 February 2023
ACCEPTED 03 July 2023
PUBLISHED 26 July 2023

CITATION
Li Y, Xiao M, Zhang Y, Li Z, Bai S, Su H, Peng R, Wang G, Hu X, Song X, Li X, Tang C, Lu G, Yin F, Zhang P and Du J (2023) Identification of two novel papillomaviruses in belugas. *Front. Microbiol.* 14:1165839. doi: 10.3389/fmicb.2023.1165839

1. Introduction

Papillomavirus (PV) can cause persistent infections in various parts of the skin and mucous membranes in humans and mammals, resulting in epidermal proliferative lesions (Cheng et al., 2023; Yang et al., 2023). Human genital PV (HPV) types are primarily responsible for cervical, vulvar, penile, and other anogenital cancers (Pino et al., 2023; Seyoum et al., 2023). Although papillomaviruses have been detected in all vertebrate populations except amphibians, most of them have been identified in humans and other mammals (Van Doorslaer et al., 2017). However, novel papillomaviruses have been recently detected in the cartilage of *Fulmarus glacialis* (Gaynor et al., 2015) and the liver of *Hemidactylus frenatus* (Agius et al., 2019) using next-generation sequencing (NGS). The increase in the number of viruses is associated with viruses that are typically found in terrestrial mammals, being found in marine mammals as well. For instance, Mäkelä et al. (2015) found respiratory diseases in marine mammals caused by influenza A and B virus infections (Mäkelä et al., 2015). Fouchier et al. (2004) found that H7N7 infected seals, which then infected human beings causing conjunctivitis (Fouchier et al., 2004). In addition, the US Centers for Disease Control and Prevention revealed that from 2007 to 2009, various belugas in North America were infected with the seal pox virus, which caused various diseases (Roess et al., 2011) and the individuals who were in close contact with belugas also tested positive for the virus. A study at the University of Florida found that the San Miguel sea lion virus (SMSV) was indistinguishable from the vesicular exanthema of swine virus (VESV), which caused vesicular inflammation in California pigs (Waltzek et al., 2012). These studies suggest the possibility of zoonotic viruses being carried by marine mammals and the risk of cross-species transmission. Cui et al. (2019) and Zhou et al. (2020) found that severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), two highly infectious and pathogenic viruses, might have originated from bats. The genetic recombination ability of coronaviruses makes them more adaptable to their hosts and increases their survival (Su et al., 2016). Therefore, the growing global virus diversity and potential spillovers require attention (Cui et al., 2019).

Papillomaviruses (PVs) are a diverse group of small, non-enveloped viruses belonging to the family *Papillomaviridae*, with double-stranded DNA genomes ranging in length from 5 to 8 kb (Zhang et al., 2022). The structurally conserved circular genome of PVs is organized into several partially overlapping open reading frames (ORFs) (Arman and Munger, 2022). These include the major genes present in all PVs: the early genes (E1 and E2), which are responsible for replication (Kraberger et al., 2022; Prabhakar et al., 2023), regulation of the life cycle, viral gene expression. The late genes (L1 and L2) are controlling the formation of viral capsids. Many PVs also carry other early genes, viral oncogenes E6 and E7 that are responsible for their ability to interfere with tumor suppressor genes such as p53 and the retinoblastoma protein (pRB) family (Nadile et al., 2022; Sofiani et al., 2023; Zhi et al., 2023). The upstream regulatory region or long control region (URR or LCR) located between the L1 and E6 ORFs is responsible for the origin of virus replication, as well as virus and cell transcription factor binding sites (TFBS) (Warburton et al., 2018; Coursey et al., 2021).

While the understanding of non-human PV types has increased in recent years, knowledge about the diversity of animal PVs remains limited (Bernard et al., 2010). The discovery of bovine and human papillomavirus sequences in cats suggests the possibility of cross-species transmission (Egberink et al., 2013). Recently, research on invertebrates and vertebrates in marine ecosystems has focused on the origin, evolution, and emergence of terrestrial viruses. Papillomaviruses were first reported in porpoises (*Phocoena phocoena*) (Gottschling et al., 2011) and subsequently found in many other small Odontoceti, including common dolphins (*Delphinus delphis*) (Robles-Sikisaka et al., 2012), common bottlenose dolphins (*Tursiops truncatus*) (Bossart et al., 2019), and burmester's porpoises (*Phocoena spinipinnis*) (Van Bressem et al., 2007; Brimer et al., 2017). In order to expand our understanding of the diversity of marine animal PVs, we conducted a marine mammal virus survey program and discovered two novel papillomaviruses (PV-HMU-1 and PV-HMU-2) from nose, throat, and anal swab samples of belugas (*Delphinapterus leucas*). We obtained and characterized the complete genome sequences of PV-HMU-1 and PV-HMU-2. The alignment results of the novel PVs revealed low amino acid identities with other known papillomaviruses and clustered with dolphin papillomaviruses in the evolutionary tree. The findings of this study deepen our understanding of the host and genetic diversity of papillomaviruses and provide essential data for the possible zoonotic transmission caused by animal-origin PVs.

2. Materials and methods

2.1. Swab sample collection

We collected a total of 21 nose, throat, and anal swabs from ten belugas and one false killer whale (*Pseudorca crassidens*), housed in aquaria at Qingdao Polar Haichang Ocean Park and Dalian Sun Asia Tourism Holding Co., China in April 2018 (**Supplementary Table 1**). The nine throat, nose, and anal swabs from mammals in Qingdao constituted pool 49 (**Table 1**), while the 12 throat and anal swabs from Dalian constituted pool 41 (**Table 2**). The sampling procedures were approved by the Ethics Committee of the Hainan Medical University (Approval number: HMUEC20180059). The collected samples were quickly immersed in the maintenance medium in virus-sampling tubes (Yocon Biology, Beijing, China), to ensure sample quality, and transported to the laboratory within 24 h using an ice-cold dry chain (Wu et al., 2018a). The samples were then stored at −80°C (Wu et al., 2018a).

2.2. Viral nucleic acid library construction

Samples from each species were combined by adding 1 mL from each sample into one fresh sample tube. Based on the swab sample locations, the 21 samples were divided into two pools and passed through 0.45 μm filters (Millipore Sigma, Burlington, MA, USA) to remove eukaryotic cells and large bacteria. The filtrate was then ultracentrifuged at 100,000 × *g* at 4°C for 3 h. The precipitates collected from the two pool samples were resuspended in 100 μL Hank's balanced salt solution and digested using DNase, which consisted of 14 U of TURBO DNase (Ambion, USA), 20 U of benzonase (Novagen, Germany), and 20 U of RNase One (Promega, USA) at 37°C for 2 h in 1 × DNase buffer to decompose and remove unprotected nucleic acids. Viral DNA and RNA were extracted using QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. First-strand cDNA was generated using Superscript III Reverse Transcriptase (Invitrogen, Thermo Fisher Scientific), as previously described (Wu et al., 2018b). Double-stranded DNA was synthesized using Klenow fragment enzyme combined with K-8N (GACCATCTAGCGACCTCCACNNNNNNNN). To eliminate excess enzymes and PCR product primers, a PCR purification kit

TABLE 1 Detection rate of PV-HMU-1 among belugas from Qingdao Ocean Park.

Numbers	Animal species	Throat swab	Nose swab	Anal swabs	Sampling date	Sampling place
1	*Delphinapterus leucas*	+	–	–	19 April, 2018	Qingdao
2	*D. leucas*	–	–	–	19 April, 2018	Qingdao
3	*Pseudorca crassidens*	–	–	–	19 April, 2018	Qingdao

TABLE 2 Detection rate of PV-HMU-2 among belugas from Dalian Ocean Park.

Numbers	Animal species	Throat swab	Anal swabs	Sampling date	Sampling place
1	*Delphinapterus leucas*	Null	–	18 April, 2018	Dalian
2	*D. leucas*	+	Null	18 April, 2018	Dalian
3	*D. leucas*	+	–	18 April, 2018	Dalian
4	*D. leucas*	–	Null	18 April, 2018	Dalian
5	*D. leucas*	–	–	18 April, 2018	Dalian
6	*D. leucas*	+	–	18 April, 2018	Dalian
7	*D. leucas*	+	Null	18 April, 2018	Dalian
8	*D. leucas*	+	–	18 April, 2018	Dalian

was used, followed by the use of magnetic beads to adsorb 300–2,000 bp fragments and dissolve them in nuclease free water. The nucleic acid library was then constructed using the DNA Library Prep Kit (Invitrogen Collibri, USA).

2.3. Next-generation sequencing

The amplified viral nucleic acid libraries were sequenced with an Illumina HiSeq 2500 sequencer using the 150 bp paired-end method (Illumina Inc., San Diego, CA, USA). To maximize the available length and total output of the raw data, each pool sample library was sequenced with separate channels and indexing. The raw sequence reads were filtered, and valid sequences were obtained using previously described criteria. The initial image analysis and base calling were performed using the GAPipeline program with default parameters. Sequences that include the primer K and adaptor are extracted using the Illumina filter (Wu et al., 2012). The sequence data were deposited into the National Center for Biotechnology Information sequence read archive under the accession number PRJNA650224.

2.4. Taxonomic assignment

Raw sequence reads were filtered to obtain valid sequences using previously described criteria (Yang et al., 2011). Assembly of quality-controlled reads of each sample into contigs was performed with Trinity (Version 2.5.1). Diamond (Version 0.9.14.115) software was used to translate contigs into amino acid sequences. Sequence similarity-based taxonomic assignments were conducted as described previously (Yang et al., 2011). Briefly, each sequence was aligned with the NCBI non-redundant protein databases (NR) using the basic local alignment search tool (Blast)[1] Blast X (E-value $< 10^{-5}$, −F: filter query sequence, default = T) to determine its viral origin. All Blast X results of contigs were annotated with taxonomy IDs MegaHit and other information obtained from the NCBI Entrez server (Tamura et al., 2013).

2.5. Genome sequencing of papillomavirus

Molecular clues obtained from metagenomic analyses were utilized to classify sequence reads into viral families or genera using MEGAN.[2] Representative reads of novel papillomaviruses were selected for PCR amplification and sequencing verification, and the complete novel PVs sequences were amplified. To amplify and sequence parts of the genome, specific nested PCR primers were designed. cDNA was generated using Superscript III Reverse Transcriptase (Invitrogen/United States) and random primers. All primer sequences were based on newly obtained reads and amplified sequences, and the primer sequences used are shown in **Supplementary Table 2**.

2.6. Genome annotation

The viral nucleotide sequences of the genomes and the amino acid sequences of the ORFs were deduced by comparing them against those of other viral families. Conserved protein families and domains were predicted using Pfam,[3] Blast p (see text footnote 1), and InterProScan 5 (see text footnote 3). The genome structure diagram was constructed using IBS 1.0.3 software, and routine sequence alignment was performed using Clustal Omega.[4] TFBS was predicted with "BDGP"[5] and "Softberry."[6]

---

[1] https://blast.ncbi.nlm.nih.gov

[2] https://megasoftware.net/show_eua

[3] http://www.ebi.ac.uk/services/proteins

[4] http://www.ebi.ac.uk/tools/

[5] http://www.fruitfly.org/seq_tools/promoter.html

[6] http://www.softberry.com/

2.7. Calculation of viral prevalence

Specific primers targeting the non-structural gene were designed based on the partial viral genomic sequences obtained from NGS for PV screening in each individual sample (**Supplementary Table 2**). PCR was performed using GoTaq Colorless Master Mix (Promega, Madison, WI, USA). The first-round PCR product (2 μl) as the template for the second round of PCR was used to achieve high specificity and sensitivity. The thermal cycling conditions for PCRs were as follows: 94°C for 5 min, followed by 35 cycles of 94°C for 30 s, 57°C for 35 s, 72°C for 30 s, and a final elongation step at 72°C for 10 min. The PCR products were analyzed using 1.5% agarose gel electrophoresis and ultraviolet imaging (Bio-Rad, USA).

2.8. Phylogenetic analysis

An amino acid alignment was performed between the novel beluga PV and other reference sequences using the Blast tool on the NCBI to identify closely related sequences. The phylogenetic tree included both mammalian and non-mammalian papillomaviruses, and representative species from each genus were selected as reference sequences. According to a previous study, each PV reference sequence is concatenated independently (Agius et al., 2019). A total of 61 PVs reference sequences were combined with the novel PVs sequences in this study to construct a phylogenetic analysis tree. MEGAX was used to align the amino acid sequences using the MUSCLE package with default parameters. The phylogenetic tree was constructed using the maximum likelihood method. The best substitution models were the Poisson model with uniform rates among sites, and this model was replicated with 1,000 bootstrap replicates to construct the phylogenetic tree.

3. Results

3.1. Discovery of two novel PVs in belugas using NGS

In this study, we obtained 2.74 GB (214,466 valid reads, 150 bp in length) and 1.39 GB (102,834 valid reads, 150 bp in length) of nucleotide data from sample pools 41 and 49, respectively. Sequences of less than 50 bp and duplicate sequences were eliminated. The archaea, bacteria, microbial eukaryotes, and those that had no significant similarity with any amino acid sequences in the viral NR protein database were eliminated. In pool 49, 14,055 virus-associated reads, accounting for 6.55% of the total reads, were identified, while in pool sample 41, 2,404 virus-associated reads, accounting for 2.33% of the total reads were identified. The viromes of both pools consisted of double-stranded (ds) DNA, single-stranded (ss) DNA, ssRNA, and retro-transcribing viruses. The virus-associated reads of two pool samples were assigned to families such as *Picornaviridae*, *Herpesviridae*, *Myoviridae*, *Podoviridae*, *Siphoviridae*, *Microviridae*, *Retroviridae*, and unclassified viruses (**Figure 1**). Moreover, swab sample pools from Qingdao and Dalian contained 418 and 54 papillomavirus reads, respectively, with PVs accounting for 2.97 and 2.24% of the total viruses in pools 49 and 41, respectively. After PCR verification, two novel PV sequences were obtained and compared with the original sequencing PV reads, which showed that the PV reads in the metagenomic sequencing results belonged to the two novel PVs. In addition, molecular clues related to mammalian *Picornaviridae* and *Herpesviridae* were found in the whale swab samples from Dalian, with 491 reads related to herpesviruses showing less than 50.68% aa identity with other known herpesviruses and being related to Beluga alpha herpes virus 1. Furthermore, the sequencing results showed 56 reads related to picornavirus, with less than 40.24% aa identity with other known picornaviruses, and verification of this virus was completed by our laboratory (Wang et al., 2021).

3.2. Genomic characterization of novel PVs

Specific primers were designed based on the NGS-generated fragment sequences to obtain the complete genome sequences of two novel PVs from samples obtained from belugas in Qingdao and Dalian Ocean Parks. The novel PVs were named PV-HMU-1 (GenBank accession: OP748391) and PV-HMU-2 (GenBank accession: OP748392), respectively. The complete genome length of PV-HMU-1 and PV-HMU-2 were 7,624 and 7,179 bp, respectively, with C + G contents of 58.77 and 54.37% (**Figure 2** and **Supplementary Table 1**). The Pfam software predicted that the observed mammalian PV genomic structures included a core set of early (E6, E1, and E2) and late (L2 and L1) open-reading frames (ORFs) (**Figure 2**). ORF prediction software revealed that PV-HMU-1 E6 (184 aa) was 552 nt (1–552), E1 (625 aa) was 1,875 nt (687–2,561), early E2 (393 aa) was 1,179 nt (2,506–3,684), late L2 (535 aa) was 1,605 nt (3,974–5,578), and late L1 (582 aa) was 1,746 nt (5,337–7,082) in length. For PV-HMU-2, E6 (233 aa) was 699 nt (1–699), E1 (622 aa) was 1,866 nt (755–2,620), early E2 (372 aa) was 1,116 nt (2,556–3,671), late L2 (450 aa) was 1,350 nt (3,787–5,136), and late L1 (496 aa) was 1,488 nt (5,120–6,607) in length.

A study by Schulz et al. (2012) used TFBS prediction to identify potential *cis*-regulatory elements in the upstream regulatory region (URR) of rodent PVs. The URR sequences of PV-HMU-1 contained important TFBS, such as activator protein 1 (AP1), which were located directly upstream of the TFIID BS (TATA box binding protein). However, the TFBS of PV-HMU-2 was not clearly predicted. Both PV-HMU-1 and PV-HMU-2 included a zinc-finger motif (C-X2-C-X29-C-X2-C) in E6, while E7 was absent (**Figure 2**). An adenosine triphosphate/guanosine triphosphate (ATP/GTP) binding site (G-X4-GKS) was present within the helicase domain of E1 in both sequences. The N-terminal, DNA-binding domain, and helicase were located at positions 689–1,058 bp| 755–1,124 bp, 1,235–1,652 bp| 1,273–1,700 bp, and 1,707–2,504 bp| 1,712–2,589 bp of the E1 region for PV-HMU-1 and PV-HMU-2, respectively. Comparing the L2 and L1 amino acid sequences of PV-HMU-1 and PV-HMU-2 with those of other papillomaviruses, PV-HMU-1 showed 41.97 and 64.99%

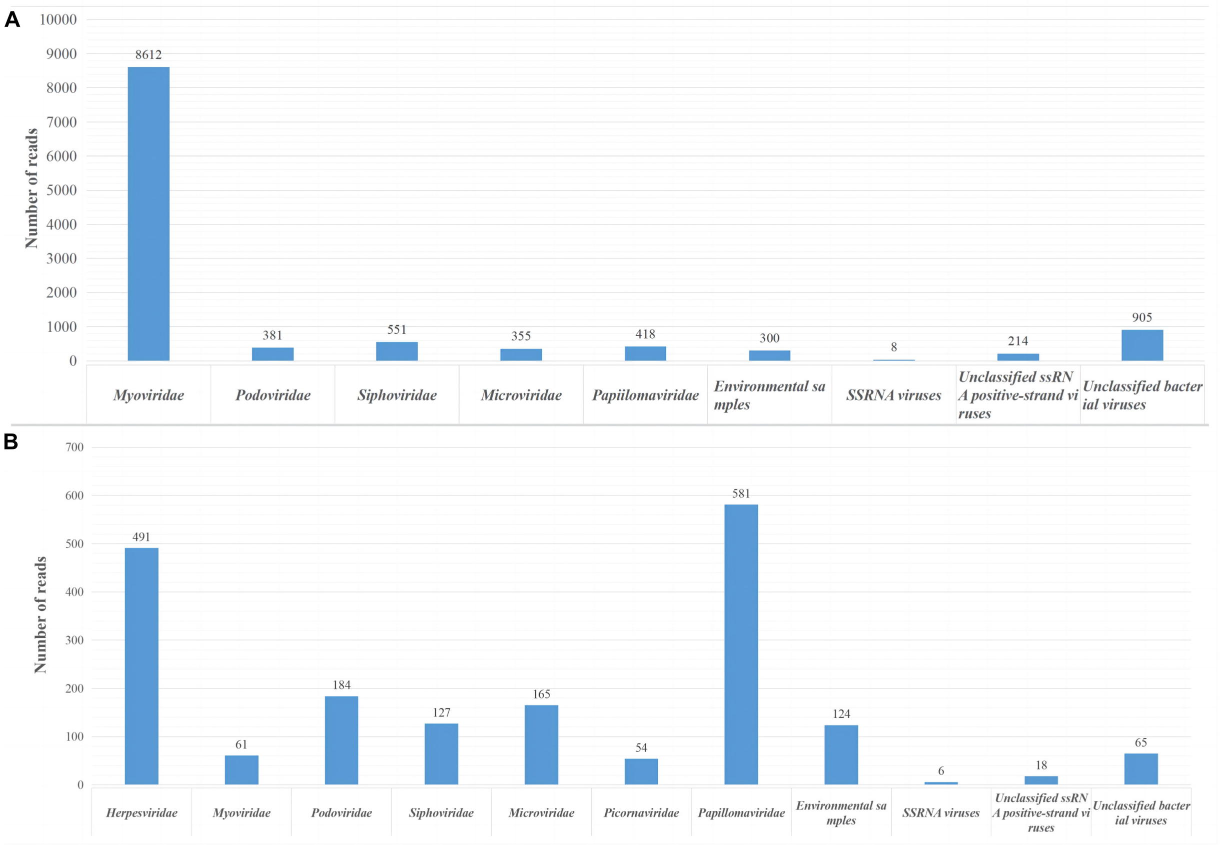


FIGURE 1
Viral reads from two sample pools using next-generation sequencing (NGS). **(A)** Identification of viral groups in swab samples from Qingdao belugas and false killer whale. **(B)** Viral content in swab samples from Dalian belugas.

identity to *Phocoena* papillomavirus 4 (NC 018076), respectively, while PV-HMU-2 showed 49.19 and 68.12% identity to Sus scrofa papillomavirus 1 (MT078972), respectively.

3.3. Phylogenetic analyses

Phylogenetic analyses were conducted using the complete E1-E2-L2-L1 amino acid sequences of PV-HMU-1 and PV-HMU-2, as well as 61 representative references related to *Papillomaviridae* in GenBank (**Figure 3**). Based on the results of Blast and preliminary evolutionary analysis, the novel PVs were attributed to the mammalian PV clade, so we mainly selected mammalian PVs as references, with a few Sauropsid and Fish PV clades selected as out-groups. More than 50 genera of PVs existed, and we selected the representative reference of each genus and the virus strains with the highest identities relative to the novel PVs to construct the evolutionary tree. The phylogenetic analysis showed that the amino acid combination of PV-HMU-1 clustered with *Tursiops truncatus* papillomavirus 2 (TtPV2) (NC 008184) and formed a relatively independent clade within the genus *Upsilonpapillomavirus*. PV-HMU-2 clustered with *Phocoena phocoena* papillomavirus 4 (PphPV4) (GU117623), the only species of *Dyopipapillomavirus*. Sus scrofa papillomavirus 1 is a member of the genus *Dyoomikronpapillomavirus* and was clustered with PV-HMU-2. PV-HMU-1 and PV-HMU-2 did not cluster on the same branch on the evolutionary tree and were far from human-associated papillomaviruses. The genome sequences of PV-HMU-1 and PV-HMU-2 carried by belugas were difficult to compare with those of other known papillomaviruses at the nucleic acid level in the NCBI nucleotide database. Genome identity and phylogenetic analyses suggested that these beluga PVs represented novel species.

3.4. Prevalence of PV infection in belugas

Gene-specific nested primers targeting the L1 gene region for PCR were designed based on the genomic sequences of novel PVs to screen for their presence in the throat, nose, and anal swabs. The throat and nose swabs of belugas in both locations showed a high infection rate of PVs, while the infection rate in anal swabs was low. Thirty-three percent (1/3) of throat swabs and nose swabs, but no anal swabs were positive for PV-HMU-1 (**Table 1**). Seventy-one percent (5/7) of throat swabs but no anal swabs were positive for PV-HMU-2 (**Table 2**). All samples collected in Qingdao and Dalian were collected three times in April, September, and December 2018 to observe the virus-carrying rate in different quarters. The throat swabs of three animals in Dalian consistently tested positive for PV-HMU-2 at all three sampling

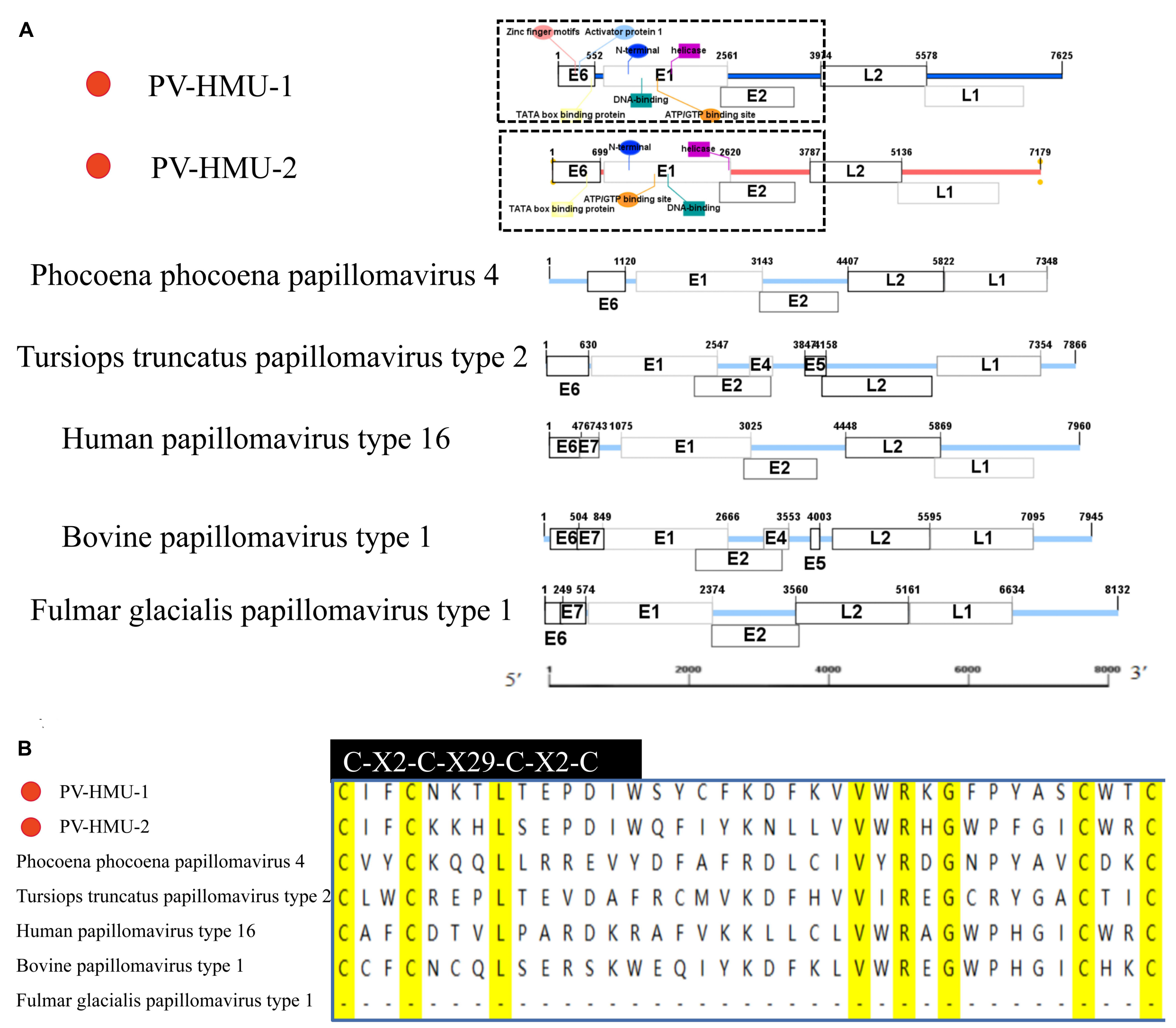


FIGURE 2

Sequence and motif annotation of PV-HMU-1 and PV-HMU-2. **(A)** Putative and confirmed early and late ORFs identified in PV-HMU-1 and PV-HMU-2 genomes using ORF predictor. We used IBS 1.0.3 software to construct the genomic structure diagram of PV-HMU-1, PV-HMU-2, *Phocoena phocoena* papillomavirus 4 (PphPV4) and *Lagenorhynchus acutus* papillomavirus (GU117624). **(B)** The E6 conserved protein domains of PV-HMU-1 and PV-HMU2 (light blue shade) was aligned with other papillomaviruses. The C-terminal zinc-binding domains for PV-HMU-1, PV-HMU-2, HPV16, BPV1, TtPV2, and PphPV4 are indicated.

times, indicating persistent viral presence. To confirm the presence of the same virus infection in different months, a set of L primers was used to screen the samples. After amplification, fragments of appropriate size were obtained and sent for sequencing. The sequencing results of two positive samples collected at different times showed an aa identity of > 99%. The positive samples were identified by amplification products of 540 bp for PV-HMU-1 and 522 bp for PV-HMU-2.

4. Discussion

The increasing contact between humans and marine mammals due to the exploration of the oceans and interest in these animals has led to a higher risk of zoonotic diseases that can endanger human health. While efforts have been made to characterize novel PV types in different animal hosts, knowledge about PV diversity in marine mammals remains limited. In recent years, many viruses, including picornaviruses, coronaviruses, morbilliviruses, and poxviruses, have been detected in dolphins (Robles-Sikisaka et al., 2012), seals (Fouchier et al., 2004), and sea lions (Waltzek et al., 2012). Due to the continuous expansion of the host range of viruses in marine mammals, it is suggested that they may serve as important hosts for many viruses, and there may still be many unknown viruses in these hosts. Reports of marine mammals acting as reservoirs for viruses are increasing. In Groth et al. (2014) obtained a complete sequence of influenza A virus (H13N2) from a sick pilot whale, while Sacristán et al. (2019) and Vargas-Castro et al. (2021) found that dolphins in the South Atlantic and whales in Spain were highly infected with herpesvirus (Sacristán et al., 2019; Vargas-Castro et al., 2021). Beluga coronavirus was detected in a study at the Washington University School of

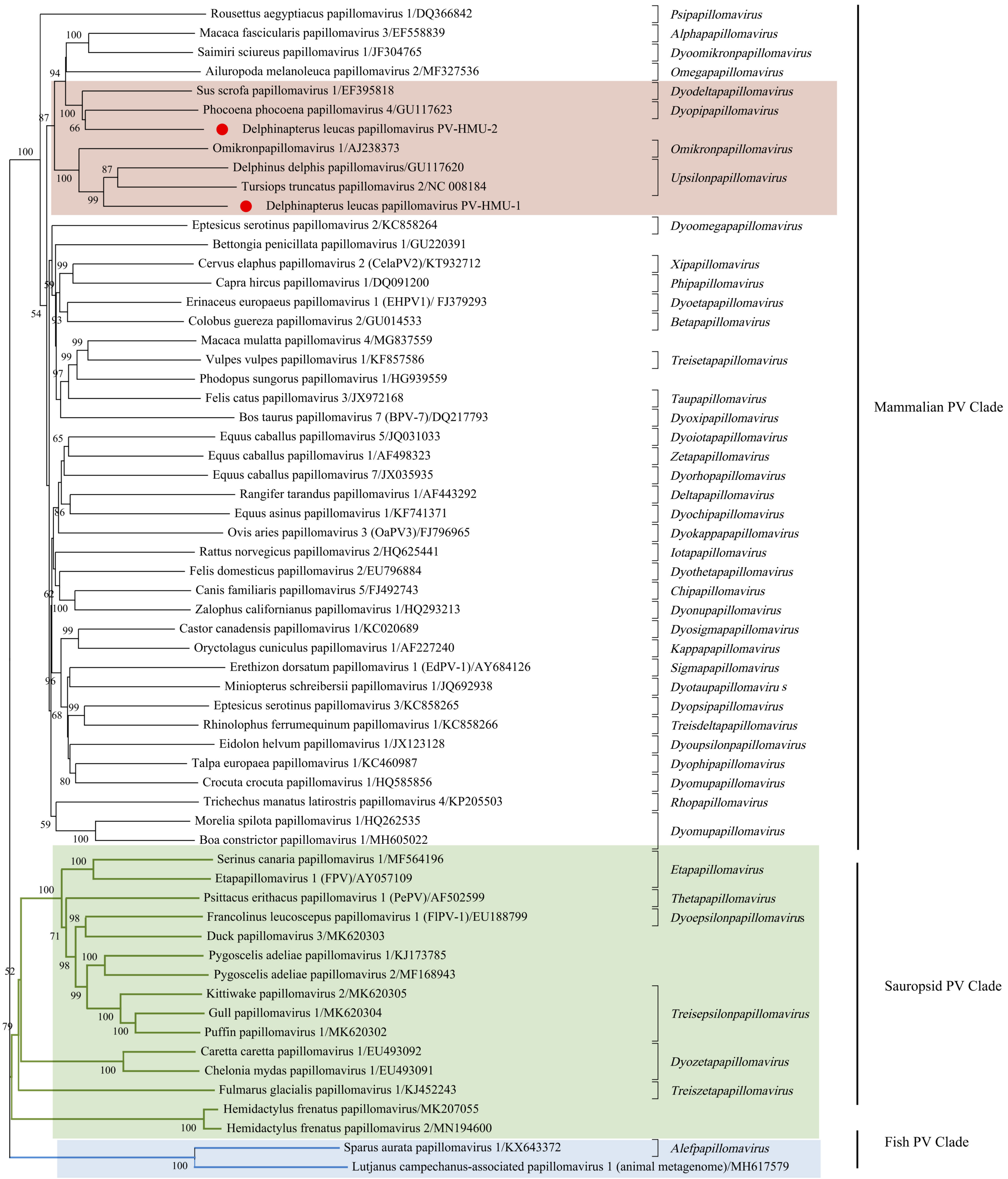


FIGURE 3
Phylogenetic tree of concatenated early (E1, E2) and late (L1) papillomavirus proteins. The evolutionary history was inferred from 61 amino acid sequences, including the novel PV-HMU-1 and PV-HMU-2, using a maximum likelihood approach. A 1,000 bootstrap replication were used to construct the tree. PV-HMU-1 and PV-HMU-2 are marked with red circles under mammalian-related PVs. The virus strains close to the novel identified PVs are marked with pink background, Sauropsid PV, and Fish PV clades are marked with green and blue backgrounds, respectively.

Medicine (Mihindukulasuriya et al., 2008), and a novel bottlenose dolphin coronavirus was detected at the University of Hong Kong (Woo et al., 2014). Changes in living habits, social behaviors, and migration patterns are significant factors in the transmission and spread of viruses among cetaceans. Monitoring the viruses carried by marine mammals is crucial for preventing potential infectious diseases of cetacean origin. In this study, we detected PV-HMU-1 and PV-HMU-2 in belugas from Qingdao and the Dalian Polar Ocean Parks and obtained their complete genomes. The genomes of PV-HMU-1 and PV-HMU-2 contain five ORFs (E6, E1, E2, L2, and L1) of mammal PVs but do not have the E7 ORF typical of *Alphapapillomavirus* (**Figure 2**). The E6 protein of the novel virus contains a zinc-finger motif that is crucial for the virus to invade the cell environment. AP1 is present at the PV-HMU-1 TFBS site and is used to regulate α-HPV expression. The two novel PVs contain an ATP/GTP binding site, an N-terminal, DNA binding region, and a helix in the E1 region. Papillomaviruses are generally transmitted through sex, close contact, and mother-to-child transmission. The 50% infection rate of PV-HMU-2 infection in individual samples reflected the possibility of a high incidence of PV infection in belugas. Belugas that tested positive for PV-HMU-1 and PV-HMU-2 were positive three times within a year, indicating that the infection was persistent. It is important to note that some of the negative results may be due to inadequate sampling.

In 2019, the ICTV report on the classification of genera highlighted that, based on multiple sequences or comparative pairs of L1 genes, most PV genus types have less than 60% sequence identity with other types, but this is not the only criterion (de Villiers et al., 2004). The L gene aa identity between PV-HMU-1 and PV-HMU-2 was 65.21%, although they were found in the same beluga host. The two novel PVs showed less than 68.12% aa identity with the L gene region of known papillomaviruses. The new viruses mainly cluster with viruses whose hosts are small toothed whales in the phylogenetic tree, which are closely related to *Upsilonpapillomavirus* and *Dyopipapillomavirus*. The clustering of PV-HMU-2 and pig PV suggests that the evolution of PV in belugas and terrestrial animals is related.

Several phylogenetic studies have shown that a complex evolutionary scenario drives PV diversity, such as recombination, co-divergence with the host, interspecies transmission, and adaptation to new environments (Geoghegan et al., 2017). Although the belugas in this study did not exhibit obvious tumors or other clinical manifestations, the high infection rate suggests a possible association between papillomavirus infection and belugas. While this study amplified the complete genome sequences of two PVs and investigated their classification, it is merely a small step toward understanding the complexities of papillomaviruses in marine mammal populations. More work is necessary to examine the evolution, transmission, and biological role of papillomaviruses in marine mammal populations, isolate and record them, and provide fundamental data for future research on the evolution, classification, and traceability of mammalian viruses.

In this study, we reported the discovery of two novel papillomaviruses in the throat and nasal swabs collected from belugas in Qingdao and Dalian. To our knowledge, belugas have never been reported as hosts for papillomaviruses before. These findings increase our understanding of the mammal-borne virus community in Qingdao and Dalian and provide preliminary data for future research on the evolution of PVs. However, further investigation is necessary to determine the pathogenicity of these two novel viruses and their potential impact on both humans and animals.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in this article/**Supplementary material**.

Author contributions

FY, JD, and PZ designed the study. PZ, SB, XS, and XL collected the specimens. YL, ZL, MX, YZ, GW, RP, HS, and XH performed the experiments. CT and GL analyzed the data. YL wrote the draft manuscript. JD and PZ edited the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (No. 32060015), the National Key R&D Program of China (grant No. 2022YFE0210300), the Key Research and Development Plan of Hainan Province (Nos. ZDYF2020150 and ZDYF2022SHFZ085), the CAMS Innovation Fund for Medical Sciences (CIFMS) (grant No. 2021-I2M-1-039), the Major Science and Technology Program of Hainan Province (No. ZDKJ202003), the Hainan Provincial Natural Science Foundation of China (Nos. 820RC628 and 823RC598), the Research Project of the Hainan Academician Innovation Platform (No. YSPTZX202004), the Hainan Talent Development Project (No. SRC200003), and the Graduate student innovation grant of Hainan Medical University (Nos. HYYS2020-33, Hys2020-360, Hys2020-376, and Qhys2021-356).

Conflict of interest

XS was employed by Dalian Sun Asia Tourism Holding Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer SS declared a past co-authorship with SB and PZ to the handling editor MH.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1165839/full#supplementary-material

References

Agius, J. E., Phalen, D. N., Rose, K., and Eden, J. S. (2019). New insights into Sauropsid Papillomaviridae evolution and epizootiology: Discovery of two novel papillomaviruses in native and invasive Island geckos. *Virus Evol.* 5:2. doi: 10.1093/ve/vez051

Arman, W., and Munger, K. (2022). Mechanistic contributions of lncRNAs to cellular signaling pathways crucial to the lifecycle of human papillomaviruses. *Viruses* 14:2439. doi: 10.3390/v14112439

Bernard, H. U., Burk, R. D., Chen, Z., van Doorslaer, K., zur Hausen, H., and de Villiers, E. M. (2010). Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology* 401, 70–79. doi: 10.1016/j.virol.2010.02.002

Bossart, G. D., Romano, T. A., Peden-Adams, M. M., Schaefer, A. M., Rice, C. D., Fair, P. A., et al. (2019). Comparative innate and adaptive immune responses in Atlantic bottlenose dolphins (*Tursiops truncatus*) with viral, bacterial, and fungal infections. *Front. Immunol.* 10:1125. doi: 10.3389/fimmu.2019.01125

Brimer, N., Drews, C. M., and Vande Pol, S. B. (2017). Association of papillomavirus E6 proteins with either MAML1 or E6AP clusters E6 proteins by structure, function, and evolutionary relatedness. *PLoS Pathog.* 13:e1006781. doi: 10.1371/journal.ppat.1006781

Cheng, D., Qiu, K., Rao, Y., Mao, M., Li, L., Wang, Y., et al. (2023). Proliferative exhausted CD8(+) T cells exacerbate long-lasting anti-tumor effects in human papillomavirus-positive head and neck squamous cell carcinoma. *Elife* 12:e82705. doi: 10.7554/eLife.82705

Coursey, T. L., Van Doorslaer, K., and McBride, A. A. (2021). Regulation of Human Papillomavirus 18 genome replication, establishment, and persistence by sequences in the viral upstream regulatory region. *J. Virol.* 95:e0068621. doi: 10.1128/jvi.00686-21

Cui, J., Li, F., and Shi, Z. L. (2019). Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* 17, 181–192. doi: 10.1038/s41579-018-0118-9

de Villiers, E. M., Fauquet, C., Broker, T. R., Bernard, H. U., and zur Hausen, H. (2004). Classification of papillomaviruses. *Virology* 324, 17–27. doi: 10.1016/j.virol.2004.03.033

Egberink, H., Thiry, E., Möstl, K., Addie, D., Belák, S., Boucraut-Baralon, C., et al. (2013). Feline viral papillomatosis: ABCD guidelines on prevention and management. *J. Feline Med. Surg.* 15, 560–562. doi: 10.1177/1098612x13489213

Fouchier, R. A., Schneeberger, P. M., Rozendaal, F. W., Broekman, J. M., Kemink, S. A., Munster, V., et al. (2004). Avian influenza A virus (H7N7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome. *Proc. Natl. Acad. Sci. U.S.A.* 101, 1356–1361. doi: 10.1073/pnas.0308352100

Gaynor, A. M., Fish, S., Duerr, R. S., Cruz, F. N. Jr., and Pesavento, P. A. (2015). Identification of a novel papillomavirus in a Northern Fulmar (*Fulmarus glacialis*) with viral production in cartilage. *Vet. Pathol.* 52, 553–561. doi: 10.1177/0300985814542812

Geoghegan, J. L., Duchêne, S., and Holmes, E. C. (2017). Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. *PLoS Pathog.* 13:e1006215. doi: 10.1371/journal.ppat.1006215

Gottschling, M., Bravo, I. G., Schulz, E., Bracho, M. A., Deaville, R., Jepson, P. D., et al. (2011). Modular organizations of novel cetacean papillomaviruses. *Mol. Phylogenet. Evol.* 59, 34–42. doi: 10.1016/j.ympev.2010.12.013

Groth, M., Lange, J., Kanrai, P., Pleschka, S., Scholtissek, C., Krumbholz, A., et al. (2014). The genome of an influenza virus from a pilot whale: Relation to influenza viruses of gulls and marine mammals. *Infect. Genet. Evol.* 24, 183–186. doi: 10.1016/j.meegid.2014.03.026

Kraberger, S., Austin, C., Farkas, K., Desvignes, T., Postlethwait, J. H., Fontenele, R. S., et al. (2022). Discovery of novel fish papillomaviruses: From the Antarctic to the commercial fish market. *Virology* 565, 65–72. doi: 10.1016/j.virol.2021.10.007

Mäkelä, S. M., Österlund, P., Westenius, V., Latvala, S., Diamond, M. S., Gale, M. Jr., et al. (2015). RIG-I signaling is essential for influenza B virus-induced rapid interferon gene expression. *J. Virol.* 89, 12014–12025. doi: 10.1128/jvi.01576-15

Mihindukulasuriya, K. A., Wu, G., St Leger, J., Nordhausen, R. W., and Wang, D. (2008). Identification of a novel coronavirus from a beluga whale by using a panviral microarray. *J. Virol.* 82, 5084–5088. doi: 10.1128/jvi.02722-07

Nadile, M., Retsidou, M. I., Gioti, K., Beloukas, A., and Tsiani, E. (2022). Resveratrol against cervical cancer: Evidence from in vitro and in vivo studies. *Nutrients* 14:5273. doi: 10.3390/nu14245273

Pino, M. D., Matas, I., Carrillo, P., Martí, C., Glickman, A., Carreras-Dieguez, N., et al. (2023). Natural history of anal HPV infection in women treated for cervical intraepithelial neoplasia. *Cancers* 15:1147. doi: 10.3390/cancers15041147

Prabhakar, A. T., James, C. D., Fontan, C. T., Otoa, R., Wang, X., Bristol, M. L., et al. (2023). Human papillomavirus 16 E2 interaction with TopBP1 is required for E2 and viral genome stability during the viral life cycle. *J. Virol.* 97:e0006323. doi: 10.1128/jvi.00063-23

Robles-Sikisaka, R., Rivera, R., Nollens, H. H., St Leger, J., Durden, W. N., Stolen, M., et al. (2012). Evidence of recombination and positive selection in cetacean papillomaviruses. *Virology* 427, 189–197. doi: 10.1016/j.virol.2012.01.039

Roess, A. A., Levine, R. S., Barth, L., Monroe, B. P., Carroll, D. S., Damon, I. K., et al. (2011). Sealpox virus in marine mammal rehabilitation facilities, North America, 2007-2009. *Emerg. Infect. Dis.* 17, 2203–2208. doi: 10.3201/eid1712.101945

Sacristán, C., Esperón, F., Ewbank, A. C., Díaz-Delgado, J., Ferreira-Machado, E., Costa-Silva, S., et al. (2019). Novel herpesviruses in riverine and marine cetaceans from South America. *Acta Trop.* 190, 220–227. doi: 10.1016/j.actatropica.2018.11.021

Schulz, E., Gottschling, M., Ulrich, R. G., Richter, D., Stockfleth, E., and Nindl, I. (2012). Isolation of three novel rat and mouse papillomaviruses and their genomic characterization. *PLoS One* 7:e47164. doi: 10.1371/journal.pone.0047164

Seyoum, A., Seyoum, B., Gure, T., Alemu, A., Belachew, A., Abeje, D., et al. (2023). Genotype heterogeneity of high-risk human papillomavirus infection in Ethiopia. *Front. Microbiol.* 14:1116685. doi: 10.3389/fmicb.2023.1116685

Sofiani, V. H., Veisi, P., Rukerd, M. R. Z., Ghazi, R., and Nakhaie, M. (2023). The complexity of human papilloma virus in cancers: A narrative review. *Infect. Agent Cancer* 18:13. doi: 10.1186/s13027-023-00488-w

Su, S., Wong, G., Shi, W., Liu, J., Lai, A. C. K., Zhou, J., et al. (2016). Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol.* 24, 490–502. doi: 10.1016/j.tim.2016.03.003

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Van Bressem, M. F., Cassonnet, P., Rector, A., Desaintes, C., Van Waerebeek, K., Alfaro-Shigueto, J., et al. (2007). Genital warts in Burmeister's porpoises: Characterization of Phocoena spinipinnis papillomavirus type 1 (PsPV-1) and evidence for a second, distantly related PsPV. *J. Gen. Virol.* 88, 1928–1933. doi: 10.1099/vir.0.82694-0

Van Doorslaer, K., Ruoppolo, V., Schmidt, A., Lescroël, A., Jongsomjit, D., Elrod, M., et al. (2017). Unique genome organization of non-mammalian papillomaviruses provides insights into the evolution of viral early proteins. *Virus Evol.* 3:vex027. doi: 10.1093/ve/vex027

Vargas-Castro, I., Melero, M., Crespo-Picazo, J. L., Jiménez, M., Sierra, E., Rubio-Guerri, C., et al. (2021). Systematic determination of herpesvirus in free-ranging cetaceans stranded in the Western Mediterranean: Tissue tropism and associated lesions. *Viruses* 13:2180. doi: 10.3390/v13112180

Waltzek, T. B., Cortés-Hinojosa, G., Wellehan, J. F. Jr., and Gray, G. C. (2012). Marine mammal zoonoses: A review of disease manifestations. *Zoonoses Public Health* 59, 521–535. doi: 10.1111/j.1863-2378.2012.01492.x

Wang, G., Huang, Y., Zhang, W., Peng, R., Luo, J., Liu, S., et al. (2021). Identification and genome analysis of a novel picornavirus from captive belugas (Delphinapterus leucas) in China. *Sci. Rep.* 11:21018. doi: 10.1038/s41598-021-00605-y

Warburton, A., Redmond, C. J., Dooley, K. E., Fu, H., Gillison, M. L., Akagi, K., et al. (2018). HPV integration hijacks and multimerizes a cellular enhancer to generate a viral-cellular super-enhancer that drives high viral oncogene expression. *PLoS Genet.* 14:e1007179. doi: 10.1371/journal.pgen.1007179

Woo, P. C., Lau, S. K., Lam, C. S., Tsang, A. K., Hui, S. W., Fan, R. Y., et al. (2014). Discovery of a novel bottlenose dolphin coronavirus reveals a distinct species of marine mammal coronavirus in Gammacoronavirus. *J. Virol.* 88, 1318–1331.

Wu, Z., Du, J., Lu, L., Yang, L., Dong, J., Sun, L., et al. (2018a). Detection of hantaviruses and arenaviruzses in three-toed jerboas from the inner mongolia autonomous region, China. *Emerg. Microbes Infect.* 7:35. doi: 10.1038/s41426-018-0036-y

Wu, Z., Liu, B., Du, J., Zhang, J., Lu, L., Zhu, G., et al. (2018b). Discovery of diverse rodent and bat pestiviruses with distinct genomic and phylogenetic characteristics in several Chinese provinces. *Front. Microbiol.* 9:2562. doi: 10.3389/fmicb.2018.02562

Wu, Z., Ren, X., Yang, L., Hu, Y., Yang, J., He, G., et al. (2012). Virome analysis for identification of novel mammalian viruses in bat species from Chinese provinces. *J. Virol.* 86, 10999–11012. doi: 10.1128/jvi.01394-12

Yang, J., Yang, F., Ren, L., Xiong, Z., Wu, Z., Dong, J., et al. (2011). Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. *J. Clin. Microbiol.* 49, 3463–3469. doi: 10.1128/jcm.00273-11

Yang, X., Li, Y., Tang, Y., Li, Z., Wang, S., Luo, X., et al. (2023). Cervical HPV infection in Guangzhou, China: An epidemiological study of 198,111 women from 2015 to 2021. *Emerg. Microbes Infect.* 12:e2176009. doi: 10.1080/22221751.2023.2176009

Zhang, M., Kiyono, T., Aoki, K., Goshima, N., Kobayashi, S., Hiranuma, K., et al. (2022). Development of an in vitro carcinogenesis model of human papillomavirus-induced cervical adenocarcinoma. *Cancer Sci.* 113, 904–915. doi: 10.1111/cas.15246

Zhi, W., Wei, Y., Lazare, C., Meng, Y., Wu, P., Gao, P., et al. (2023). HPV-CCDC106 integration promotes cervical cancer progression by facilitating the high expression of CCDC106 after HPV E6 splicing. *J. Med. Virol.* 95:e28009. doi: 10.1002/jmv.28009

Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273. doi: 10.1038/s41586-020-2012-7

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 30 August 2023
DOI 10.3389/fmicb.2023.1252272

Canine circovirus among dogs and cats in China: first identification in cats

Xiangyu Xiao†, Yan chao Li†, Feng pei Xu, Xiangqi Hao, Shoujun Li and Pei Zhou*

Guangdong Provincial Pet Engineering Technology Research Center, College of Veterinary Medicine, South China Agricultural University, Guangzhou, Guangdong, China

OPEN ACCESS

EDITED BY
Shailendra Saxena,
King George's Medical University, India

REVIEWED BY
Jing Yang,
National Institute for Communicable Disease Control and Prevention (China CDC), China
Julian Ruiz-Saenz,
Cooperative University of Colombia, Colombia

*CORRESPONDENCE
Pei Zhou
✉ zhoupei@scau.edu.cn

†These authors have contributed equally to this work and share first authorship

RECEIVED 03 July 2023
ACCEPTED 15 August 2023
PUBLISHED 30 August 2023

CITATION
Xiao X, Li Yc, Xu Fp, Hao X, Li S and Zhou P (2023) Canine circovirus among dogs and cats in China: first identification in cats. *Front. Microbiol.* 14:1252272.
doi: 10.3389/fmicb.2023.1252272

Canine circovirus (CanineCV) is a virus associated with respiratory and digestive diseases in dogs and often occurs in coinfections with other pathogens, thereby aggravating the symptoms of infected dogs. CanineCV was first reported in the United States in 2012. Subsequently, it was reported among dogs in Europe, Asia, and South America. To investigate the prevalence of CanineCV in dogs in China, 331 dog samples were collected in this study. The PCR results showed that 9.06% (30/331, 95% CI = 6.2% ~ 12.7%) of the dog samples were CanineCV positive. CanineCV has also been detected in some carnivorous wild animals, indicating the potential risk of cross-species transmission of this virus. And, cats are also one of the most common pets in our daily lives, who is close contact with dogs. Thus, this study first investigated the prevalence of CanineCV in cats. The PCR results showed that 3.42% (14/409, 95% CI = 1.9% ~ 5.7%) of the cat samples were CanineCV positive. Moreover, 14 canine-derived CanineCV whole genomes and the first cat-derived CanineCV whole genome were obtained in this study. Rep and Cap are the major nonstructural proteins and structural proteins of CanineCV, respectively. In nucleic acid homology analyses, these 15 CanineCV strains showed a high degree of variation in Rep (85.9 ~ 99%) and Cap (85.6 ~ 100%). In phylogenetic analyses, the 15 CanineCV strains clustered into 3 different genotypes (genotypes 1, 3, and 4). Among them, the first cat-derived CanineCV belonged to CanineCV-3. In addition, 4 genetic recombination events were predicted in these 15 CanineCV strains, occurring in multiple regions of the genome. In conclusion, this study is the first to provide evidence of CanineCV infection in cats and successfully obtained the first whole genome of cat-derived CanineCV. The complex circulation and high prevalence of CanineCV among dogs and cats emphasize the importance of continuous monitoring of this virus in various animal species.

KEYWORDS

canine circovirus, dogs and cats, high prevalence, genotypes, genetic recombination

1. Introduction

Canine circovirus (CanineCV) is a member of the genus Circovirus, family Circoviridae (Kapoor et al., 2012). Circoviruses also include porcine circovirus (Meng, 2013; Palinski et al., 2017), goose circovirus (GoCV) (Guo et al., 2011), duck circovirus (DuCV) (Hong et al., 2018), and other bird viruses (Todd et al., 2001a,b; Hong et al., 2018; Gorbalenya et al., 2020). The genome of CanineCV is an approximately 2062–2064 nt DNA genome and consists mainly of 3 open reading frames (ORFs). ORF1 encodes the replicate protein (Rep) and ORF2 the capsid protein (Cap); ORF3 encodes a 105 amino acids (aa) of unknown function (Gomez-Betancur et al., 2023). Pathogenicity varies among circoviruses. Most members of the genus Circovirus infect susceptible animals and can induce lymphoid tissue damage, leading to immunosuppression and a range of serious secondary infections. For example, DuCV can cause immunosuppression, which increases the risk of coinfection with other pathogens (Yuan et al., 2022). Porcine circovirus 2 (PCV2) is related to postweaning multisystemic wasting disease syndrome (PMWS) (Allan and Ellis, 2000; Rakibuzzaman and Ramamoorthy, 2021). In addition, some circoviruses only cause subclinical infections in susceptible animals, such as porcine circovirus 1 (PCV1), which does not cause cytopathic effects, although it induces the production of serum antibodies in swine (Rakibuzzaman and Ramamoorthy, 2021). CanineCV is associated with respiratory diseases and diarrhea in dogs (Li et al., 2013; Decaro et al., 2014; Hsu et al., 2016; Anderson et al., 2017; Cruz et al., 2020; Dankaona et al., 2022), and CanineCV may also cause immunosuppression due to coinfection with other pathogens, leading to aggravation of clinical symptoms (Dowgier et al., 2017; Kotsias et al., 2019). As CanineCV could not effectively replicate on cells, its viral isolation and cultivation *in vitro* have been unsuccessful (Gomez-Betancur et al., 2023).

CanineCV was first identified among dogs in the United States in 2012 (Kapoor et al., 2012). Subsequently, it was reported among dogs in Europe, Asia, and South America (Gomez-Betancur et al., 2023). Hence, the natural host of CanineCV is dog. However, the presence of CanineCV has been detected in various wild animal species. Italian researchers have identified the presence of CanineCV in organ samples from wolves and badgers in indian (Zaccaria et al., 2016). British researchers have found that the nucleotide homology between fox circovirus and CanineCV is approximately 89% (Bexton et al., 2015). CanineCV was detected in both red and arctic foxes by Norwegian researchers, who further determined that the transmission of CanineCV within the arctic fox population can be traced back to 1996 (Urbani et al., 2021). The above findings demonstrate the risk of cross-species transmission of CanineCV, highlighting the need for further investigation into its prevalence among both wild and domestic animals. Herein, we describe an investigation of the prevalence of CanineCV in dogs conducted in Shanghai, Guangzhou, Foshan, and Zhengzhou in China, and this study also represents the first investigation of CanineCV prevalence in cats.

2. Materials and methods

2.1. Sample collection

From November 2021 to September 2022, a total of 740 samples were collected from 331 dogs (21 fecal samples, 16 nasal swabs, and 294 serum samples) and 409 cats (9 fecal samples, 64 nasal swabs, and 336 serum samples) in Shanghai, Guangzhou, Foshan, and Zhengzhou in China. Essential background data on the animals were also documented. The samples were stored at −80°C until processing. Each sample was collected after permission was obtained from the owner, and the procedures met the requirements of the Experimental Animal Welfare Ethics Committee of South China Agricultural University.

2.2. Pretreatment of clinical samples

Nasal swabs and fecal samples: The samples were vortexed for 20 s and then centrifuged at 4°C and 12,000 r/min for 10 min, and the supernatant was transferred to a 1.5 mL centrifuge tube. Blood samples were centrifuged at 4°C and 4,000 r/min for 10 min, and the supernatant was transferred to a 1.5 mL centrifuge tube. All samples were stored at −80°C.

2.3. Nucleic acid extraction and reverse transcription

To detect CanineCV and the DNA virus associated with it, total viral nucleic acid was extracted from clinical samples. Nucleic acid was extracted using a RaPure Virus DNA/RNA Kit (Magen, Guangzhou, China) according to the manufacturer's instructions. To detect CanineCV-associated RNA viruses, the extracted viral nucleic acids were reverse-transcribed. cDNA of CanineCV-positive samples was obtained using 5 × ABScript III RT Mix (Abclonal, Wuhan, China) according to the manufacturer's instructions.

2.4. PCR detection of CanineCV and other pathogens

First, CanineCV was detected using primers reported in previous research (q-Rep218F 5′-GTTCGCCGWTGGTGCTT and q-Rep218R 5′-CCCGAGCAGGCTCAAAAT) (Hao et al., 2022). The amplification protocol included predenaturation at 98°C for 1 min, followed by 35 cycles of denaturation at 98°C for 20 s, annealing at 57°C for 30 s, and extension at 72°C for 15 s, with a final extension at 72°C for 5 min. The amplified band of the positive sample was 218 bp.

Other viral pathogens [Canine parvovirus virus 2 (CPV-2), Canine corona virus (CCoV), Canine astrovirus (CaAstV), Canine calicivirus (CaCV), Canine rotavirus (CRV), Canine distemper virus (CDV), Feline calicivirus (FCV) and Feline panleukopenia virus (FPV)]were also detected using specific primers and PCR programs, as previously reported (Gentsch et al., 1992; Jiang et al., 1999; Zhu et al., 2011; Hao et al., 2019; Chang et al., 2021; Zobba et al., 2021; Xiao et al., 2023). All primers (Table 1) were synthesized by Sangon Biotech (Shanghai, China).

2.5. Amplification and sequencing of the CanineCV complete genome

According to the detected results, CanineCV-positive samples were selected to perform amplification of the full-length CanineCV genome. First, based on rolling circle amplification (RCA), Phi29 DNA Polymerase (Transgen, Beijing, China) was used to amplify the CanineCV genomes in the samples (Rector et al., 2004). Subsequently, one pair of primers was designed to amplify the full genome of CanineCV (CanineCV-F 5′-TCTCGCTCCGCCCGAAGCAC and CanineCV-R 5′-ACGCACCGCCTCCCCTCGCA), and the viral genome was amplified by PCR using ApexHF HS DNA Polymerase (AG, Hunan, China). The 2,123 bp band was purified and cloned into the M5 Hiper pTOPO-Blunt vector (Mei5bio, Beijing, China). The ligation products were transformed into DH5α cells, and positive clones tested by PCR were sent to Sangon Biotech for sequencing. Finally, genome sequences were assembled using SnapGene software (version 4.2.4).

TABLE 1 Primers used for PCR detection.

Pathgeon	Primer name	Primer sequence (5′-3′)	PCR products
CPV-2	CPV2-F	AAGACGTGCAAGCGAGTCC	337 bp
	CPV2-R	GAGCGAAGATAAGCAGCGTAA	
CCoV	CCoV-F	AGGAAGGCAACAATCCAATA	477 bp
	CCoV-R	GCCACCTCTGATGGACGA	
CaAstV	CaAstV-F	CAANTCACAACCCAAAACAAA	480 bp
	CaAstV-R	TTTTNACNATCACTGCTAGNG	
CaCV	CaCV-F	GATTACTCCASSTGGGAYTCMAC	319 bp
	CaCV-R	TGACGATTTCATCATCMCCRTA	
CRV	CRV-F	ATTTCGGACCATTTATAACC	876 bp
	CRV-R	TGGCTTCGCCATTTTATAGACA	
CDV	CDV-F	AGATTCAGCCATTTGTAGCCA	794 bp
	CDV-R	GTTGGACTACCTGAGCCCTA	
FCV	FCV-F	AACCTGCGCTAACGTGCT	1940 bp
	FCV-R	TGWATTCCCATGTAGGAGGC	
FPV	FPV-F	AAGACGTGCAAGCGAGTCC	337 bp
	FPV-R	GAGCGAAGATAAGCAGCGTAA	

TABLE 2 Positive detection rates of CanineCV in different samples.

Species	Health status	Number of positive samples	Number of samples	Positive detection rate (%)	95% CI	Correlation coefficient	Pearson chi-square test	Phi
Dog	Healthy	6	69	8.7	3.3% ~ 18%	0.221	0.012 ($\chi^2 = 6.354$)	0.226 ($p = 0.012$)
	Sick	14	55	25.45	14.7% ~ 39.0%			
Cat	Healthy	1	90	1.11	0% ~ 6%	0.086	0.3 ($\chi^2 = 1.074$)	0.086 ($p = 0.3$)
	Sick	2	55	3.64	0.4% ~ 12.5%			

2.6. Genetic characterization and phylogenetic analyses

Nucleotide sequences were edited using the BioEdit program and aligned by the ClustalW method. Homology analysis was performed with the DNAStar software package (MegAlign). The obtained complete genomes were aligned with 199 reference sequences of CanineCV and 13 reference sequences of other circoviruses in the GenBank database using the MAFFT algorithm.[1] A phylogenetic tree was constructed using MEGA (version 7.0.26), and evolutionary distances were computed using the neighbor-joining method (p-distance model). Statistical support was provided by 1,000 bootstrap replicates. Other circoviruses were selected as an outgroup to determine the roots of the evolutionary tree.

2.7. Recombination analyses

RDP (version 4) was used to analyze the whole genomes of all CanineCV strains that had been aligned with MAFFT, and all sequences were downloaded from GenBank. Recombination events were considered only if they were detected by at least four of seven programs (RDP, GENECONV, Maxchi, Chimera, 3Seq, Bootscan, and SiSscan) with a p value <0.05. The recombinant breakpoints were confirmed using SimPlot, with a sliding window of 200 bp (step:10 bp). And phylogenetic analysis was used for further verification.

3. Results

3.1. Prevalence of CanineCV in dogs and cats

PCR detection showed that the prevalence of CanineCV was 9.06% (30/331, 95% CI = 6.2% ~ 12.7%) in dogs and 3.42% (14/409, 95% CI = 1.9% ~ 5.7%) in cats. The prevalence of CanineCV was 25.45% (14/55, 95% CI = 14.7% ~ 39.0%) among dogs with diarrheic or respiratory symptoms; the prevalence among healthy dogs was 8.7% (6/69, 95% CI = 3.3% ~ 18%). The prevalence of CanineCV was 3.64% (2/55, 95% CI = 0.4% ~ 12.5%) among cats with diarrheic or respiratory symptoms, whereas it was 1.11% (1/90, 95% CI = 0% ~ 6%) among healthy cats (Tables 2, 3). These data show that the prevalence of CanineCV among animals with diarrheic or respiratory symptoms was higher than that among healthy animals, indicating that

---

[1] https://mafft.cbrc.jp/alignment/software/

TABLE 3 General information and detection results for the positive samples in this study.

Sample name	Location	Species	Sample type	Other pathogens detected
1*	Shanghai	Dog	Feces	CPV-2
2*	Guangzhou	Dog	Feces	CPV-2
3	Guangzhou	Dog	Serum	-
4	Guangzhou	Dog	Serum	-
5	Guangzhou	Dog	Serum	-
6	Guangzhou	Dog	Serum	-
7*	Guangzhou	Dog	Serum	-
8	Guangzhou	Dog	Serum	-
9	Guangzhou	Dog	Serum	-
10*	Guangzhou	Dog	Serum	-
11	Guangzhou	Dog	Serum	-
12	Guangzhou	Dog	Serum	CPV-2
13	Guangzhou	Dog	Serum	-
14	Guangzhou	Dog	Serum	-
15*	Guangzhou	Dog	Serum	-
16	Guangzhou	Dog	Serum	-
17	Zhengzhou	Dog	Feces	CCoV+CaCV
18	Zhengzhou	Dog	Feces	CaCV
19	Zhengzhou	Dog	Feces	CPV-2 + CCoV+CaCV
20	Zhengzhou	Dog	Feces	CCoV+CaCV
21	Guangzhou	Dog	Serum	CPV-2
22	Guangzhou	Dog	Serum	CPV-2
23*	Guangzhou	Dog	Serum	CPV-2
24*	Guangzhou	Dog	Serum	CPV-2
25	Zhengzhou	Dog	Feces	CPV-2 + CaCV
26	Zhengzhou	Dog	Feces	-
27	Zhengzhou	Dog	Feces	CPV-2 + CCoV
28	Guangzhou	Dog	Serum	-
29*	Foshan	Dog	Feces	-
30	Guangzhou	Dog	Nasal swab	CCoV
31	Guangzhou	Cat	Serum	-
32	Guangzhou	Cat	Serum	-
33	Guangzhou	Cat	Serum	-
34	Guangzhou	Cat	Serum	FPV
35	Guangzhou	Cat	Serum	-
36	Guangzhou	Cat	Serum	FPV
37	Guangzhou	Cat	Serum	-
38	Guangzhou	Cat	Serum	FPV
39	Guangzhou	Cat	Serum	-
40	Guangzhou	Cat	Serum	-
41	Guangzhou	Cat	Serum	FPV
42	Guangzhou	Cat	Serum	FPV
43	Foshan	Cat	Nasal swab	-
44*	Guangzhou	Cat	Nasal swab	-

"-" denotes negative; "*" denotes samples that successfully amplified the full genome sequence; canine coronavirus (CCoV), canine parvovirus 2 (CPV-2), canine calicivirus (CaCV), feline panleukopenia virus (FPV).

CanineCV is associated with animal diseases. To elucidate the possibility of a correlation between CanineCV and diarrhea and respiratory disease, the data were processed for calculating the Phi coefficient of association. The correlation coefficient of CanineCV and canine diarrhea and respiratory disease was 0.221. The *p* value of pearson chi-square test was 0.012 ($\chi^2 = 6.354$) and Phi was 0.226 ($p = 0.012$), respectively. The result suggested a positive association between CanineCV and canine diarrhea and respiratory disease. The correlation coefficient of CanineCV and feline diarrhea and respiratory disease was 0.086. The *p* value of pearson chi-square test was 0.3 ($\chi^2 = 1.074$) and Phi was 0.086 ($p = 0.3$), respectively. The result suggested no association between CanineCV and feline diarrhea and respiratory disease.

Of all positive samples, 46.67% (14/30) dog-positive samples were coinfected with other canine pathogens (CPV-2 or CCoV or CaCV); 35.71% (5/14) cat-positive samples were coinfected with FPV.

3.2. Whole genome of CanineCV

Fifteen full genomes of CanineCV strains were obtained, as amplified from 9 positive samples: isolates SH1, SH2, SH3, GZ1, GZ2, GZ3, GZ4, GZ5, GZ6, GZ7, GZ8, GZ9, FS1, FS2, and FS3 (GenBank ID OP575971-OP575985), with GZ9 being amplified from a cat sample named 44.

All 15 CanineCV genomes were 2063 nt in length, consistent with most CanineCV strains in the GenBank database. Sequence analysis revealed a nucleotide composition of 52 ~ 53% GC and 47 ~ 48% AT in these CanineCV strains. Nucleic acid Homology analysis showed that the 15 CanineCV strains shared 86.6% ~ 100% identity and 82.1% ~ 99.3% identity with published sequences in GenBank (Table 4). Moreover, two main ORFs indicated that Rep and Cap had a high degree of variation at both the nt and aa levels.

3.3. Phylogenetic analysis

The NJ phylogenetic tree was constructed based on CanineCV whole genomes. Phylogenetic analysis showed that all CanineCV strains segregated into 5 clades, corresponding to 5 genotypes (Figure 1). The CanineCV sequences identified in this study were divided into 3 genotypes (CanineCV-1, CanineCV-3, and CanineCV-4). Among, the first cat-derived CanineCV belongs to genotype 3 (Figure 1).

3.4. Recombination analyses

Four potential recombination events with a transferred fragment ranging in size from 11.3% ~ 61.6% of the genome were detected using RDP4 software (Table 5), and the results were further confirmed using Simplot (Figure 2) and phylogenetic analyses (Figure 3). Notably, recombination events occurred not only in the 2 ORF genes but also in other parts of the genome. Moreover, events 1, 2, and 3 were intergenotypic recombination events, whereas event 4 was a recombination event that occurred within genotype 3. Interestingly, the major parent and minor parent of recombination event 4 were isolated from Italian wolf and dog, respectively. In addition, the major parents of other recombination events were isolated from dogs in Canada, China, and the United States. The minor parents of other recombination events were isolated from dogs in China and Italy.

4. Discussion

CanineCV has been discovered in dogs and some carnivorous wild animals. In this study, the prevalence of CanineCV was 25.45% (14/55, 95% CI = 14.7% ~ 39.0%) among dogs with diarrheic or respiratory symptoms; the prevalence among healthy dogs was 8.7% (6/69, 95% CI = 3.3% ~ 18%). The high prevalence of CanineCV among dogs is consistent with previous reports from Iran and Thailand (Turan and Işıdan, 2020; Beikpour et al., 2022). Our previous study described the first identification and full genomic characterization of feline stool-associated circular DNA virus in cats in China (Hao et al., 2021). Due to the close contact between cats and dogs, this study also investigated presence of CanineCV in cats. The prevalence of CanineCV in cats, 2.07%, was identified for the first time in this study. These data remind us of the importance of paying close attention to the prevalence of CanineCV in dogs and cats.

The CanineCV prevalence in animals with diarrhea or respiratory symptoms was higher than that in healthy animals, which is consistent with previous studies (Li et al., 2013; Decaro et al., 2014; Hsu et al., 2016; Anderson et al., 2017; Cruz et al., 2020). Moreover, 43.18% (19/44) of positive samples showed coinfection with other pathogens, which is also consistent with previous studies conducted in Italy and China (Hsu et al., 2016; Beikpour et al., 2022). Our previous study showed that CanineCV occurs in coinfections with other viruses, and immunosuppression may occur, aggravating clinical symptoms and mortality (Hao et al., 2022). Therefore, CanineCV detection in dogs and cats with diarrhea or respiratory symptoms is necessary.

TABLE 4 Homology analysis of CanineCV.

Homology analysis	Whole genome (%)	Rep (%)		Cap (%)	
	nt	nt	aa	nt	aa
CanineCV obtained in this study	86.6 ~ 100	85.4 ~ 100	92.9 ~ 100	84.3 ~ 100	93.3 ~ 100
CanineCV obtained in this study with published sequences in the GenBank database	82.1 ~ 99.3	79.6 ~ 99.1	85.9 ~ 99	79.2 ~ 99.4	85.6 ~ 100

CanineCV is divided into five genotypes (CanineCV-1, −2, −3, −4, and − 5) based on the whole genome (Urbani et al., 2021). The CanineCV-1 genotype is mainly reported in the USA, Europe, and Asia. The CanineCV-2, −3, and − 4 genotypes are mainly found in Asia (Niu et al., 2020; Urbani et al., 2021). Additionally, the CanineCV-5 genotype has been detected in Europe and North America (Urbani et al., 2021). This study obtained 15 genomes of CanineCV, and homology analysis demonstrated that all these strains shared a sequence identity ranging from 86.6 to 100%. The strains from Shanghai were genotype 3, whereas the strains from Guangdong were genotypes 1, 3, and 4. Notably, the strain from a cat was classified as genotype 3. This result provides evidence for the prevalence of at least 3 genotypes of CanineCV in Guangdong, indicating the complex circulation of CanineCV among dogs and cats.

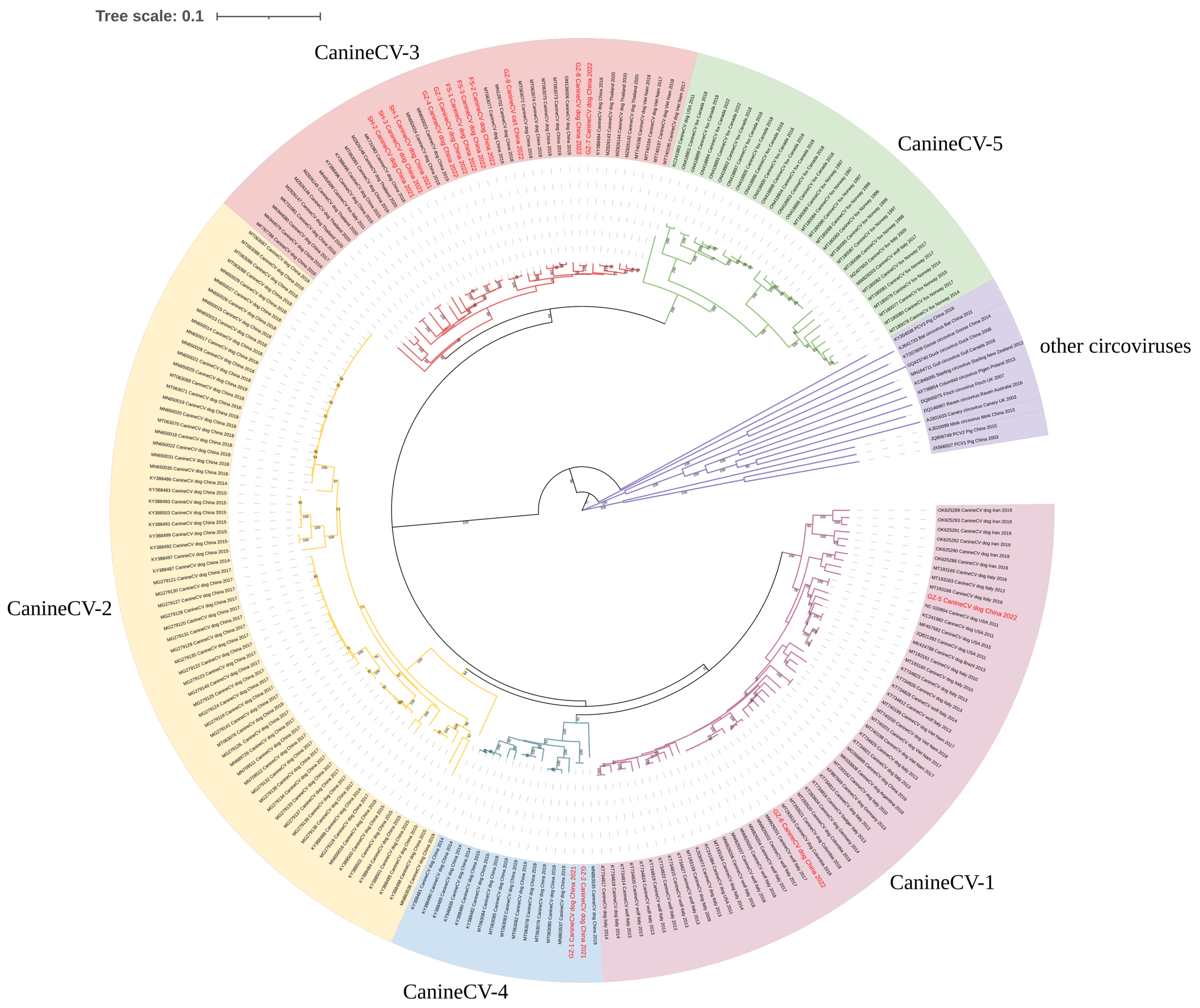


FIGURE 1
Neighbor-joining phylogenetic tree based on the full genome of CanineCV. Statistical support was provided by bootstrapping with 1,000 replicates. The red font indicates the strains obtained in this study.

TABLE 5 Information on recombination events detected using the RDP (R), GENECONV (G), Maxchi (M), Chimera (C), 3Seq (Q), Bootscan (B), and SiSscan (S) methods implemented in the computer program RDP4.

Event	Recombination sequence	Major parent	Minor parent	Position (in alignment)	Detection method ($p < 0.05$)
1	SH-1, SH-2, SH-3, GZ-3, GZ-4, GZ-7, GZ-8, GZ-9	ON418894 (Canada)	MN650016 (China)	343–1,062	RGBMCST
2	GZ-1, GZ-2	MT063085 (China)	MT193162 (Italy)	1794–2028	RGBMCST
3	FS-1, FS-2, FS-3	KC241983 (United States)	MN650016 (China)	354–1,062	RGBMCST
4	GZ-5	MW829207 (Italy)	MT193163 (Italy)	616–1887	RMST

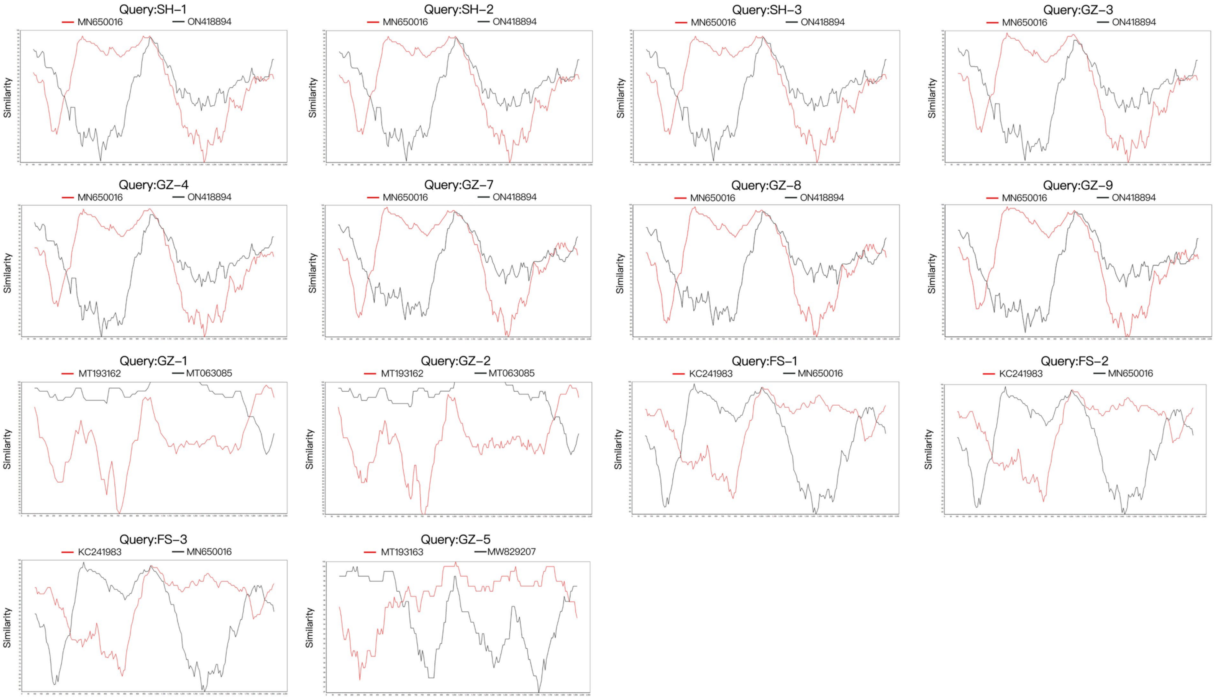


FIGURE 2
Recombination analysis of the 15 full genomes of CanineCV strains using SimPlot v3.5.1 with a window and step sizes of 200 nt and 10 nt, respectively.

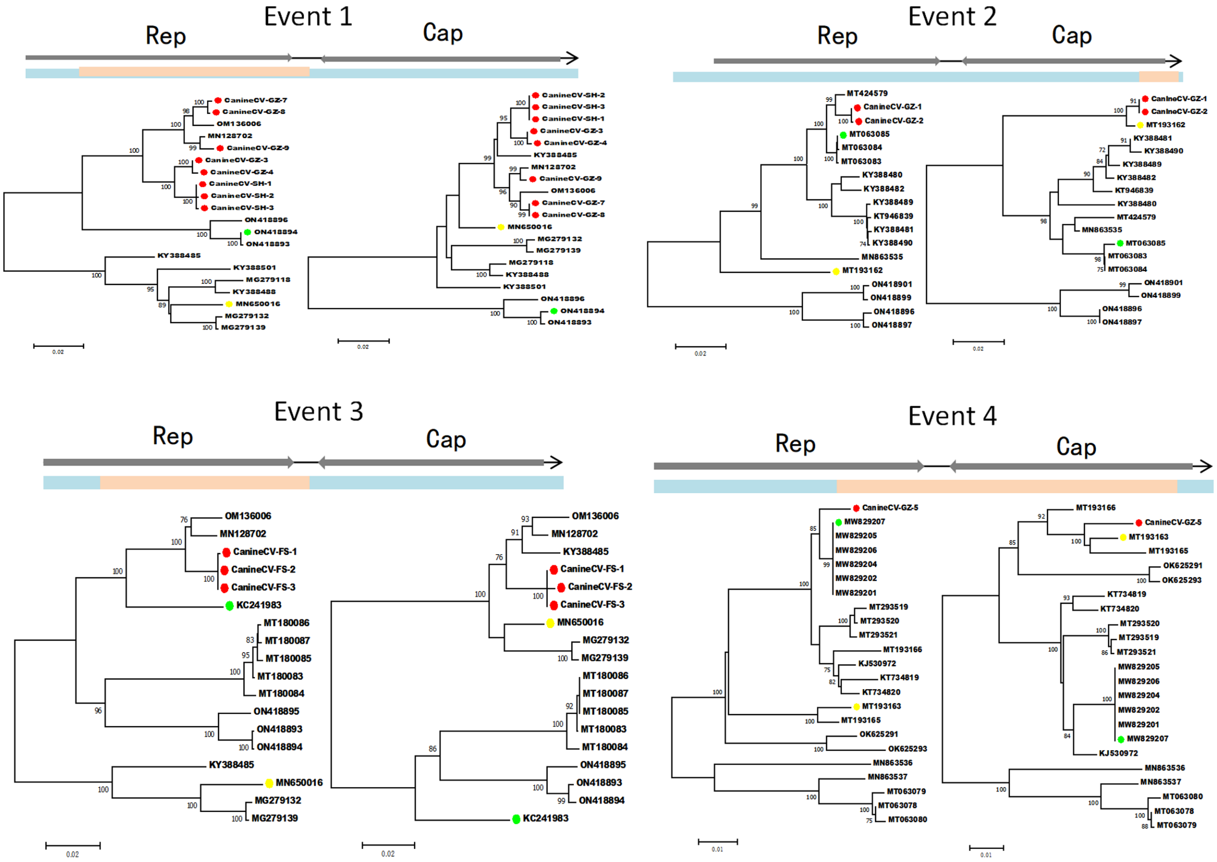


FIGURE 3
Phylogenetic trees illustrate the potential recombination events detected in this study. Blue indicates the sequence region derived from the major parent; pink indicates the sequence region derived from the minor parent. The dots marked in green, yellow, and red indicate the potential major parent, minor parent, and recombinant sequences, respectively. The phylogenetic trees were reconstructed using the neighbor-joining method. Statistical support was provided by bootstrapping with 1,000 replicates. The numbers (>70) above branches indicate percent bootstrap values.

Several circoviruses, including CanineCV, have undergone genetic recombination during evolution (Piewbang et al., 2018; Stenzel et al., 2018; Sun et al., 2019; Niu et al., 2020; Dankaona et al., 2022; Yuan et al., 2022). Four potential recombination events were detected in this study, including intragenotype and intergenotype recombination. Moreover, recombination events occurred not only in the 2 ORF genes but also in other parts of the genome, which is consistent with previous results (Piewbang et al., 2018; Sun et al., 2019; Tuong et al., 2021). More importantly, the major and minor parents in these 4 recombination events were not only from China but also from Europe and North America. Therefore, it is necessary to closely monitor the genomic recombination of CanineCV.

In conclusion, this study is the first to provide evidence of CanineCV infection in cats and successfully obtained the first whole genome of cat-derived CanineCV. The complex circulation and high prevalence of CanineCV among dogs and cats emphasize the importance of continuous monitoring of this virus in various animal species.

Data availability statement

The datasets presented in this study are deposited in the NCBI GenBank repository, accession numbers OP575971-OP575985.

Ethics statement

The animal studies were approved by the Experimental Animal Welfare Ethics Committee of South China Agricultural University. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

XX and YL participated in writing original draft and writing–review and editing. XX, YL, FX, and XH performed the experiments. SL and PZ contributed to conceptualization, methodology, validation, resources, writing–review and editing, supervision, and funding acquisition. All authors read and approved the final manuscript.

Funding

This work was supported by the Natural Science Foundation Guangdong province (2023A1515012171 and 2022A1515010733) and Guangzhou Municipal Science and Technology Bureau (SL2022A04J00674).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Allan, G. M., and Ellis, J. A. (2000). Porcine circoviruses: a review. *J. Vet. Diagn. Investig.* 12, 3–14. doi: 10.1177/104063870001200102

Anderson, A., Hartmann, K., Leutenegger, C. M., Proksch, A. L., Mueller, R. S., and Unterer, S. (2017). Role of canine circovirus in dogs with acute haemorrhagic diarrhoea. *Vet. Rec.* 180:542. doi: 10.1136/vr.103926

Beikpour, F., Ndiana, L. A., Sazmand, A., Capozza, P., Nemati, F., Pellegrini, F., et al. (2022). Detection and genomic characterization of canine circovirus in Iran. *Animals* 12:507. doi: 10.3390/ani12040507

Bexton, S., Wiersma, L. C., Getu, S., van Run, P. R., Verjans, G. M. G. M., Schipper, D., et al. (2015). Detection of circovirus in foxes with meningoencephalitis, United Kingdom, 2009–2013. *Emerg. Infect. Dis.* 21, 1205–1208. doi: 10.3201/eid2107.150228

Chang, L., Qi, Y., Liu, D., Du, Q., Zhao, X., and Tong, D. (2021). Molecular detection and genotyping of bovine viral diarrhea virus in western China. *BMC Vet. Res.* 17. doi: 10.1186/s12917-021-02747-7

Cruz, T. F., Batista, T. N., Vieira, E. M., Portela, L. M. F., Baccarin, A. M., Gradiz, J. J., et al. (2020). Genomic characterization of canine circovirus detected in a dog with intermittent hemorrhagic gastroenteritis in Brazil. *Ciência Rural* 50. doi: 10.1590/0103-8478cr20190909

Dankaona, W., Mongkholdej, E., Satthathum, C., Piewbang, C., and Techangamsuwan, S. (2022). Epidemiology, genetic diversity, and association of canine circovirus infection in dogs with respiratory disease. *Sci. Rep.-Uk* 12:15445. doi: 10.1038/s41598-022-19815-z

Decaro, N., Martella, V., Desario, C., Lanave, G., Circella, E., Cavalli, A., et al. (2014). Genomic characterization of a circovirus associated with fatal hemorrhagic enteritis in dog, Italy. *PLoS One* 9:e105909. doi: 10.1371/journal.pone.0105909

Dowgier, G., Lorusso, E., Decaro, N., Desario, C., Mari, V., Lucente, M. S., et al. (2017). A molecular survey for selected viral enteropathogens revealed a limited role of canine circovirus in the development of canine acute gastroenteritis. *Vet. Microbiol.* 204, 54–58. doi: 10.1016/j.vetmic.2017.04.007

Gentsch, J. R., Glass, R. I., Woods, P., Gouvea, V., Gorziglia, M., Flores, J., et al. (1992). Identification of group a rotavirus gene 4 types by polymerase chain reaction. *J. Clin. Microbiol.* 30, 1365–1373. doi: 10.1128/JCM.30.6.1365-1373.1992

Gomez-Betancur, D., Vargas-Bermudez, D. S., Giraldo-Ramírez, S., Jaime, J., and Ruiz-Saenz, J. (2023). Canine circovirus: an emerging or an endemic undiagnosed enteritis virus? *Front. Vet. Sci.* 10. doi: 10.3389/fvets.2023.1150636

Gorbalenya, A. E., Krupovic, M., Mushegian, A., Kropinski, A. M., Siddell, S. G., Varsani, A., et al. (2020). The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. *Nat. Microbiol.* 5, 668–674. doi: 10.1038/s41564-020-0709-x

Guo, J., Tian, J., Tan, X., Yu, H., Ding, S., Sun, H., et al. (2011). Pathological observations of an experimental infection of geese with goose circovirus. *Avian Pathol.* 40, 55–61. doi: 10.1080/03079457.2010.538371

Hao, X., Li, Y., Chen, H., Chen, B., Liu, R., Wu, Y., et al. (2022). Canine circovirus suppresses the type I interferon response and protein expression but promotes CPV-2 replication. *Int. J. Mol. Sci.* 23:6382. doi: 10.3390/ijms23126382

Hao, X., Li, Y., Hu, X., Fu, X., Dong, J., Zhang, H., et al. (2021). Feline stool-associated circular DNA virus (FESCV) in diarrheic cats in China. *Front. Vet. Sci.* 8:694089. doi: 10.3389/fvets.2021.694089

Hao, X., Liu, R., He, Y., Xiao, X., Xiao, W., Zheng, Q., et al. (2019). Multiplex PCR methods for detection of several viruses associated with canine respiratory and enteric diseases. *PLoS One* 14:e213295. doi: 10.1371/journal.pone.0213295

Hong, Y. T., Kang, M., and Jang, H. K. (2018). Pathogenesis of duck circovirus genotype 1 in experimentally infected Pekin ducks. *Poult. Sci.* 97, 3050–3057. doi: 10.3382/ps/pey177

Hsu, H., Lin, T., Wu, H., Lin, L., Chung, C., Chiou, M., et al. (2016). High detection rate of dog circovirus in diarrheal dogs. *BMC Vet. Res.* 12. doi: 10.1186/s12917-016-0722-8

Jiang, X., Huang, P. W., Zhong, W. M., Farkas, T., Cubitt, D. W., and Matson, D. O. (1999). Design and evaluation of a primer pair that detects both Norwalk- and Sapporo-like caliciviruses by RT-PCR. *J. Virol. Methods* 83, 145–154. doi: 10.1016/s0166-0934(99)00114-7

Kapoor, A., Dubovi, E. J., Henriquez-Rivera, J. A., and Lipkin, W. I. (2012). Complete genome sequence of the first canine circovirus. *J. Virol.* 86:7018. doi: 10.1128/JVI.00791-12

Kotsias, F., Bucafusco, D., Nuñez, D. A., Lago Borisovsky, L. A., Rodriguez, M., and Bratanich, A. C. (2019). Genomic characterization of canine circovirus associated with fatal disease in dogs in South America. *PLoS One* 14:e218735. doi: 10.1371/journal.pone.0218735

Li, L., McGraw, S., Zhu, K., Leutenegger, C. M., Marks, S. L., Kubiski, S., et al. (2013). Circovirus in tissues of dogs with vasculitis and hemorrhage. *Emerg. Infect. Dis.* 19, 534–541. doi: 10.3201/eid1904.121390

Meng, X. J. (2013). Porcine circovirus type 2 (pcv2): pathogenesis and interaction with the immune system. *Annu. Rev. Anim. Biosci.* 1, 43–64. doi: 10.1146/annurev-animal-031412-103720

Niu, L., Wang, Z., Zhao, L., Wang, Y., Cui, X., Shi, Y., et al. (2020). Detection and molecular characterization of canine circovirus circulating in northeastern China during 2014–2016. *Arch. Virol.* 165, 137–143. doi: 10.1007/s00705-019-04433-4

Palinski, R., Piñeyro, P., Shang, P., Yuan, F., Guo, R., Fang, Y., et al. (2017). A novel porcine circovirus distantly related to known circoviruses is associated with porcine dermatitis and nephropathy syndrome and reproductive failure. *J. Virol.* 91. doi: 10.1128/JVI.01879-16

Piewbang, C., Jo, W. K., Puff, C., van der Vries, E., Kesdangsakonwut, S., Rungsipipat, A., et al. (2018). Novel canine circovirus strains from Thailand: evidence for genetic recombination. *Sci. Rep. UK.* 8. doi: 10.1038/s41598-018-25936-1

Rakibuzzaman, A., and Ramamoorthy, S. (2021). Comparative immunopathogenesis and biology of recently discovered porcine circoviruses. *Transbound. Emerg. Dis.* 68, 2957–2968. doi: 10.1111/tbed.14244

Rector, A., Tachezy, R., and Van Ranst, M. (2004). A sequence-independent strategy for detection and cloning of circular DNA virus genomes by using multiply primed rolling-circle amplification. *J. Virol.* 78, 4993–4998. doi: 10.1128/jvi.78.10.4993-4998.2004

Stenzel, T., Dziewulska, D., Muhire, B., Hartnady, P., Kraberger, S., Martin, D., et al. (2018). Recombinant goose circoviruses circulating in domesticated and wild geese in Poland. *Viruses* 10:107. doi: 10.3390/v10030107

Sun, W., Zhang, H., Zheng, M., Cao, H., Lu, H., Zhao, G., et al. (2019). The detection of canine circovirus in Guangxi, China. *Virus Res.* 259, 85–89. doi: 10.1016/j.virusres.2018.10.021

Todd, D., Weston, J., Ball, N. W., Borghmans, B. J., Smyth, J. A., Gelmini, L., et al. (2001a). Nucleotide sequence-based identification of a novel circovirus of canaries. *Avian Pathol.* 30, 321–325. doi: 10.1080/03079450120066322

Todd, D., Weston, J. H., Soike, D., and Smyth, J. A. (2001b). Genome sequence determinations and analyses of novel circoviruses from goose and pigeon. *Virology* 286, 354–362. doi: 10.1006/viro.2001.0985

Tuong, N. M., Piewbang, C., Rungsipipat, A., and Techangamsuwan, S. (2021). Detection and molecular characterization of two canine circovirus genotypes co-circulating in Vietnam. *Vet. Quart.* 41, 232–241. doi: 10.1080/01652176.2021.1967511

Turan, T., and Işıdan, H. (2020). Molecular characterization of canine astrovirus, vesivirus and circovirus, isolated from diarrheic dogs in Turkey. *Iran. J. Vet. Res.* 21, 172–179. doi: 10.22099/IJVR.2020.35522.5212

Urbani, L., Tryland, M., Ehrich, D., Fuglei, E., Battilani, M., and Balboni, A. (2021). Ancient origin and genetic segregation of canine circovirus infecting arctic foxes (vulpes lagopus) in svalbard and red foxes (vulpes vulpes) in northern Norway. *Transbound. Emerg. Dis.* 68, 1283–1293. doi: 10.1111/tbed.13783

Xiao, X., Hao, X., Chen, B., Zhou, P., and Li, S. (2023). Two multiplex pcr methods for detecting several pathogens associated with feline respiratory and intestinal tracts. *Vet. Sci.* 10:14. doi: 10.3390/vetsci10010014

Yuan, S., Yao, X. Y., Yang, H. H., Zhang, Y. Q., Liu, H., Sun, J., et al. (2022). Research note: genetic diversity of duck circoviruses circulating in partial areas of Guangdong province, Southern China. *Poult. Sci.* 101:102032. doi: 10.1016/j.psj.2022.102032

Zaccaria, G., Malatesta, D., Scipioni, G., Di Felice, E., Campolo, M., Casaccia, C., et al. (2016). Circovirus in domestic and wild carnivores: an important opportunistic agent? *Virology* 490, 69–74. doi: 10.1016/j.virol.2016.01.007

Zhu, A. L., Zhao, W., Yin, H., Shan, T. L., Zhu, C. X., Yang, X., et al. (2011). Isolation and characterization of canine astrovirus in China. *Arch. Virol.* 156, 1671–1675. doi: 10.1007/s00705-011-1022-z

Zobba, R., Visco, S., Sotgiu, F., Pinna Parpaglia, M. L., Pittau, M., and Alberti, A. (2021). Molecular survey of parvovirus, astrovirus, coronavirus, and calicivirus in symptomatic dogs. *Vet. Res. Commun.* 45, 31–40. doi: 10.1007/s11259-020-09785-w

Frontiers in Microbiology

TYPE Original Research
PUBLISHED 31 August 2023
DOI 10.3389/fmicb.2023.1239420

EDITED BY
Wenshi Wang,
Xuzhou Medical University, China

REVIEWED BY
Qingfeng He,
Fudan University, China
Rui Qi,
Lanzhou University, China

*CORRESPONDENCE
Sen Wang
✉ njwangsen@163.com
Yuxin Chen
✉ yuxin.chen@nju.edu.cn

†These authors have contributed equally to this work and share first authorship

RECEIVED 13 June 2023
ACCEPTED 14 August 2023
PUBLISHED 31 August 2023

CITATION
Huang T, Fan Y, Xia Y, Xu X, Chen X, Ye H, Chen Y and Wang S (2023) Association of low HDL-c levels with severe symptoms and poor clinical prognosis in patients with severe fever and thrombocytopenia syndrome.
Front. Microbiol. 14:1239420.
doi: 10.3389/fmicb.2023.1239420

Association of low HDL-c levels with severe symptoms and poor clinical prognosis in patients with severe fever and thrombocytopenia syndrome

Taihong Huang[1†], Yinyin Fan[2†], Yanyan Xia[1], Xuejing Xu[1], Xinyue Chen[1], Hongling Ye[1], Yuxin Chen[1*] and Sen Wang[1*]

[1]Department of Clinical Laboratory Medicine, Nanjing Drum Tower Hospital Clinical College of Nanjing Medical University, Nanjing, China, [2]Department of Pancreatic Surgery, Nanjing Drum Tower Hospital Clinical College of Jiangsu University, Nanjing, China

Background: Severe fever with thrombocytopenia syndrome (SFTS) is an acute infectious disease caused by a novel bunyavirus, characterized by high fever, thrombocytopenia, and multiple organ damage. Disturbances in lipid metabolism often occur during viral infections, but the changes and clinical significance of lipid profiles in SFTS patients remain unclear. This study aimed to investigate the alterations in lipid profiles and their clinical significance in SFTS patients.

Methods: A total of 157 SFTS patients and 157 healthy controls were enrolled in this study. Serum lipid levels were collected and analyzed among different groups and prognosis categories. Receiver operating characteristic (ROC) curve analysis was performed to assess the ability of lipid levels in distinguishing between severe and mild cases, as well as surviving and non-surviving patients. Pearson correlation analysis was used to examine the associations between lipid levels and clinical laboratory parameters.

Results: SFTS patients exhibited significantly lower levels of HDL-c, LDL-c, cholesterol, APoAI, and ApoB compared to healthy controls, while triglyceride levels were significantly higher. Serum HDL-c and ApoAI demonstrated good performance as indicators for distinguishing between survivors and non-survivors (AUC of 0.87 and 0.85, respectively). Multivariate regression analysis indicated that HDL-c independently acts as a protective factor in patients with SFTS. HDL-c levels showed decline in non-survivors but recovered in survivors. Moreover, HDL-c exhibited significant correlations with various clinical laboratory parameters (IL-6, CRP, AST, TT, APTT, PLT, ALB, and CD4).

Conclusion: This study identified abnormalities in serum lipid metabolism among SFTS patients. HDL-c and ApoAI levels hold potential as biomarkers for distinguishing survivors from non-survivors. Additionally, HDL-c and ApoAI may serve as therapeutic targets for the management of SFTS patients.

KEYWORDS

severe fever with thrombocytopenia syndrome, novel bunyavirus, HDL, APoAI, biomarker

1. Introduction

Severe fever with thrombocytopenia syndrome (SFTS) is an emerging infectious disease caused by a novel bunyavirus. Over the past few years, SFTS has garnered global attention due to its significant impact. The initial recognition and naming of SFTS occurred in 2009 when a patient presented with clinical manifestations of persistent high fever, general malaise, and thrombocytopenia (Yu et al., 2011). Since then, an increasing number of similar cases have been reported, particularly in the Asian region, notably in China, Japan, and South Korea (Yoshikawa et al., 2015; Yun et al., 2017). The cases of SFTS are primarily distributed in rural areas within mountainous and hilly regions, occurring sporadically. This disease is most prevalent during the summer, and the population is generally susceptible. SFTS virus (SFTSV) is an RNA virus with a single-stranded genome, primarily transmitted through tick bites (Ren et al., 2021). Ticks serve as the main vectors responsible for infecting hosts, including humans, livestock, and wild animals, thereby facilitating the transmission of SFTSV (Luo et al., 2015). The clinical presentation of SFTS primarily encompasses abrupt onset of high fever, fatigue, headache, muscle and joint pain, alongside notable thrombocytopenia. Patients frequently exhibit bleeding manifestations, including epistaxis, skin congestion, and gum bleeding. Severe cases may involve gastrointestinal bleeding, central nervous system symptoms, and multiple organ failure (Wang et al., 2022). Although the clinical manifestations of SFTS bear similarities to other viral hemorrhagic fevers, this disease demonstrates a more rapid progression and higher case fatality rate (Zhao et al., 2021). The underlying mechanisms of SFTSV infection remain incompletely elucidated, and regrettably, there is currently no specific therapy or vaccine available for combating SFTS. Consequently, precise evaluation of disease severity and accurate prediction of disease progression assume paramount importance.

Abnormal lipid metabolism is pivotal in the development and progression of a range of diseases, including cardiovascular disease, metabolic syndrome, and diabetes (Tenenbaum et al., 2006; Chen et al., 2022). High-density lipoprotein (HDL) stands as a key player in lipid metabolism, regulating cholesterol regulation and demonstrating noteworthy capacities in anti-infection, anti-inflammation, anti-apoptosis, and anti-oxidation functions (Grunfeld et al., 1999; Murphy et al., 2008). Viral infections have been found to impact lipid metabolism, leading to alterations in serum lipid concentrations, particularly cholesterol levels, among patients infected with human immunodeficiency virus (HIV) and hepatitis C virus (HCV) (Kulasegaram et al., 2005; Bassendine et al., 2013). In the case of SARS-CoV-2 infection, COVID-19 patients experience significant decreases in low-density lipoprotein cholesterol (LDL-C) and high-density lipoprotein cholesterol (HDL-C), with low levels of HDL-C being associated with a poorer prognosis (Sun et al., 2020; Masana et al., 2021). The understanding of lipid metabolism and its clinical implications in patients with SFTS is currently lacking. This study aimed to investigate the relationship between lipid levels and disease prognosis among SFTS patients.

2. Materials and methods

2.1. Patients and control participants

This study analyzed a total of 157 hospitalized patients who were diagnosed with fever and thrombocytopenia at the Department of Infectious Diseases in Nanjing Drum Tower Hospital between January 2021 and December 2022. This study was approved by the Institutional Review Board (IRB) of Nanjing Drum Tower Hospital (2022–238-02), Nanjing, China. All patients underwent testing for SFTSV RNA using real-time reverse transcription polymerase chain reaction, confirming SFTSV infection. The patient cohort comprised 80 males and 77 females, with an average age of 61.8 ± 11.1 years. Based on the prognosis of SFTS patients, they were categorized into two groups: survivors and non-survivors. Severe cases were defined as patients meeting any of the following criteria (Deng et al., 2013): multi-organ dysfunction, acute respiratory distress syndrome (ARDS), sepsis, disseminated intravascular coagulation (DIC), failure of one or more organs (such as heart failure, acute renal failure, or liver failure), infection-induced toxic shock, or death. A total of 157 individuals were recruited from the Physical Examination Center of Nanjing Drum Tower Hospital to serve as the control group. The healthy control group included 83 males and 74 females, with an average age of 60.7 ± 9.2 years. No significant variations were observed in terms of sex or age between the control group and the patient group. Standardized forms were utilized to gather clinical data from the electronic medical record system, encompassing demographic information, laboratory results, and clinical symptoms. The demographic and clinical characteristics of the study population are summarized in Table 1.

2.2. Detection and collection of clinical parameters

Cholesterol, high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), apolipoprotein AI (ApoAI), apolipoprotein B (ApoB), immunoglobulins (IgG, IgA, IgM), albumin (ALB), C-reactive protein (CRP), complement proteins C3 and C4, alanine aminotransferase (ALT), and aspartate aminotransferase (AST) levels, lactate dehydrogenase (LDH) along with other biochemical indicators, were measured using a biochemical analyzer (Beckman AU5400, Germany). CRP, IgG, IgA, IgM, C3, C4 were detected using the immunoturbidimetric method. Prothrombin time (PT), activated partial thromboplastin time (APTT), Fibrinogen and D-dimer were measured using Sysmex CS-5100 automated coagulation analyzer. Blood cell counts, hemoglobin, and other hematological parameters were determined using an automated hematology analyzer (Sysmex Corporation, Japan). All data were extracted from the hospital's laboratory information system and medical record system.

2.3. Statistical analysis

Statistical comparisons between SFTS patients and healthy controls were conducted using Student's t-test. Differences were deemed statistically significant if the p-value was less than 0.05 ($p < 0.05$). All data measurements are presented as mean ± standard

TABLE 1 Baseline characteristics for patients with SFTS.

Parameters	Survival	Non-survival	*p*	Mild symptoms	Severe symptoms	*p*
No.	129	28	–	101	56	–
Male/Female (n)	66/63	14/14	–	51/50	29/27	–
Age (years)	60.8 ± 11.4	66.5 ± 8.3	0.01	60.6 ± 11.6	64.1 ± 9.8	0.06
BMI (kg/m^2)	23.6 ± 3.3	23.5 ± 3.4	0.94	23.6 ± 3.3	23.4 ± 3.3	0.90
Days of hospital stay	11.5 ± 6.5	7.2 ± 7.1	0.002	11.1 ± 6.8	10.1 ± 6.9	0.36
Time from onset to admission (days)	11.2 ± 9.3	10.7 ± 11.2	0.81	11.8 ± 10.4	9.8 ± 8.1	0.23
History *n* (%)						
Hypertension	42 (33%)	12 (43%)	0.29	35 (35%)	19 (34%)	0.93
Diabetes	10 (8%)	5 (18%)	0.09	10 (10%)	5 (9%)	0.84
Cardiovascular disease	51 (40%)	13 (46%)	0.50	31 (31%)	33 (59%)	0.0006
Cerebrovascular disease	31 (24%)	15 (54%)	0.001	21 (21%)	25 (45%)	0.002
kidney disease	38 (29%)	17 (61%)	0.002	28 (28%)	27 (48%)	0.003
Liver Disease	108 (84%)	27 (96%)	0.08	84 (83%)	51 (91%)	0.17
Cancer	3 (2%)	0 (0%)	0.42	3 (3%)	0 (0%)	0.19
Laboratory findings						
WBC ($\times10^9$/L)	5.4 ± 4.2	6.1 ± 4.2	0.41	5.4 ± 4.1	5.7 ± 4.5	0.61
NEU ($\times10^9$/L)	3.9 ± 4.3	3.9 ± 3.8	0.95	3.9 ± 4.3	3.9 ± 4.2	0.96
LYM ($\times10^9$/L)	1.2 ± 1.1	1.0 ± 0.7	0.35	1.2 ± 1.2	1.0 ± 0.6	0.26
HGB (g/L)	122.3 ± 22.0	124.6 ± 22.1	0.61	122.4 ± 22.5	123.3 ± 21.1	0.79
PLT ($\times10^9$/L)	104.6 ± 88.1	52.0 ± 48.7	0.003	108.0 ± 86.5	72.5 ± 77.1	0.01
ALT (U/L)	114.8 ± 188.6	96.2 ± 80.8	0.61	113.2 ± 206.2	108.4 ± 93.6	0.87
AST (U/L)	177.1 ± 194.8	425.6 ± 363.6	<0.0001	167.0 ± 179.6	319.6 ± 324.3	0.0002
LDH (U/L)	711.8 ± 552.8	2096 ± 2025	<0.0001	723.2 ± 579.8	1,370 ± 1,610	0.0004
ALB (g/L)	33.9 ± 4.3	29.2 ± 3.8	<0.0001	34.5 ± 4.3	30.7 ± 4.1	<0.0001
GLB (g/L)	28.8 ± 7.8	27.9 ± 6.1	0.56	28.2 ± 7.7	29.5 ± 7.2	0.29
CRP (mg/L)	12.1 ± 22.2	34.1 ± 40.9	0.0001	9.7 ± 16.6	27.3 ± 38.3	0.0001
IL-6 (pg/ml)	30.3 ± 63.8	878.1 ± 2,335	0.02	21.6 ± 23.9	542.4 ± 1822	0.12

The data was presented as the mean ± standard deviation (SD). BMI (Body Mass Index), WBC (White Blood Cell), NEU (Neutrophil), LYM (LYM), HGB (Hemoglobin), PLT (Platelets), ALT (Alanine Aminotransferase), AST (Aspartate Aminotransferase), LDH (Lactate Dehydrogenase), ALB (Albumin), GLB (Globulin), CRP (C-Reactive Protein).

deviation (m ± S). Pearson correlation analysis was employed to explore the relationships among clinical parameters, with correlations considered significant if the r value was greater than 0.2 and the p-value was less than 0.05 ($r > 0.2$, $p < 0.05$). Following univariate analysis, variables exhibiting statistical significance were chosen for subsequent multivariate logistic regression analysis. The outcomes of both univariate and multivariate regression analyses are presented as hazard ratios (HR) along with corresponding 95% confidence intervals (CI). A significance level of $p < 0.05$ was employed to determine statistical significance. Statistical analysis and data visualization were performed using SPSS 29.0 and GraphPad Prism 8.3 software.

3. Results

3.1. Clinical characteristics of the patients

This study included 157 SFTS patients, comprising 129 (82%) survivors and 28 (18%) non-survivors, with 56 (36%) classified as severe cases and 101 (64%) as mild cases. A comprehensive overview of the clinical characteristics of these patients is presented in Table 1. In this study, it was observed that non-survivors were characterized by advanced age compared to survivors. Additionally, the non-survivor group exhibited a significantly shorter duration of hospital stay compared to the survivor group. A higher proportion of non-survivors had a medical history of cerebrovascular disease and renal disease, in contrast to the survivor group. However, no significant differences were found between the survivor and non-survivor groups regarding the presence of hypertension and diabetes. In terms of clinical laboratory parameters, significant differences were noted in platelet count (PLT), AST, LDH, ALB, and CRP levels between the survivor group and the non-survivor group, as well as between the mild disease group and the severe disease group.

3.2. Altered lipid metabolism patterns observed in SFTS patients

To investigate the levels and clinical significance of lipids in SFTS patients, we conducted a comprehensive analysis of serum lipids including HDL-c, LDL-c, cholesterol, triglycerides, APoAI, and ApoB in both SFTS patients and healthy controls. Our findings reveal substantial deviations in lipid profiles among SFTS patients compared to healthy controls. Specifically, we observed significant reductions in HDL-c, LDL-c, cholesterol, APoAI, and ApoB levels in the serum of SFTS patients. Conversely, triglyceride levels were significantly elevated (Figure 1). These alterations in lipid metabolism unequivocally signify a severe disturbance in the patient's lipid profile following SFTSV infection.

3.3. Significant decrease in HDL-c levels observed in severe and non-survivor SFTS patients

To investigate the clinical significance of lipid metabolism-related indicators in SFTS patients, we conducted further analysis to compare lipid levels between severe and mild patients, as well as between survivors and non-survivors. Our findings reveal distinct patterns in lipid profiles within these patient groups. In severe patients, both HDL-c and LDL-c levels were significantly lower compared to those in mild patients, with the most noticeable decrease observed in HDL-c levels. Correspondingly, the main apolipoprotein of HDL, APoAI, also exhibited a significant decrease in critically ill patients. Conversely, no significant differences were observed in cholesterol, triglycerides, and APoB levels between the two groups (Figure 2). Additionally, we examined lipid levels in surviving patients and non-surviving patients. The results indicate significant reductions in HDL-c, LDL-c, cholesterol, APoAI, and APoB levels among non-surviving patients, while no significant difference was observed in triglyceride levels between the two groups (Supplementary Figure S1).

3.4. Prognostic potential of lipid metabolism parameters in predicting SFTS patient outcomes

To investigate the predictive performance of lipid metabolism parameters as biomarkers for SFTS prognosis, we conducted ROC curve analysis to assess their ability to distinguish between severe and mild cases, as well as predict mortality versus survival. The results indicated that the effectiveness of lipid metabolism parameters in differentiating between severe and mild conditions was not remarkable. Among these parameters, HDL-c and ApoAI exhibited the highest performance, with an area under the curve (AUC) of 0.69 (Figure 3A). However, when used as markers to predict survival and non-survival, HDL-c and ApoAI demonstrated greater discriminatory ability, with AUCs of 0.87 and 0.84, respectively (Figure 3B). This suggests that HDL is the most suitable biomarker for assessing the prognosis of SFTS patients among the lipid metabolism parameters. Using a cutoff value of 0.71, it achieves a sensitivity of 82.1% and specificity of 77.5%.

3.5. Analyzing prognostic factors for poor outcomes in SFTS patients

We further evaluated the laboratory parameters influencing the poor prognosis in SFTS patients using univariate and multivariate regression analysis. As shown in Table 2, the univariate regression analysis revealed

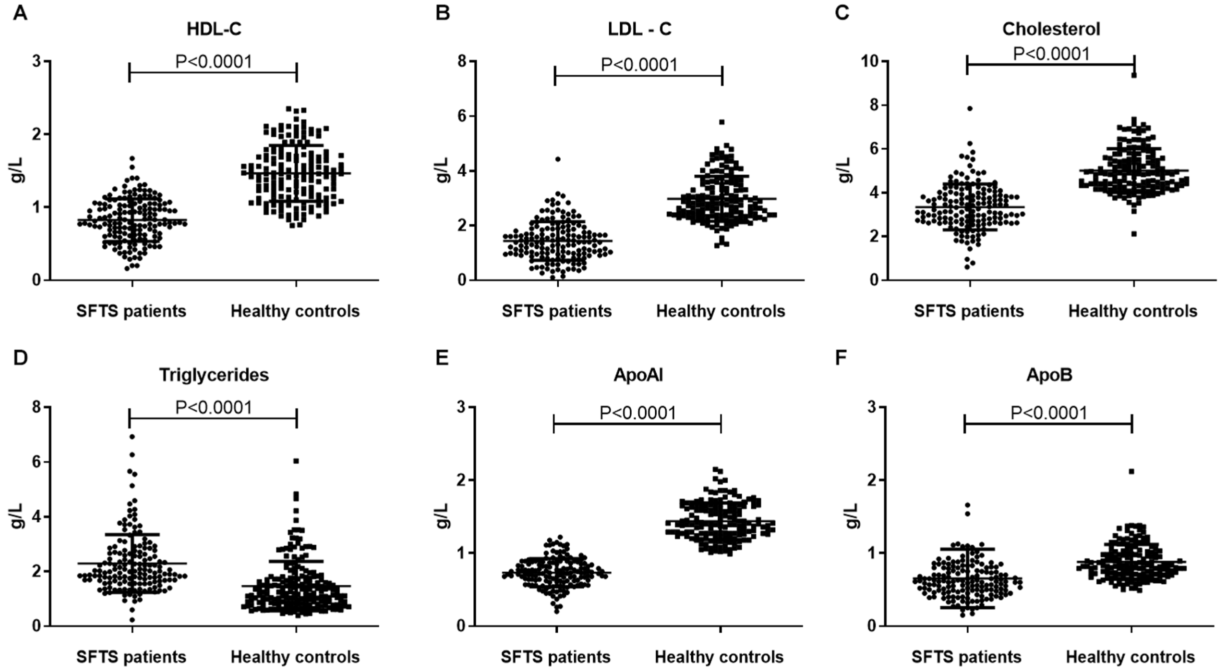


FIGURE 1
Comparative analysis of lipid profiles between SFTS patients and healthy controls. Comparison of serum lipid profiles (HDL-c, LDL-c, Cholesterol, Triglycerides, APoAI, and ApoB) between SFTS patients ($n = 157$) and healthy controls ($n = 157$) **(A–F)**. Data are expressed using mean ± standard deviation.

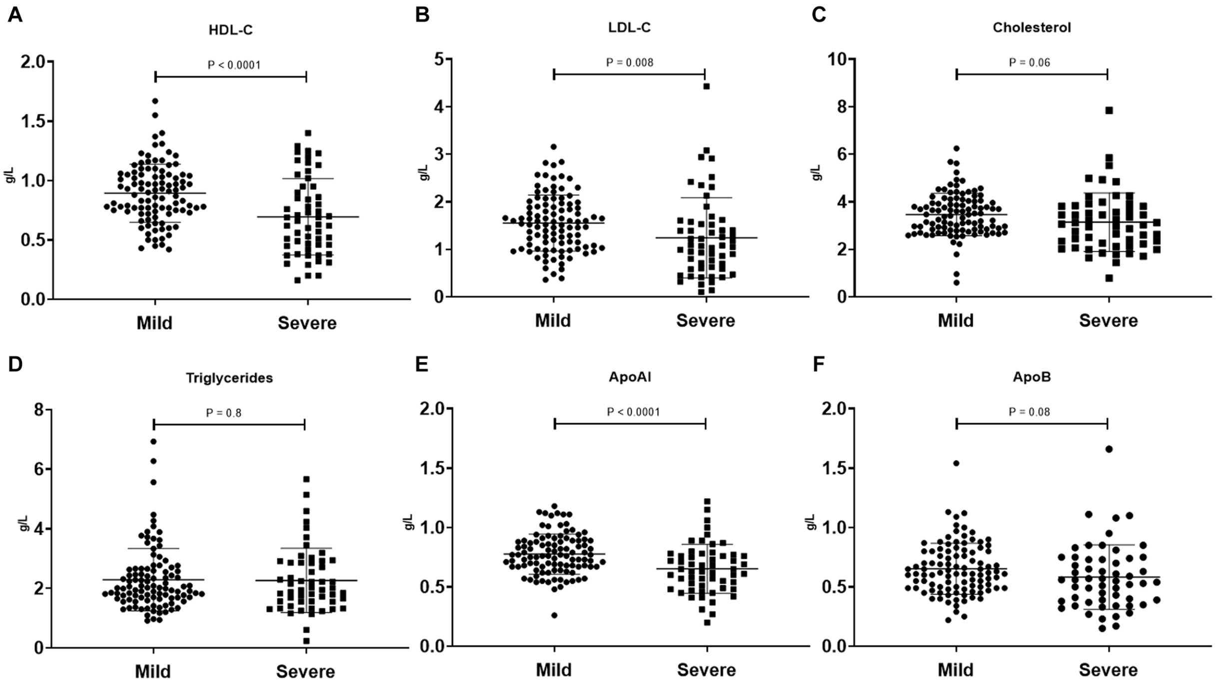


FIGURE 2
Comparison of serum lipid profiles (HDL-c, LDL-c, Cholesterol, Triglycerides, APoAI, and ApoB) between mild SFTS patients (*n* = 101) and severe SFTS patients (*n* = 56) **(A–F)**. Data are expressed using mean ± standard deviation.

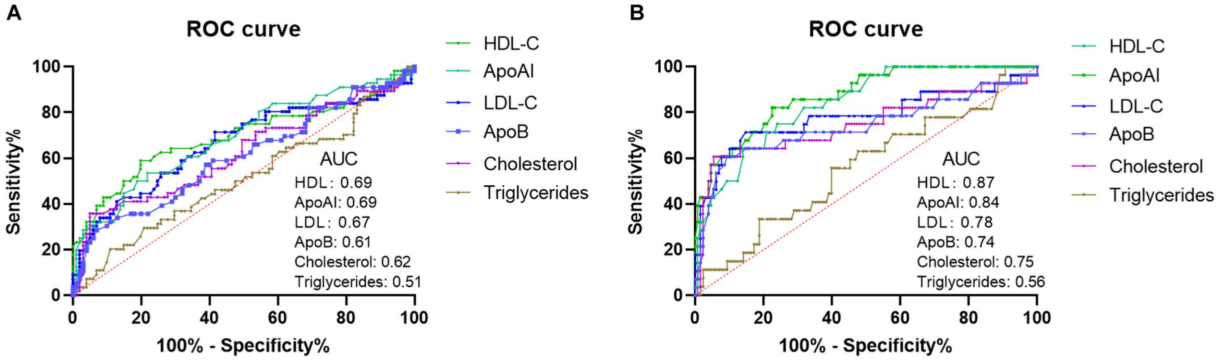


FIGURE 3
Analysis of ROC curves to assess the discriminatory ability of lipid profiles (HDL-c, LDL-c, Cholesterol, Triglycerides, APoAI, and ApoB) in distinguishing mild SFTS patients from severe SFTS patients **(A)**. Analysis of ROC curves to assess the ability of lipid profiles in distinguishing SFTS survivors from SFTS non-survivors **(B)**.

significant associations between outcomes in SFTS patients and Age, HDL-c, PLT, AST, LDH, Cholesterol, LDL-c, CRP, PT, APTT, Fibrinogen, D-dimer, and Albumin. Subsequently, these factors were included in the multivariate regression analysis, which identified HDL-c as an independent protective factor for the prognosis of SFTS patients.

3.6. Dynamic changes in HDL-c levels among non-survivors and recovered patients

To further explore the association between HDL-c levels and the prognosis of SFTS patients, we examined the dynamic changes of HDL-c levels in both non-survivors and survivors. Among the non-survivor group, the study findings revealed that HDL-c levels in most non-surviving patients exhibited a decline as the disease progressed, with follow-up HDL-c levels significantly lower than the admission levels (Figure 4A). Conversely, within the survivor group, a notable rise in HDL-c levels was observed as the patients' condition improved, with statistically significant results (Figure 4B). In addition, we also dynamically analyzed the HDL-c levels of patients in the survivor group and non-survivor group according to the days after admission. The results showed that the level of HDL-c in the non-survivor group continued to decline after hospitalization (Supplementary Figure S2).

TABLE 2 Univariate and multivariate regression analysis of laboratory parameters in SFTS patients and their prognostic implications.

Laboratory parameters	Univariable analysis		Multivariable analysis	
	HR (95% CI)	*p* value	HR (95% CI)	*p* value
Age	1.053 (1.010–1.098)	0.015	1.050 (0.962–1.146)	0.277
HDL-c	0.000 (0.000–0.009)	<0.001	0.000 (0.000–0.008)	0.007
PLT	0.984 (0.972–0,995)	0.005	0.983 (0.955–1.011)	0.220
AST	1.003 (1.002–1.005)	<0.001	0.998 (0.988–1.007)	0.631
LDH	1.001 (1.001–1.002)	<0.001	1.000 (0.998–1.002)	0.995
Cholesterol	0.434 (0.263–0.715)	0.001	2.052 (0.361–11.652)	0.417
LDL-c	0.204 (0.089–0.469)	<0.001	0.168 (0.005–5.271)	0.311
CRP	1.022 (1.009–1.036)	0.001	0.971 (0.933–1.010)	0.144
PT	2.798 (1.742–4.495)	<0.001	2.826 (0.729–10.963)	0.133
APTT	1.131 (1.080–1.185)	<0.001	1.034 (0.952–1.124)	0.425
Fibrinogen	0.497 (0.285–0.868)	0.014	1.719 (0.481–6.138)	0.404
D-dimer	1.15 (1.062–1.245)	<0.001	1.104 (0.971–1.254)	0.130
Albumin	0.744 (0.656–0.843)	<0.001	0.907 (0.687–1.198)	0.493

HR (hazard ratio), HDL-c (high-density lipoprotein cholesterol), PLT (Platelets), AST (Aspartate Aminotransferase), LDH (Lactate Dehydrogenase), LDL-c (high-density lipoprotein cholesterol), CRP (C-Reactive Protein), PT (Prothrombin time), APTT (activated partial thromboplastin time).

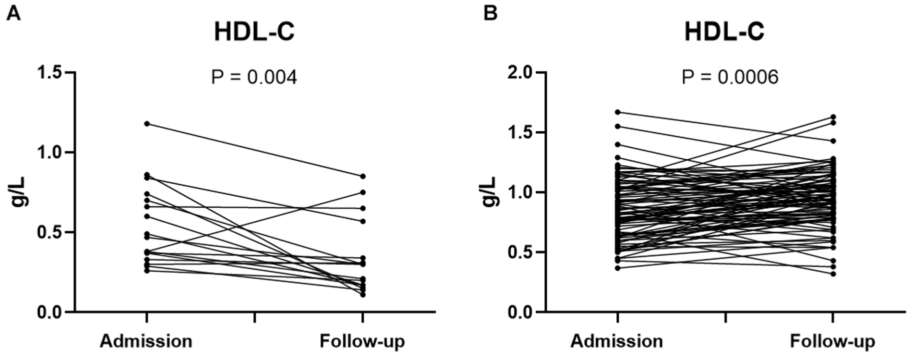


FIGURE 4
Dynamic changes of HDL-c level in admission and follow-up of SFTS non-survivors (*n* = 17) **(A)**. Dynamic changes of HDL-c level in SFTS surviving patients at admission and follow-up (*n* = 89) **(B)**. Statistical analysis using paired *T* test. Exclusion of patients with single HDL-c test from the analysis.

3.7. Correlation analysis between HDL-c levels and clinical laboratory parameters

During viral infections, several clinical laboratory parameters undergo significant changes, which are believed to be correlated with disease severity. To further investigate the clinical significance of HDL-c in SFTS patients, we conducted an in-depth analysis to explore the correlation between HDL-c levels and key clinical laboratory parameters. The results revealed a significant negative correlation between HDL-c and IL-6, CRP, AST, TT, APTT (Figures 5A–E), while demonstrating a significant positive correlation with PLT, ALB, and CD4 counts (Figures 5F–H). These findings provide additional evidence supporting the correlation between HDL-c levels and disease severity in SFTS patients.

4. Discussion

Various viral infections, including dengue virus, HIV, and SARS-CoV-2, have been associated with significant alterations in lipid profiles, which are considered prognostic indicators of disease (Maggi et al., 2017; Lima et al., 2019; Rezaei et al., 2022). In COVID-19 patients, LDL-C, HDL-C, and cholesterol levels were found to be decreased, particularly in severe and critically ill cases. Furthermore, among ICU-hospitalized COVID-19 patients, it has been observed that decreased HDL levels are linked to higher mortality rates (Wang et al., 2021).

The level and clinical significance of lipid profiles in SFTS patients have received limited attention. This study revealed significant disturbances in lipid metabolism hospitalized SFTS patients. Notably, LDL-c, HDL-c, and total cholesterol levels were significantly reduced, while triglyceride levels were significantly increased. These findings resemble the lipid metabolism disruptions observed in patients with HIV and COVID-19. Among critically ill patients, the most notable decrease was observed in HDL-c levels, and ROC curve analysis identified HDL-c as the most efficient marker in distinguishing non-survivors from survivors. Consequently, our research primarily focused on investigating HDL. We found a higher prevalence of cerebrovascular disease in severe or non-survivor patients, potentially due to immune and blood coagulation imbalances that worsen the

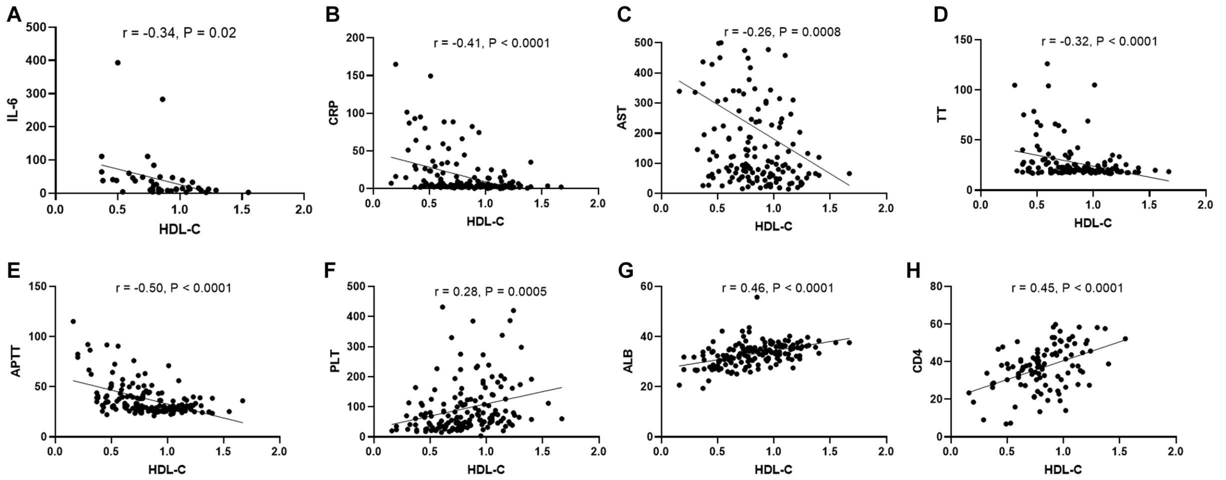


FIGURE 5
Correlation between HDL-c level and other laboratory parameters in SFTS patients. Correlation analysis of serum HDL-c level with IL-6, CRP, AST, TT, APTT，PLT, ALB, and CD4 counts in SFTS patients, respectively **(A–H)**.

condition. Univariate analysis linked age, HDL-c, PLT, and AST to patient prognosis. However, multivariate analysis, excluding other factors, highlighted HDL-c as an independent protective factor consistently associated with patient prognosis. We conducted a dynamic analysis of patients' HDL-c levels. Among non-surviving patients, HDL-c levels did not increase but rather exhibited a decreasing trend as hospitalization time progressed. While there were no statistically significant differences between each time point, this might be attributed to the fact that not all patients underwent multiple HDL-c tests. Additionally, variations in the duration of hospital stays could have led to some mutual offsetting in HDL-c levels calculated based on days since admission. Our paired analysis, focusing on admission and follow-up as two time points, revealed a significant increase in HDL-c levels among survivors, whereas HDL-c levels in non-survivors notably decreased. Furthermore, HDL-c displayed significant correlations with several important clinical laboratory parameters, providing insights into the disease severity. Therefore, we propose that HDL-c holds potential as a prognostic marker for assessing the prognosis of SFTS patients.

HDL possesses multiple functions that contribute to its protective effects. One of its main functions is reverse cholesterol transport, where the protein ApoAI binds with free cholesterol in tissue cells and transports it to the liver, thus reducing overall cholesterol levels and delaying the onset and progression of coronary heart disease (Tosheska Trajkovska and Topuzovska, 2017). Furthermore, HDL has been shown to prevent systemic endotoxemia by binding and neutralizing lipopolysaccharide (LPS), serving as a crucial mechanism for its antibacterial effect (Parker et al., 1995). HDL also exerts anti-inflammatory effects by downregulating Toll-like receptor (TLR)-induced pro-inflammatory cytokines through the transcriptional regulator activating transcription factor 3 (ATF3) (de Nardo et al., 2014). Additionally, HDL exhibits antioxidant and anti-apoptotic properties (Nofer et al., 2001; Robbesyn et al., 2003). Given these functions, HDL can provide protective benefits for patients. In SFTS patients, we observed lower levels of HDL in severe and non-survivor patients, whereas mild and recovered patients exhibited higher HDL-c levels. Notably, HDL-c demonstrated a significant negative correlation with inflammatory indicators such as IL-6, CRP, and AST. These findings further suggest that HDL may play a protective role in SFTS patients.

The decrease in HDL-c levels during infection can be attributed to several mechanisms. Studies have indicated that pro-inflammatory cytokines like IL-6 can directly inhibit the activity of apolipoprotein synthase, leading to a reduction in HDL-C and apoA-1 production (Pirillo et al., 2015). Since HDL-c exerts anti-inflammatory effects, the deficiency of HDL-C exacerbates cytokine overproduction and further depletion of HDL-C. Inflammatory cytokines, such as TNF, can also decrease the activity of LCAT, resulting in reduced cholesterol lipid accumulation within HDL (Ly et al., 1995). Moreover, the downregulation of ABCA1 and ABCG1 by endotoxins can contribute to decreased HDL cholesterol levels (Schmitz and Langmann, 2005).

Additionally, in SFTS patients, it is noteworthy that while cholesterol levels significantly decreased, triglyceride levels were significantly higher compared to healthy controls, consistent with observations in COVID-19 patients. This phenomenon may be attributed to increased hepatic VLDL production and secretion during infection, which stimulates triglyceride synthesis (Feingold et al., 1992). Simultaneously, the infection and inflammation may inhibit the activity of lipoprotein lipase, leading to reduced clearance of triglycerides.

In addition to its potential as a prognostic biomarker for SFTS patients, this study also introduces a novel treatment concept. Studies have indicated that Omega-3 polyunsaturated fatty acids (PUFA) can improve lipid metabolism by reducing triglyceride levels and increasing HDL levels, while also reducing inflammatory responses (Ni et al., 2020). This approach has shown positive effects in the treatment of COVID-19 (Sorokin et al., 2020; Faguer et al., 2022). Furthermore, the use of statins to increase HDL levels is also a current research focus in the context of COVID-19 (Vahedian-Azimi et al., 2021). Presently, treatment options for SFTS patients are limited. Taking into account the potential protective role of HDL in SFTS patients, increasing HDL levels could potentially offer novel therapeutic strategies for managing SFTS. Of course, this requires further investigation into the biological mechanisms and treatment effects, through additional animal experiments and clinical trial data. This research direction is also highly meaningful.

In conclusion, this study revealed significant abnormalities in the serum lipid profile of SFTS patients, with a notable decrease in HDL-c levels observed in severe and deceased patients. HDL-c emerges as a potential biomarker for predicting poor prognosis in SFTS patients. These findings suggest that elevating HDL levels could be a novel therapeutic strategy for the treatment of SFTS patients.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Institutional Review Board (IRB) of Nanjing Drum Tower Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

SW, YC, and TH contributed to the design of the study and supervised the scientific work. YF, YX, XX, XC, and HY contributed to the analysis and interpretation of the data. TH and YF drafted the manuscript and SW revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grants from Nanjing Medical Science and technique Development Foundation (QRX17142, YKK21066), Clinical Trials from the Affiliated Drum Tower Hospital, Medical School of Nanjing University (2022-LCYJ-PY-40). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1239420/full#supplementary-material

References

Bassendine, M. F., Sheridan, D. A., Bridge, S. H., Felmlee, D. J., and Neely, R. D. (2013). Lipids and HCV. *Semin. Immunopathol.* 35, 87–100. doi: 10.1007/s00281-012-0356-2

Chen, M. Y., Meng, X. F., Han, Y. P., Yan, J. L., Xiao, C., and Qian, L. B. (2022). Profile of crosstalk between glucose and lipid metabolic disturbance and diabetic cardiomyopathy: inflammation and oxidative stress. *Front Endocrinol (Lausanne).* 13:983713. doi: 10.3389/fendo.2022.983713

de Nardo, D., Labzin, L. I., Kono, H., Seki, R., Schmidt, S. V., Beyer, M., et al. (2014). High-density lipoprotein mediates anti-inflammatory reprogramming of macrophages via the transcriptional regulator ATF3. *Nat. Immunol.* 15, 152–160. doi: 10.1038/ni.2784

Deng, B., Zhou, B., Zhang, S., Zhu, Y., Han, L., Geng, Y., et al. (2013). Clinical features and factors associated with severity and fatality among patients with severe fever with thrombocytopenia syndrome Bunyavirus infection in Northeast China. *PLoS One* 8:e80802. doi: 10.1371/journal.pone.0080802

Faguer, S., Del Bello, A., Danet, C., Renaudineau, Y., Izopet, J., and Kamar, N. (2022). Apolipoprotein-A-I for severe COVID-19-induced hyperinflammatory states: a prospective case study. *Front. Pharmacol.* 13:936659. doi: 10.3389/fphar.2022.936659

Feingold, K. R., Staprans, I., Memon, R. A., Moser, A. H., Shigenaga, J. K., Doerrler, W., et al. (1992). Endotoxin rapidly induces changes in lipid metabolism that produce hypertriglyceridemia: low doses stimulate hepatic triglyceride production while high doses inhibit clearance. *J. Lipid Res.* 33, 1765–1776. doi: 10.1016/S0022-2275(20)41334-3

Grunfeld, C., Marshall, M., Shigenaga, J. K., Moser, A. H., Tobias, P., and Feingold, K. R. (1999). Lipoproteins inhibit macrophage activation by lipoteichoic acid. *J. Lipid Res.* 40, 245–252. doi: 10.1016/S0022-2275(20)33363-0

Kulasegaram, R., Peters, B. S., and Wierzbicki, A. S. (2005). Dyslipidaemia and cardiovascular risk in HIV infection. *Curr. Med. Res. Opin.* 21, 1717–1725. doi: 10.1185/030079905X65529

Lima, W. G., Souza, N. A., Fernandes, S., Cardoso, V. N., and Godoi, I. P. (2019). Serum lipid profile as a predictor of dengue severity: a systematic review and meta-analysis. *Rev. Med. Virol.* 29:e2056. doi: 10.1002/rmv.2056

Luo, L. M., Zhao, L., Wen, H. L., Zhang, Z. T., Liu, J. W., Fang, L. Z., et al. (2015). *Haemaphysalis longicornis* ticks as reservoir and vector of severe fever with thrombocytopenia syndrome virus in China. *Emerg. Infect. Dis.* 21, 1770–1776. doi: 10.3201/eid2110.150126

Ly, H., Francone, O. L., Fielding, C. J., Shigenaga, J. K., Moser, A. H., Grunfeld, C., et al. (1995). Endotoxin and TNF lead to reduced plasma LCAT activity and decreased hepatic LCAT mRNA levels in Syrian hamsters. *J. Lipid Res.* 36, 1254–1263. doi: 10.1016/S0022-2275(20)41133-2

Maggi, P., Di Biagio, A., Rusconi, S., Cicalini, S., D'Abbraccio, M., D'Ettorre, G., et al. (2017). Cardiovascular risk and dyslipidemia among persons living with HIV: a review. *BMC Infect. Dis.* 17:551. doi: 10.1186/s12879-017-2626-z

Masana, L., Correig, E., Ibarretxe, D., Anoro, E., Arroyo, J. A., Jerico, C., et al. (2021). Low HDL and high triglycerides predict COVID-19 severity. *Sci. Rep.* 11:7217. doi: 10.1038/s41598-021-86747-5

Murphy, A. J., Woollard, K. J., Hoang, A., Mukhamedova, N., Stirzaker, R. A., McCormick, S. P., et al. (2008). High-density lipoprotein reduces the human monocyte inflammatory response. *Arterioscler. Thromb. Vasc. Biol.* 28, 2071–2077. doi: 10.1161/ATVBAHA.108.168690

Ni, C., Cao, J., Li, D., Wu, W., Cao, L., and Zhu, C. (2020). Parenteral nutrition effects of Omega-3 fatty acids on C-reactive protein, high-density lipoprotein, lymphocyte characteristics and the treatment of critically ill patients. *Cell. Mol. Biol.* 66, 52–56. doi: 10.14715/cmb/2020.66.3.8

Nofer, J. R., Levkau, B., Wolinska, I., Junker, R., Fobker, M., von Eckardstein, A., et al. (2001). Suppression of endothelial cell apoptosis by high density lipoproteins (HDL) and HDL-associated lysosphingolipids. *J. Biol. Chem.* 276, 34480–34485. doi: 10.1074/jbc.M103782200

Parker, T. S., Levine, D. M., Chang, J. C., Laxer, J., Coffin, C. C., and Rubin, A. L. (1995). Reconstituted high-density lipoprotein neutralizes gram-negative bacterial lipopolysaccharides in human whole blood. *Infect. Immun.* 63, 253–258. doi: 10.1128/iai.63.1.253-258.1995

Pirillo, A., Catapano, A. L., and Norata, G. D. (2015). HDL in infectious diseases and sepsis. *Handb. Exp. Pharmacol.* 224:483. doi: 10.1007/978-3-319-09665-0_15

Ren, F., Shen, S., Ning, Y., Wang, Q., Dai, S., Shi, J., et al. (2021). Non-structural proteins of severe fever with thrombocytopenia syndrome virus suppress RNA synthesis in a transcriptionally active cDNA-derived viral RNA synthesis system. *Front. Microbiol.* 12:709517. doi: 10.3389/fmicb.2021.709517

Rezaei, A., Neshat, S., and Heshmat-Ghahdarijani, K. (2022). Alterations of lipid profile in COVID-19: a narrative review. *Curr. Probl. Cardiol.* 47:100907. doi: 10.1016/j.cpcardiol.2021.100907

Robbesyn, F., Garcia, V., Auge, N., Vieira, O., Frisach, M. F., Salvayre, R., et al. (2003). HDL counterbalance the proinflammatory effect of oxidized LDL by inhibiting intracellular reactive oxygen species rise, proteasome activation, and subsequent NF-kappaB activation in smooth muscle cells. *FASEB J.* 17, 743–745. doi: 10.1096/fj.02-0240fje

Schmitz, G., and Langmann, T. (2005). Transcriptional regulatory networks in lipid metabolism control ABCA1 expression. *Biochim. Biophys. Acta* 1735, 1–19. doi: 10.1016/j.bbalip.2005.04.004

Sorokin, A. V., Karathanasis, S. K., Yang, Z. H., Freeman, L., Kotani, K., and Remaley, A. T. (2020). COVID-19-associated dyslipidemia: implications for mechanism of impaired resolution and novel therapeutic approaches. *FASEB J.* 34, 9843–9853. doi: 10.1096/fj.202001451

Sun, J. T., Chen, Z., Nie, P., Ge, H., Shen, L., Yang, F., et al. (2020). Lipid profile features and their associations with disease severity and mortality in patients with COVID-19. *Front Cardiovasc Med.* 7:584987. doi: 10.3389/fcvm.2020.584987

Tenenbaum, A., Fisman, E. Z., Motro, M., and Adler, Y. (2006). Atherogenic dyslipidemia in metabolic syndrome and type 2 diabetes: therapeutic options beyond statins. *Cardiovasc. Diabetol.* 5:20. doi: 10.1186/1475-2840-5-20

Tosheska Trajkovska, K., and Topuzovska, S. (2017). High-density lipoprotein metabolism and reverse cholesterol transport: strategies for raising HDL cholesterol. *Anatol. J. Cardiol.* 18, 149–154. doi: 10.14744/AnatolJCardiol.2017.7608

Vahedian-Azimi, A., Mohammadi, S. M., Banach, M., Beni, F. H., Guest, P. C., Al-Rasadi, K., et al. (2021). Improved COVID-19 outcomes following statin therapy: an updated systematic review and meta-analysis. *Biomed. Res. Int.* 2021:1901772. doi: 10.1155/2021/1901772

Wang, G., Deng, J., Li, J., Wu, C., Dong, H., Wu, S., et al. (2021). The role of high-density lipoprotein in COVID-19. *Front. Pharmacol.* 12:720283. doi: 10.3389/fphar.2021.720283

Wang, Y., Song, Z., Xu, X., Wei, X., Yuan, H., Liang, H., et al. (2022). Clinical symptoms associated with fatality of severe fever with thrombocytopenia syndrome: a systematic review and meta-analysis. *Acta Trop.* 232:106481. doi: 10.1016/j.actatropica.2022.106481

Yoshikawa, T., Shimojima, M., Fukushi, S., Tani, H., Fukuma, A., Taniguchi, S., et al. (2015). Phylogenetic and geographic relationships of severe fever with thrombocytopenia syndrome virus in China, South Korea, and Japan. *J. Infect. Dis.* 212, 889–898. doi: 10.1093/infdis/jiv144

Yu, X. J., Liang, M. F., Zhang, S. Y., Liu, Y., Li, J. D., Sun, Y. L., et al. (2011). Fever with thrombocytopenia associated with a novel bunyavirus in China. *N. Engl. J. Med.* 364, 1523–1532. doi: 10.1056/NEJMoa1010095

Yun, S. M., Park, S. J., Park, S. W., Choi, W., Jeong, H. W., Choi, Y. K., et al. (2017). Molecular genomic characterization of tick-and human-derived severe fever with thrombocytopenia syndrome virus isolates from South Korea. *PLoS Negl. Trop. Dis.* 11:e5893. doi: 10.1371/journal.pntd.0005893

Zhao, J., Lu, Q., Li, H., Yuan, Y., Cui, N., Yuan, C., et al. (2021). Sex differences in case fatality rate of patients with severe fever with thrombocytopenia syndrome. *Front. Microbiol.* 12:738808. doi: 10.3389/fmicb.2021.738808

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 08 November 2023
DOI 10.3389/fmicb.2023.1249794

Willow (*Salix* spp.) bark hot water extracts inhibit both enveloped and non-enveloped viruses: study on its anti-coronavirus and anti-enterovirus activities

Dhanik Reshamwala[1], Sailee Shroff[1], Jaana Liimatainen[2], Jenni Tienaho[2], Mira Laajala[1], Petri Kilpeläinen[2], Anneli Viherä-Aarnio[2], Maarit Karonen[3], Tuula Jyske[2] and Varpu Marjomäki[1]*

[1]Department of Biological and Environmental Sciences, Nanoscience Center, University of Jyväskylä, Jyväskylä, Finland, [2]Natural Resources Institute Finland (Luke), Helsinki, Finland, [3]Natural Chemistry Research Group, Department of Chemistry, University of Turku, Turku, Finland

Introduction: Recurring viral outbreaks have a significant negative impact on society. This creates a need to develop novel strategies to complement the existing antiviral approaches. There is a need for safe and sustainable antiviral solutions derived from nature.

Objective: This study aimed to investigate the antiviral potential of willow (*Salix* spp.) bark hot water extracts against coronaviruses and enteroviruses. Willow bark has long been recognized for its medicinal properties and has been used in traditional medicines. However, its potential as a broad-spectrum antiviral agent remains relatively unexplored.

Methods: Cytopathic effect inhibition assay and virucidal and qPCR-based assays were used to evaluate the antiviral potential of the bark extracts. The mechanism of action was investigated using time-of-addition assay, confocal microscopy, TEM, thermal, and binding assays. Extracts were fractionated and screened for their chemical composition using high-resolution LC-MS.

Results: The native *Salix* samples demonstrated their excellent antiviral potential against the non-enveloped enteroviruses even at room temperature and after 45 s. They were equally effective against the seasonal and pandemic coronaviruses. Confocal microscopy verified the loss of infection capacity by negligible staining of the newly synthesized capsid or spike proteins. Time-of-addition studies demonstrated that *Salix* bark extract had a direct effect on the virus particles but not through cellular targets. Negative stain TEM and thermal assay showed that antiviral action on enteroviruses was based on the added stability of the virions. In contrast, *Salix* bark extract caused visible changes in the coronavirus structure, which was demonstrated by the negative stain TEM. However, the binding to the cells was not affected, as verified by the qPCR study. Furthermore, coronavirus accumulated in the cellular endosomes and did not proceed after this stage, based on the confocal studies. None of the tested commercial reference samples, such as salicin, salicylic acid, picein, and triandrin, had any antiviral activity. Fractionation of the extract and subsequent MS analysis revealed that most of the separated fractions were very effective against enteroviruses and contained several different chemical groups such as hydroxycinnamic acid derivatives, flavonoids, and procyanidins.

OPEN ACCESS

EDITED BY
Shailendra Saxena,
King George's Medical University, India

REVIEWED BY
Ángel L. Álvarez,
University of Oviedo, Spain
Manoj Baranwal,
Thapar Institute of Engineering & Technology, India

*CORRESPONDENCE
Varpu Marjomäki
✉ varpu.s.marjomaki@jyu.fi

RECEIVED 29 June 2023
ACCEPTED 18 September 2023
PUBLISHED 08 November 2023

CITATION
Reshamwala D, Shroff S, Liimatainen J, Tienaho J, Laajala M, Kilpeläinen P, Viherä-Aarnio A, Karonen M, Jyske T and Marjomäki V (2023) Willow (*Salix* spp.) bark hot water extracts inhibit both enveloped and non-enveloped viruses: study on its anti-coronavirus and anti-enterovirus activities. *Front. Microbiol.* 14:1249794. doi: 10.3389/fmicb.2023.1249794

Conclusion: *Salix* spp. bark extracts contain several virucidal agents that are likely to act synergistically and directly on the viruses.

KEYWORDS

antivirals, nature-based enteroviruses, coronaviruses, *Salix* spp., broad-spectrum

1. Introduction

The emergence of viral outbreaks leading to epidemics and pandemics causes a huge strain on the global economy and public health. The recent pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has been a catastrophic event, and as of June 2022, it has caused over 6.9 million deaths worldwide (WHO, 2022). SARS-CoV-2 belongs to the β-coronavirus genus, which also includes SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV). It is an enveloped, positive-sense single-stranded RNA (+ssRNA) virus with a diameter of 60–140 nm (Pizzato et al., 2022). Transmission of SARS-CoV-2 occurs through aerosol, the fecal-oral route, and surface contamination (Karia et al., 2020; Zhou et al., 2021) to cause lower respiratory tract infections. The group of beta coronaviruses (HCoVs) also includes several seasonal coronaviruses that cause the common cold. HCoV strain OC43 is responsible for 15–30% of mild upper respiratory tract infections in humans (Gaunt et al., 2010). Both belong to the β-coronavirus genus and are closely related genetically (Lu et al., 2020). Moreover, both of these viruses replicate in the human respiratory epithelium and spread via aerosols and droplets (Kutter et al., 2018). Enteroviruses are also positive-sense single-stranded RNA (+ssRNA) viruses, but in many ways different from coronaviruses: They are non-enveloped, much smaller in size (30 nm in diameter), and stay very stable and infectious on surfaces and in the environment. They are responsible for causing acute infections such as flu, meningitis, pancreatitis, and myocarditis. They are also associated with chronic infections like type 1 diabetes (Nekoua et al., 2022), asthma, and chronic obstructive pulmonary disease (COPD) (Kurai et al., 2013). Enteroviruses include several serotypes that infect through the fecal–oral route but also viruses that infect through the respiratory route, such as rhinoviruses and enterovirus D68, also called the "new polio" (Cassidy et al., 2018).

Antiviral agents, such as vaccines, drugs, and virucides, help in reducing viral transmission. Virucides are used to reduce the viral load on the surface and in the environment. They are used as disinfectants for surface sterilization of biological and medicinal products. Additionally, they have been used to inactivate viruses in foodstuffs, detergents, and cosmetics (Galabov, 2007). However, the majority of the virucides are chemical disinfectants, which are hazardous in nature and cause environmental contamination. In addition, they cause side effects on human health, such as skin irritation. Moreover, non-enveloped viruses like enteroviruses are largely resistant to chemical disinfectants (Chan and Abu Bakar, 2005; Sauerbrei and Wutzler, 2010). Even though vaccines are an effective weapon against virus infection, it is not feasible to develop a vaccine against all the enteroviruses. In addition, the process of vaccine development and approval also takes time. Currently, there are no clinically approved drugs for enteroviruses. Thus, there is a great need to find broadly acting antiviral agents that would lower the infectivity of viruses around us and that could complement the vaccines and drugs in the combat against viruses. Natural products are a rich source of bioactive compounds. Out of 1,881 approved drugs from the start of January 1981 to the end of September 2019, a total of 41.8% are either biological macromolecules, unaltered natural products, botanical drugs, or natural product derivatives. If synthetic products mimicking natural compounds are also considered, the share increases to 64.3% (Newman and Cragg, 2020). Various natural products have been reported to exhibit antiviral activity, and they are an interesting source of novel antivirals because of their availability, tolerability, and expected low side effects (Kumar and Pandey, 2013; Goh et al., 2020). As an example, flavonoids are a diverse group of plant secondary metabolites known for their antioxidant, anti-inflammatory, anticarcinogenic, and other therapeutic properties (Kumar and Pandey, 2013; Panche et al., 2016). While often the mechanism of antiviral action remains unknown, natural products have been reported to interact with the viral life cycle by either targeting viral entry, replication, assembly, or release (Lin et al., 2014; Linnakoski et al., 2018).

In our previous study, we showed that *Salix* bark hot water extracts are highly effective against non-enveloped enteroviruses (Coxsackie virus A9) and not cytotoxic in the used concentrations (Tienaho et al., 2021). Interestingly, none of the tested reference compounds, such as triandrin, salicin, salicylic acid, or picein, showed antiviral activities, suggesting that the bioactive properties of *Salix* clone bark extracts could be due to the synergistic effects of different bioactive agents such as tannins and other polyphenols. In the present study, *Salix* bark hot water extracts were tested for their antiviral activity against HCoV-OC43 and SARS-CoV-2, and their mechanism of action was elucidated against the coronaviruses and for the previously tested enteroviruses. Bark extracts from most of the willow clones tested showed antiviral potency against both viruses by having a direct effect on the virus particles. The extracts caused clustering of both the viruses but halted infection in different ways for non-enveloped and enveloped viruses: through the increased stability of enteroviruses structure, but through the compromised structure of coronaviruses.

2. Methods

2.1. Cells

Human alveolar basal epithelial adenocarcinoma (A549), Vero E6, and human lung fibroblasts (MRC-5) cells were obtained from the American Type Culture Collection (ATCC) (Manassas, VA, USA). The A549 cells were propagated in Dulbecco's modified Eagle medium (DMEM) (Gibco, Paisley, UK), whereas MRC-5 and Vero E6 cell lines were grown in Eagle's minimum essential medium (MEM) (Gibco, Paisley, UK). Both MEM and DMEM were supplemented with 10% fetal bovine serum (FBS, Gibco, Paisley, UK), 1% L-GlutaMAX (Gibco, Paisley, UK), and 1% antibiotics (penicillin/streptomycin) (Gibco, Paisley, UK) and stored in a humidified 5% CO_2 incubator at 37°C.

2.2. Viruses

Coxsackievirus B3 (CVB3; Nancy strain) and Coxsackievirus A9 strain (CVA9; Griggs strain) were obtained from ATCC. They were produced and purified as described before (Myllynen et al., 2016; Ruokolainen et al., 2019), with one exception of adding 0.1% (v/v) TWEEN$^{®}$ 80 (Sigma-Aldrich, Steinheim, Germany) during the freeze–thaw cycle. Seasonal human coronavirus HCoV-OC43 (ATCC) was used as a crude or purified preparation. SARS-CoV-2 (SARS-CoV-2/Finland/1/2020) was isolated from the first COVID-19 patient in Finland (Haveri et al., 2020).

2.3. *Salix* sample collection and preparation

The *Salix* clone sample collection and extraction have been previously published by Tienaho and colleagues (Tienaho et al., 2021). For this study, 16 *Salix* clones were chosen (Table 1). Out of these, 12 were examples of widely distributed native Finnish species: three *S. myrsinifolia* Salisb. clones, four *S. phylicifolia* L. clones, and three natural and two artificial hybrids of these. Four clones originated from the Swedish willow breeding program. Detailed information on the growing media, growth coordinates, and handling was published earlier (Tienaho et al., 2021). In brief, the native willow clones were harvested in May 2019 as 2-year-old 1–1.5 m in length coppice with 0.5–2 cm diameter at the base, and commercial willow samples were grown by Carbons Finland Ltd. and cut down in March 2019 when 3-year-old and the 3-m-long sample shoots were cut to ca. 40-cm-long pieces. The harvested willows were packed separately in plastic bags and immediately frozen at −20°C before handling. Two shoots of each willow were debarked 50 cm from the base and pooled. The obtained bark was cut into small pieces, frozen at −80°C, freeze-dried before being ground with a Moulinex grinder into 1- to 2-mm pieces, and kept frozen at −80°C until extracted using an ASE-350 accelerated solvent extractor (Dionex, Sunnyvale, CA, United States). The bark sample (1 g) was placed in a stainless-steel extraction vessel (22 ml). The sample was then extracted three times for 15 min with hot water at 90°C, and the extract was stored at −20°C before further analyses. In addition, clone 16 was extracted in a larger-scale extraction vessel. For this experiment, willow clone 16 stand was cut down in October 2020 by Carbons Finland Ltd. from a willow bank that was partly harvested in 2018 and 2019. The shoots were debarked immediately and frozen at −20°C. The bark was ground with a Kamas cutting mill with a 2-cm sieve and extracted in a 2-L stirring reactor (Polyclave, Büchi, Switzerland) with hot water (80°C) (Tienaho et al., 2021). Total dissolved solids (TDS) of extracts varied from 4.1 to 8.6 mg/ml.

2.4. Extraction and fractionation of willow stem

For a pilot-scale extraction, 1-year-old shoots of the commercial willow variety, Klara, were harvested by Carbons Finland Ltd. The growing site was a peat field at Aitomäki, Kouvola, in south-eastern Finland (N60°52′0.01″ E26°41′60.00″). The shoots were cut in September 2022 and immediately transported to the piloting site in Bioruukki, Espoo, Finland. Whole willow shoots, without prior debarking, were milled with a shredder (Viking GE 150, VIKING GmbH). Shredded willow was collected into a bag and stored at −30°C before the extraction. The moisture content of the shredded material was 52.5 wt%, determined by oven drying at 105°C overnight.

For hot water extraction, a 64.8 kg batch (34.0 kg o.d.) of freshly shredded willow was added into a 300 L extraction system (Kilpeläinen et al., 2014). Water was pre-heated to 135°C to obtain the targeted 90°C extraction temperature. The average temperature during the extraction was 92°C, the pressure during the extraction was 10 bars, and the extraction time was 60 min. Extract (216 kg) was collected into an intermediate bulk container. Extract's total dissolved solids (TDS) was 1.19 wt%, indicating that 76 mg/g (o.d.) of the original shredded willow sample was obtained. Finally, the extract was lyophilized.

The fractioning of the stem extract was performed according to previously reported methods with some changes (Salminen and Karonen, 2011; Tian et al., 2018). 5 g of willow extract was dissolved in 50 ml of water using an ultrasonic bath. Extract was applied into a Sephadex LH-20 column (dimensions of the resin bed: 5.0 cm i.d. × 21 cm) and eluted successively with water (500 mL), aqueous ethanol (20, 40, 60, and 80% ethanol, 500 mL for each), and aqueous acetone (30, 50, and 70% acetone, 500 mL for each, except 700 for 70% acetone). Eight fractions were collected. Fractioning was repeated two times, and the equivalent fractions were combined. Fractions were concentrated by rotary evaporation at 45°C and finally lyophilized.

2.5. Commercial substances and samples

Commercial substances were used as references in the antiviral screening assays. Salicin and picein (purity >98%) were purchased from Merck Life Science Oy. Salicylic acid (purity >99%) was obtained from VWR Chemicals, and triandrin (purity 85%) was obtained from Molport EU. Additionally, Salixin Organic Powder (48^{TM}) and Salixin Organic Extract (800NP^{TM}) were supplied by Søren Fisker (Salixin A/S) and Pia Wikström (OY CELEGO AB) and were also tested for their antiviral efficacy along with the reference substances.

2.6. Antiviral activity assay

The screening of the bark extracts from 16 *Salix* clones to determine their antiviral activity against HCoV-OC43 was performed using the cytopathic effect (CPE) inhibition assay, modified from our previous study (Reshamwala et al., 2021). In

TABLE 1 *Salix* spp. clones used in this study.

Sample number	Species	Type	Clone
1	*S. myrsinifolia*	Native	E6682
2	*S. myrsinifolia*	Native	E6771
3	*S. myrsinifolia*	Native	E6948
4	*S. phylicifolia*	Native	E6666
5	*S. phylicifolia*	Native	K2191
6	*S. phylicifolia*	Native	K2218
7	*S. phylicifolia*	Native	K2277
8	*S. myrsinifolia* × *phylicifolia*	Native hybrid	K2183
9	*S. myrsinifolia* × *phylicifolia*	Native hybrid	K2269
10	*S. myrsinifolia* × *phylicifolia*	Native hybrid	K2341
11	(K2183 *S. myrs.* × *phyl.)* × S15136 *S. gmelinii**	Artificial hybrid	V7545
12	(K2183 *S. myrs.* × *phyl.)* × P6011 *S. gmelinii**	Artificial hybrid	V7546
13		Commercial clone	Scherenee
14		Commercial clone	Tordis
15		Commercial clone	Tora
16		Commercial clone	Klara

*S. gmelinii Pall. is former S. dasyclados Wimm (Väre et al., 2021).

brief, MRC-5 cells at a density of 15,000 cells/well were cultured in 100 µl of MEM supplemented with 10% FBS, 1% GlutaMAX, and 1% penicillin/streptomycin antibiotics on a 96-well flat-bottomed microtiter plate (Sarstedt, Numbrecht, Germany) for 24 h at 37°C. The next day, the virus was pre-treated with *Salix* bark extract (1% v/v) by preparing a virus–extract mix in 2% MEM and incubating it for 1 h at 34°C. The virus titer in the virus–extract mix was 2 × 10^4 PFU/ml. A virus without the extract was used as a positive control, and a mock infection without the virus and extract was used as a negative control for the experiment. Reference substances and commercial samples mentioned above were also tested at different concentrations against the virus. Following this, the virus–extract mix was added to the cells (MOI of 0.1) for 2 h at 34°C. After the incubation, cells were aspirated, and fresh media was added. Finally, cells were incubated for 5 days at 34°C or until the cytopathic effect was observed. *Salix* bark extract was also tested against CVB3 using the CPE inhibition assay. The experiment was performed similarly as described in a previously published article (Tienaho et al., 2021). The only difference was the CVB3 titer and *Salix* extract amount, which was 2 × 10^6 PFU/ml and 0.1% v/v, respectively, in the *Salix*–virus mix, and the final MOI was 10. Once the cytopathic effect was observed under the light microscope, cells were fixed and stained for 10 min using the CPE dye (0.03% crystal violet, 2% ethanol, and 3.5% formaldehyde). The stained viable cells were then washed two times with water, following which they were lysed using a lysis buffer (0.8979 g of sodium citrate and 1 N HCl in 47.5% ethanol). Finally, the absorbance of the viable cells in the 96-well plate was measured spectrophotometrically at 570 nm using the PerkinElmer VICTOR^{TM} X4 multilabel reader (PerkinElmer, Turku, Finland). The assay was performed two times independently.

2.7. Antiviral activity assay for SARS-CoV-2

Vero E6 cells at a density of 50,000 cells/well were cultured in 100 µl of MEM supplemented with 10% FBS, 1% GlutaMAX, and 1% penicillin/streptomycin antibiotics on a 96-well flat-bottomed microtiter plate for 24 h at 37°C. The following day, SARS-CoV-2 was pre-treated with 1% v/v of *Salix* bark extract of clone 5 or P-16 by preparing a virus–extract mix in 2% MEM and incubating it for 1 h at 34°C. Handling of the virus was carried out at the BSL-3 facility at the University of Helsinki, Finland. The virus titer in the virus–extract mix was 20 PFU/ml. After the incubation, the virus–extract mix was added to cells (MOI-0.00002) for 2 h at 34°C. Following the incubation, the cells were aspirated, fresh media was added, and they were incubated for 3 days at 34°C. Finally, the supernatant solution from the cells was collected and transferred to a new 96-microtiter plate for the extraction of viral RNA. The extraction was done using a Chemagic Viral RNA/DNA Kit (PerkinElmer, Turku, Finland). Once the viral RNA was extracted, we performed a real-time reverse transcriptase polymerase chain reaction (RT-qPCR) to qualitatively detect viral nucleic acid. This was performed using a SARS-CoV-2 RT-qPCR reagent kit (PerkinElmer, Turku, Finland). To compare the relative amounts of RNA in samples, we make use of the fact that 1 difference in Cq (cycle quantification) value means ∼2× difference in RNA amount. An equation (RNA difference = $0.9646e^{0.6948x}$, x is the difference in the Cq values between the mean of test samples and the mean of virus control] was deduced by using Cq differences down from 10 (10 cycle difference meaning ∼1000 difference in relative RNA amount). This was used in our calculations to gain a value for RNA difference, of which a log value was then calculated to describe the difference in virus amounts.

2.8. Time and temperature assay

The time and temperature studies were performed using the CPE inhibition assay as described above. The only modification was in the incubation temperature (room temperature and 34°C) and time interval (45 s and 5 min) between the virus and *Salix* bark extracts. The room temperature (RT) monitored by the sensor was 21 ± 1°C.

2.9. Time-of-addition studies

Time-of-addition studies of the *Salix* bark extract were performed using the CPE inhibition assay as described above. For this assay, three modes of infection were designed. In the pre-infection mode, cells were incubated with the *Salix* bark extract (1% v/v) for 1 h at 34°C. After the incubation, cells were washed briefly on ice and then infected with HCoV-OC43 (MOI 0.01) for another 1 h at 34°C. Following the infection, cells were washed, and fresh media was added and incubated for 5 days at 34°C. In the co-infection mode, a mix of the virus (1.6×10^3 PFU/ml) and the *Salix* bark extract (1% v/v) was prepared and added directly to cells for 1 h. Following the incubation, cells were aspirated, and fresh media was added and incubated for 5 days at 34°C. In the post-infection mode, cells were infected with HCoV-OC43 (MOI 0.01) for 1 h at 34°C. After the infection, the excess virus was removed by repeated washing. Then, media containing the *Salix* bark extract (1% v/v) was added (after 1 h) and the cells were incubated for 5 days at 34°C. The schematic showing the experimental design for time-of-addition studies is shown below. Virus control (without the *Salix* bark extract), *Salix* bark extract control (without virus), and mock infection were used as controls during each of the different modes of infection studied. Extracts from willow clones 5, 8, 10, and 16 (Table 1) were used for this assay. This experiment was performed two times independently.

2.10. Virucidal assay (endpoint dilution assay)

Quantification of the reduction in viral infectivity after treatment with the *Salix* bark extract was performed using a virucidal assay modified by Alvarez and colleagues (Álvarez et al., 2022). Briefly, MRC-5 cells were seeded at a density of 15,000 cells/well on the 96-well flat-bottomed microtiter plate and incubated for 24 h in 5% CO_2 and 37°C. The next day, HCoV-OC43 (1:2 dilution) was mixed with *Salix* bark extracts of clone 8 or 10 (50% v/v) and incubated for 15 min at RT. A virus control with a similar amount of virus without the *Salix* bark extract was also incubated. After the incubation, the virus–extract mix was diluted 100 times using 2% MEM. Following this, we performed 10-fold serial dilutions (10 dilutions in total). Each dilution was added in replicates of eight to the cells and incubated for 5 days at 34°C. Following the incubation, the cells were stained with the crystal violet dye for 10 min to differentiate between the healthy and infected cells. Finally, the virus titers were calculated using the Reed–Muench method (Reed and Muench, 1938) and expressed as particle-forming units (PFU) per ml.

2.11. Particle stability thermal release assay (PaSTRy) for enterovirus

The PaSTRy assay was performed as described before (Martikainen et al., 2015). The assay is based on recording the temperature at which the viral RNA becomes accessible to Sybr Green II (SGII), and the emitted fluorescence is detected by the CFX Real-Time PCR instrument (Bio-Rad C100, Helsinki, Finland). A reaction mixture of 50 μl containing 1 μg of CVA9 and *Salix* bark extract of clones 5, 10, or 16 (10% v/v) in PBS were incubated for 1 h at 37°C. After the incubation, 10× SGII (Invitrogen) diluted in double-distilled water (ddH_2O) was added to the reaction mix and then aliquoted into a thin-walled PCR plate (Agilent, Amstelveen, Netherlands). The thermal cycler recorded the fluorescence in quadruple from 20 to 90°C with 0.5°C intervals. The fluorescence data output was extracted from the BioRad CFX manager (2.1 software, accessed on 1 March 2022) and processed in GraphPad PRISM. The relative fluorescence emission (RFU) was plotted as a function of temperature to obtain the melt curve, and the melting temperature could be determined from the melt peak, which was plotted using the derivative of the RFU as a function of temperature [d (RFU)/dT].

2.12. Negative staining for transmission electron microscopy

To visually understand the effect of the *Salix* bark extract compounds on the structure of enteroviruses and coronaviruses, we imaged the *Salix* extract-treated viruses under the transmission electron microscope (TEM) JEM-1400 (JEOl, Tokyo, Japan). Prior to sample preparation with the coronaviruses and enteroviruses, formvar-coated copper grids were glow discharged (EMS/SC7620 mini-sputter coater) for 30 s and placed on a parafilm inside a Petri dish. A reaction mixture containing four parts of HCoV-OC43 or CVA9 (OC43 and CVA9 stock infectivity: 7.43×10^9 PFU/ml and 5.47×10^9 PFU/ml, respectively) and one part of *Salix* bark extract of clones 5 or 10 were added onto the grid and mixed gently with a pipette. An untreated virus of the same amount was used as a virus control. The treated and untreated virus samples were incubated inside a Petri dish at RT for 15 min. Subsequently, the excess of the sample was blotted away using Whatman paper (Whatman 3 MM). The samples were stained with 5 μl of 1% (w/v) phosphotungstic acid for 10 s, and the excess stain was blotted away using Whatman paper. The samples were left to air dry in a grid box overnight before imaging with the TEM, equipped with a field emission gun and LaB_6 filament, operating at a voltage of 80 kV in the BF-TEM imaging mode. The images were taken with a bottom-mounted Quemesa CCD camera with a resolution of 4008 × 2664 pixels.

2.13. Binding assay for coronaviruses

MRC-5 cells at a density of 70,000 cells/well were cultured in MEM supplemented with 10% FBS, 1% GlutaMAX, and 1% penicillin/streptomycin antibiotics in a 24-well flat-bottomed plate (Fischer Scientific) for 24 h in 5% CO_2 and 37°C. The next day, 10% v/v of Clone 5 or 10 was pre-treated with the virus (6.98 $\times 10^6$ PFU/ml) for 1 h at 37°C. A virus control with the same amount of virus under the same conditions was also used. The pre-treated virus and virus control were then added to the MRC-5 cells at a MOI of 50. The virus was allowed to bind to the cells for 1 h on ice under rocking conditions. Post that, the media was taken out, and three gentle washes with 0.5% BSA/PBS, 5 min each, were given to remove any unbound viruses. After the last wash, the cells were detached from the bottom using buffer AVL from the RNA isolation kit (Qiagen, Hilden, Germany, ref. 52906), and the RNA was isolated. The isolated RNA was converted into cDNA using a Promega kit and reverse primer ($5'$-AATGTAAAGATGRCCGCGTATT) (Merck). The cDNA was amplified using the BioRad kit, reverse primer, forward primer ($5'$-TGTTAGGCCRATAATTGAGGAC) (Merck), and running it through the Touch Thermal Cycler (Bio-Rad C1000, Helsinki, Finland). The amplification steps were as follows: 10 min at 95°C, 40 cycles of 15 s at 95°C and 1 min at 50°C, 5 s at 72°C, 1 min at 95°C, followed by cooling at 12°C for 10 min.

2.14. Immunofluorescence labeling and microscopy

MRC-5 cells were seeded at a density of 8000 cells/well on the 96-well flat-bottomed microtiter plate (Fisher Scientific, Finland) and incubated for 24 h in 5% CO_2 and 37°C. The next day, the virus was pre-treated with *Salix* bark extract (1% v/v) by preparing a virus–extract mix in 2% MEM and incubating it for 1 h at 34°C. The virus–extract mix was added to the cells (MOI of 50) for 1 h at 4°C or at RT, after which the excess virus was washed with PBS. Next, fresh 2% MEM was added, and infection was allowed to proceed at 34°C for 1 h or overnight before fixing with 4% paraformaldehyde for 30 min. The cells were then permeabilized with 0.2% Triton X-100. Following this, they were treated with primary antibodies: rabbit antibody against the S-protein of HCoV-OC43 (a kind gift from Professor Ilkka Julkunen, University of Turku, Finland), mouse J2 antibody against the dsRNA of the virus (Scicons, Hungary), and mouse tubulin antibody (Santa Cruz Biotechnology, USA).

Salix bark extract (1% v/v) was also studied for its effect on CVA9 using confocal microscopy. Here, the *Salix*–virus mix was prepared in a buffer (PBS with 2 mM $MgCl_2$) and incubated for 1 h at 37°C. After the incubation, the mix was diluted with 10× DMEM, then added to A549 cells (MOI of 100) and incubated for 6 h at 37°C before fixation as described above. The primary antibodies used were rabbit antibodies labeling the CVA9 capsid and mouse J2 against the dsRNA of the virus. After 1 h of incubation, cells were washed with PBS to remove excess primary antibody and then treated for 30 min with secondary antibodies: goat anti-rabbit Alexa 555 (Invitrogen Life Technologies, USA) or goat anti-mouse Alexa 488 (Invitrogen, Life Technologies, USA). Secondary antibodies were washed with PBS, and cell nuclei were stained with DAPI (Molecular Probes, Life Technologies, USA) in PBS.

Samples were imaged with a Nikon A1R confocal microscope. The imaging was carried out with the 40× objective (NA 1.25), 405 nm diode laser, 488 nm multiline argon laser, and 561 nm sapphire laser. Laser power and detector amplification settings were optimized for each channel. Virus protein and dsRNA channels were adjusted according to the cell control to exclude antibody background. Images were visualized using the software Fiji2 (ImageJ). CellProfiler 4.2.1 was used to determine the number of infected cells in a sample. First, the nuclei were identified as primary objects using the Otsu thresholding method. Next, the infected cells were identified as secondary objects using the previously identified nuclei as a reference. The manual thresholding method was used to differentiate the background from the virus protein signal. Finally, the area and intensity of the secondary objects were measured, and the data were exported to Excel, where a threshold was set manually to differentiate infected from non-infected cells. Quantification was done to calculate the infection (%) by comparing the infected cells of the virus control with those of the test samples. At least 500 cells per sample were analyzed in the HCoV-OC43 and CVA9 assays.

2.15. Chemical characterization

Willow bark and stem extracts and fractions were analyzed by high-resolution LC–MS according to an earlier method (Karonen et al., 2021). Briefly, the UPLC-DAD-ESI-QOrbitrap-MS/MS instrument consisted of an Acquity UPLC system (Waters Corp.) coupled to a quadrupole-Orbitrap mass spectrometer (QExactiveTM, Thermo Fisher Scientific GmbH). The column was an Acquity UPLC BEH Phenyl (2.1 × 100 mm, 1.7 µm, Waters Corp.), and acetonitrile and 0.1% aqueous formic acid were used as eluents. The UV and MS data were acquired throughout the analysis. Negative ionization was used with a spray voltage of −3.0 kV and in-source collision-induced dissociation (CID) set at 30 eV. The mass range of orbitrap was m/z 150–2250 for the full scan.

2.16. Statistical analysis

Statistical analysis was performed using GraphPad Prism 6 (GraphPad Software, San Diego, CA, USA). Data are presented as mean + standard error (SEM). One-way ANOVA, followed by the Bonferroni test ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and $^{****}p < 0.0001$), was used to assess the statistical significance of the differences between treated and untreated virus samples.

3. Results

3.1. Determining the broad-spectrum antiviral activity of *Salix* bark extract

First, the antiviral potential of the *Salix* bark extracts was evaluated against the seasonal human coronavirus OC43 using the

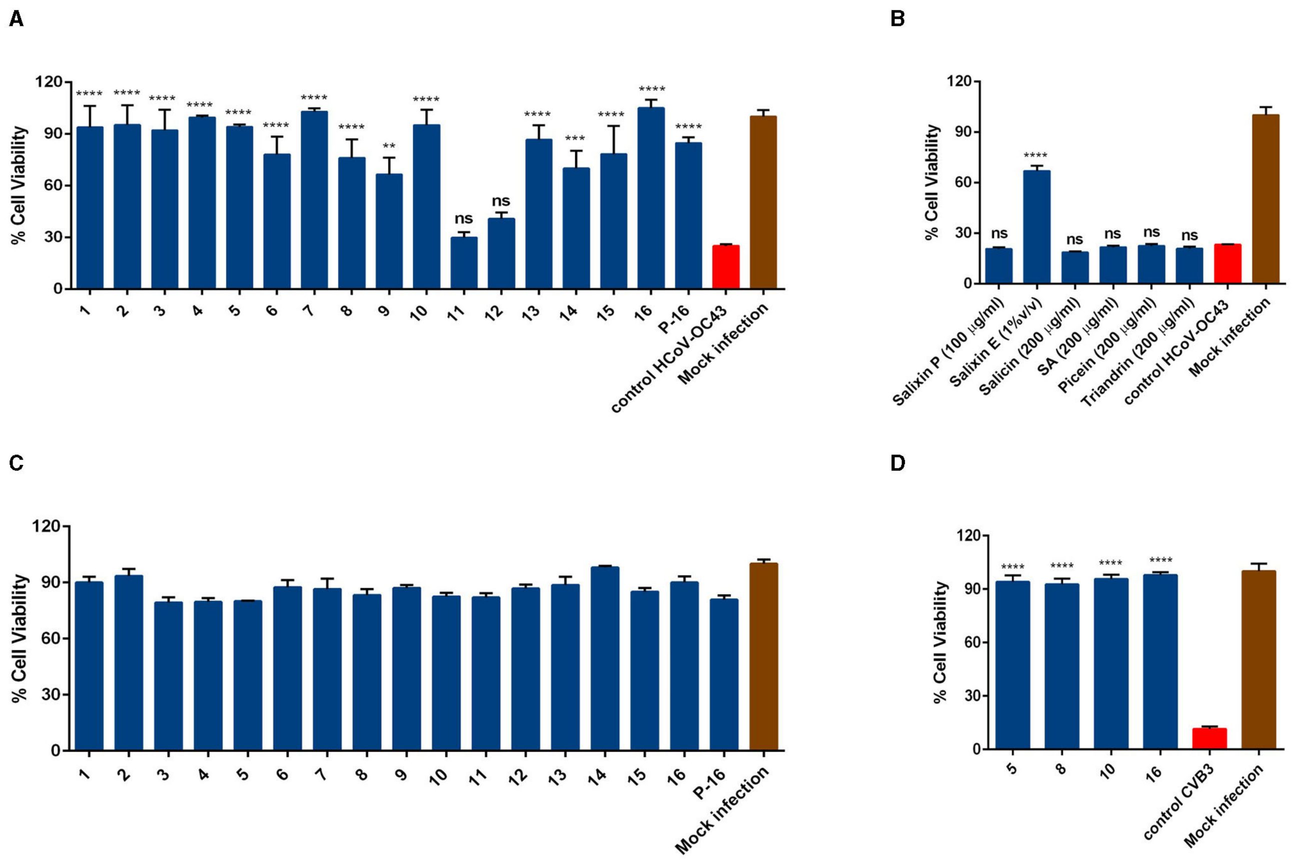


FIGURE 1
Testing **(A)** antiviral activity of bark extracts of *Salix* clones (1% v/v) **(B)** reference compounds (salicin, salicylic acid, picein, and triandrin) and Salixin organic powder and extract against HCoV-OC43 **(C)** Testing the cytotoxicity of bark extract of *Salix* clones (1% v/v) on MRC-5 cells **(D)** Testing the antiviral activity of *Salix* bark extracts of different clones (0.1% v/v) against CVB3. All the experiments were performed using the CPE inhibition assay. Virus control and test samples are normalized against the mock infection. The results are the mean of two independent experiments and are shown as average values + standard error of the mean (SEM). $^{**}p_{\text{ANOVA}} < 0.01$, $^{***}p_{\text{ANOVA}} < 0.001$, and $^{****}p_{\text{ANOVA}} < 0.0001$. ns, not significant; P-16, 2-L scale clone 16; Salixin P, Salixin Organic Powder 48^{TM}; Salixin E, Salixin Organic Extract 800NP^{TM}; SA, salicylic acid.

CPE inhibition assay. The virus was pre-treated with extract at 34°C for 1 h before being added to the cells. The bark extracts of all 16 *Salix* clones were tested at 1% v/v concentration. A virus sample without the extract was used as a positive control, and a mock infection without the virus and extract was used as a negative control for the experiment. The screening result showed that the bark extract of all the willow clones (except for 11 and 12) at 1% v/v concentration was able to protect the MRC-5 cells from HCoV-OC43 infection (Figure 1A). Research on *Salix* spp. bark extracts over the years has identified several key compounds that could play a major role in their pharmacological properties (Julkunen-Tiitto, 1985; Kammerer et al., 2005; Dou et al., 2018). Hence, we tested some of these reference substances (salicin, salicylic acid, picein, and triandrin) against the virus to understand if these substances contributed to the antiviral activity of the extract. However, none of the reference substances were able to block the virus infection, even if we had 10-fold higher concentrations of the reference compounds in comparison to the bark extracts (Figure 1B). We also tested Salixin Organic Powder (48^{TM}) and Salixin Organic Extract (800NP^{TM}) along with the reference compounds. Like the willow clones, Salixin Organic Extract showed antiviral activity against the virus, whereas the powder sample did not protect the cells from HCoV-OC43 infection. Cytotoxicity studies for bark extract of all 16 *Salix* clones (1% v/v) were also carried out similarly to the screening antiviral assay. From the result (Figure 1C), it was evident that the extract of none of the clones was toxic to the MRC-5 cells at the tested concentration. In our previous article (Tienaho et al., 2021), we had already shown the antiviral activity of the *Salix* bark extract against CVA9. We further wanted to assess whether the bark extracts of different willow clones are equally effective against another serotype of enterovirus. So, we tested some of the clones (5, 8, 10, and 16) against CVB3 and found that the bark extracts of all the clones tested were able to inhibit CVB3 infection and protect the A549 cells (Figure 1D). Thus, *Salix* bark extracts displayed broad-spectrum antiviral activity as it inhibited both enveloped as well as non-enveloped viruses.

Next, we wanted to observe the effect of willow bark extracts after only one infection cycle. This was studied using confocal microscopy, where the spike (S) protein of the virus was used as a marker of the virion, and cells were labeled using antibodies against tubulin and DAPI stain for the nucleus. The virus was first treated with willow bark extract of clones 5 or 10 for 1 h at 34°C, after

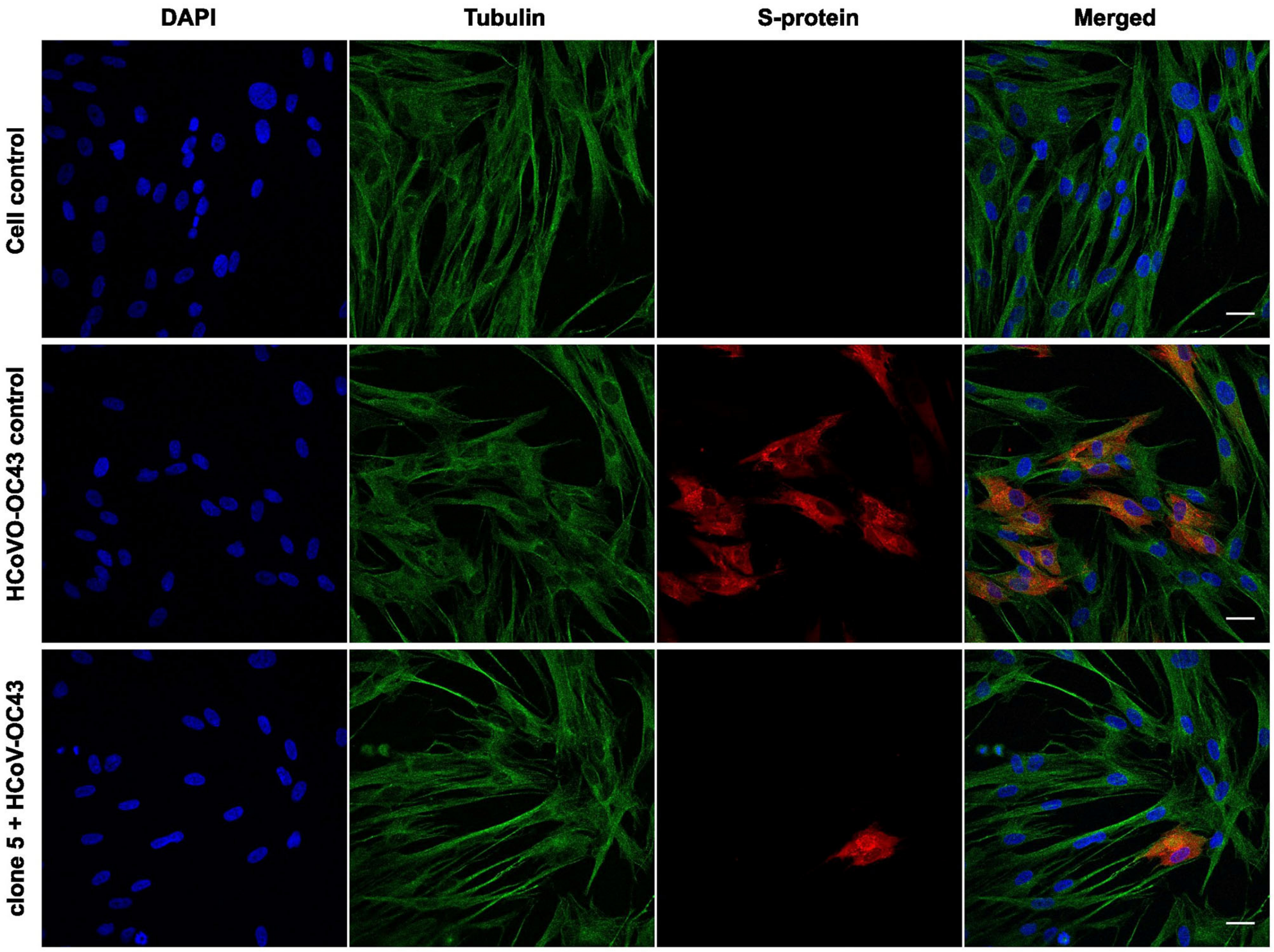


FIGURE 2
Studying the effect of *Salix* bark extract of clone 5 on HCOV-OC43 using immunofluorescence microscopy. Untreated and treated viruses were added to cells for 1 h at 4°C. After removing the unbound virus, cells were incubated overnight at 34°C before being fixed. The virus was labeled with the spike protein antibody and the cells with the tubulin antibody. Scale bar: 30 μm.

which the mixture was added to the cells, and infection was allowed to proceed for 15 h at 34°C. As expected, the mock infection showed no presence of the virus inside the cells, and the spike antibody did not cause any significant background fluorescence. The virus control showed the presence of ample amounts of S-protein inside the cells, confirming successful entry and infection of the virus in 42% of the MRC-5 cells (calculated from 350 cells) (Figure 2). However, when the virus was treated with the bark extract, the infection drastically decreased, and only 1% and 4% (at least 500 cells calculated in total) of MRC-5 cells were infected after treatment with clones 5 or 10, respectively (clone 5 shown in Figure 2). These results support the observation that the willow extract can efficiently decrease the infection of HCoV-OC43 and show in more detail that the virus protein production is halted.

In addition to HCoV-OC43, CVA9 was treated with the willow bark extract, and the infection was followed for one infection cycle (6 h). To evaluate the state of infection, both the viral capsid protein VP1 and the replication intermediate, dsRNA, were immunolabeled, and the results were detected by confocal microscopy. The virus was treated with the extract for 1 h at 37°C, after which the mixture was added to A549 cells, and infection was followed for 6 h. The control virus showed high infection in 37% of the cells (calculated from 500 cells) as the cytoplasm was full of VP1 protein and dsRNA was clearly visible (Figure 3). In contrast, none of the cells were infected in the extract-treated samples (at least 500 cells were calculated) as the signal of both VP1 protein and dsRNA was undetectable (clone 5 is shown in Figure 3).

To better demonstrate the antiviral efficacy against coronavirus, we also performed a virucidal assay. In the assay, HCoV-OC43 was pre-treated with *Salix* bark extract for 15 min at RT, followed by making serial dilutions and adding them to the cells. The result (Table 2) showed an approximately 3–4 log reduction in the virus titer for clones 8 and 10, respectively, compared to the non-treated virus. This demonstrated the exceptional potency of the bark extracts in reducing virus infectivity.

Though *Salix* bark extracts worked against the HCoV-OC43, a good surrogate for the more serious and fatal SARS-CoV-2, we wanted to evaluate their antiviral efficacy also against SARS-CoV-2. The antiviral activity of the *Salix* bark extract against SARS-CoV-2 was determined by running a RT-qPCR. Here, the viral RNA was

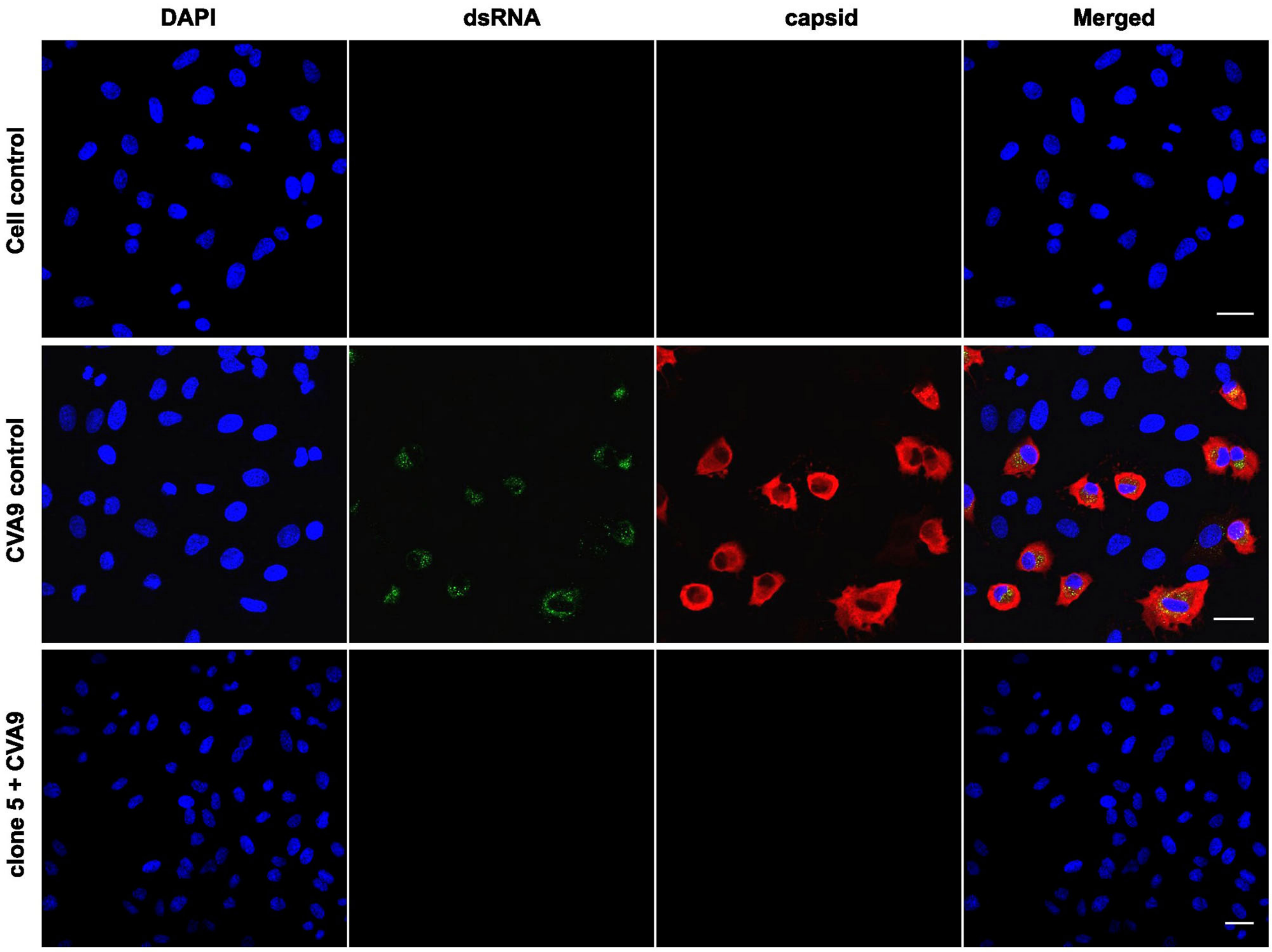


FIGURE 3
Studying the effect of *Salix* bark extract of clone 5 on CVA9 using immunofluorescence microscopy. Untreated and treated viruses were added to cells for 1 h at 37°C. Afterward, the cells were incubated for 6 h at 37°C before being fixed. The virus was labeled with the capsid protein antibody and the replication intermediate (dsRNA) antibody. Scale bar: 30 μm.

TABLE 2 Quantifying the reduction in virus infectivity using a virucidal assay.

Sample type	Virus titer (PFU/ml)
Virus control HCoV-OC43	3.88×10^{10}
HCoV-OC43 treated with clone 8	1.77×10^{7}
HCoV-OC43 treated with clone 10	1.09×10^{6}

The virus was treated with bark extract of willow clone 8 or 10 for 1 h, which led to a 3–4 log reduction in virus titer compared to virus control.

extracted from the supernatant of the infected cells and subjected to quantification by RT-qPCR to determine the presence of viral RNA. Cq values represent the number of PCR cycles taken to exceed the fluorescent intensity threshold line for detecting fluorescent signals from the sample. Cq values are inversely proportional to the amount of presence of viral RNA (cDNA) in the sample. Therefore, the lower the Cq values, the higher the amount of presence of RNA, and vice versa. The virus control had a Cq value of 14.82, suggesting a high amount of viral RNA. However, Cq values increased drastically for viruses treated with *Salix* bark extract of clones 5 and P-16 to 38.33 and 36.90, respectively (Table 3). This indicated that the amount of viral RNA was reduced significantly when it was treated with the willow bark extract. Cq values were also used to calculate the logarithmic reduction in the virus infectivity (Table 3). Based on these calculations, bark extracts caused a 6-log reduction in the viral RNA, which means a very high antiviral effect.

3.2. Effect of time and temperature on the antiviral activity of willow clones

We further wanted to study the impact of different time and temperature on the antiviral efficacy of the *Salix* bark extracts of different willow clones. So, we pre-treated the virus (HCoV-OC43) with the extract for different time periods (5 min and 45 s) and at different temperatures (34°C and RT) before adding it to the cells. When the bark extracts were incubated with the virus at 34°C for 5 min, the extracts were able to protect the cells from virus infection (Figure 4A). The antiviral efficacy was retained even when

TABLE 3 Cq mean values of the test and virus samples obtained from the qPCR are shown.

Sample	Cq mean value	Difference in Cq value compared to VC	Log difference
5	38.3304	23.5093	7.0782
P-16	36.9032	22.0821	6.6475
VC	14.8211	–	–

These mean Cq values were used to calculate the difference between test and virus samples and were further used to calculate the RNA difference. Using the RNA difference, logarithmic reduction was also calculated and is depicted in the table. VC, virus control; P-16 = 2-L scale clone 16.

the 5-min incubation was done at RT (Figure 4A). Remarkably, the extracts were equally effective in blocking the virus infection when the incubation time interval was further reduced to 45 s (Figure 4B). However, a slight reduction in the antiviral activity was evident after this very short incubation time. Overall, the *Salix* bark extracts showed excellent antiviral efficacy against HCoV-OC43 at different conditions.

3.3. Time-of-addition studies demonstrate direct action of *Salix* bark extract on coronaviruses

To elucidate the mechanism through which *Salix* bark extract blocks the virus infection, we first performed time-of-addition studies. In this assay, three modes of infection were studied (schematic shown in Figure 5A). During pre-infection mode, the extract was added to the cells for 1 h, and after the incubation, the cells were infected with the virus. In case of co-infection, a mix of virus and extract was prepared and added directly to the cells. For the post-infection mode, cells were first infected with the virus for 1 h, after which the extract was added. These studies revealed that when the *Salix* bark extract was added to the cells before or after the virus infection, it was unable to lower the antiviral activity (Figures 5B, D). However, when cells were co-infected with the clone and virus at the same time, the *Salix* bark extract was effective in protecting the cells from HCoV-OC43 infection (Figure 5C). Based on these results, it was evident that the *Salix* bark extracts do not have any effect through the host cells at used concentrations, neither do they interfere with cellular steps of viral infection. Instead, it has a direct effect on the HCoV-OC43 surface and protects the cells from viral infection.

3.4. Structural studies with TEM reveal direct effects of *Salix* bark extract treatment on coronaviruses and enteroviruses

The use of negative staining with heavy metal stains along with transmission electron microscopy (TEM) is one of the key imaging techniques that allows the direct visualization of the virus to observe morphological changes the antiviral may cause in the virus structure. To understand the effect of our willow clones on enveloped (OC43) and non-enveloped viruses (CVA9), we pre-treated both viruses with the willow bark extracts and then negatively stained them with 1% phosphotungstic acid. Negatively stained TEM samples of the untreated CVA9, which served as the control virus, showed intact virus particles that have a dark stain around the capsid and a bright center (Figure 6C). Only a small percentage of empty (red circle) capsids were observed. The empty enteroviruses have a darker center due to large openings in the virion and the flow of heavy metal to the inside. The control enteroviruses appeared to be separate from each other and largely spread on the TEM grid. Instead, in the presence of the willow bark extracts, the negatively stained virus samples showed large aggregates of the CVA9 virus, with a dark heavy metal stain around them (Figure 6D). Remarkably, the aggregates did not show dark centers or a darker appearance, strongly suggesting that they stayed intact, like in the virus control.

The negatively stained images of control OC43 showed roughly spherical to elliptical-shaped viruses with patchy dark centers and peplomers (spike proteins) sticking out from the membrane surface (Figure 6A). In addition, the viruses in the control sample were separated from each other on the TEM grid. In contrast, the coronaviruses in the willow bark extract-treated samples appeared to aggregate into clusters (Figure 6B) with a heavy metal stain around them, just like the CVA9. However, what was interesting is that the envelope fringe of the coronavirus looked distorted to some extent, and some of the peplomers had shed from the membrane surface. Also, the inside of the virus appeared to have more stains, suggesting obvious disintegration of the virus structure. These results thus altogether suggest that the willow bark extracts cause aggregation of both non-enveloped and enveloped viruses, increased stability for enteroviruses, and structural disintegration of coronaviruses.

3.5. Antiviral effects on enteroviruses are caused by increased stability of the virions (PaSTRy assay)

The PaSTRy assay is a good method that is used to study capsid stability among non-enveloped viruses. The assay uses an RNA-binding fluorescent dye (SGII) and a qPCR-based method to determine the melting temperature (Tm) at which the viral genome is released from the capsid. The presence of an inhibitor can typically affect the stability of the capsid by increasing the temperature at which the genome is released. In Figure 7, the red line shows a typical melt curve that is achieved using 1 μg of enterovirus (CVA9). The bell shape of the curve comes from the increase in fluorescence as the capsid proteins unfold during the heating process, making the viral RNA more accessible to the SGII for binding. The melt curve determined by RNA release performed in the presence of the bark extract of different willow clones (yellow, blue, and green colored lines) shows a reduction in the fluorescence peak compared to the control virus. These results suggest that the *Salix* bark extract does not allow the viral capsid to open even at high temperatures and prevents SGII access to the viral genome. When we plotted the fluorescence data as a negative first derivative plot per unit change in temperature, we were able to get the melt

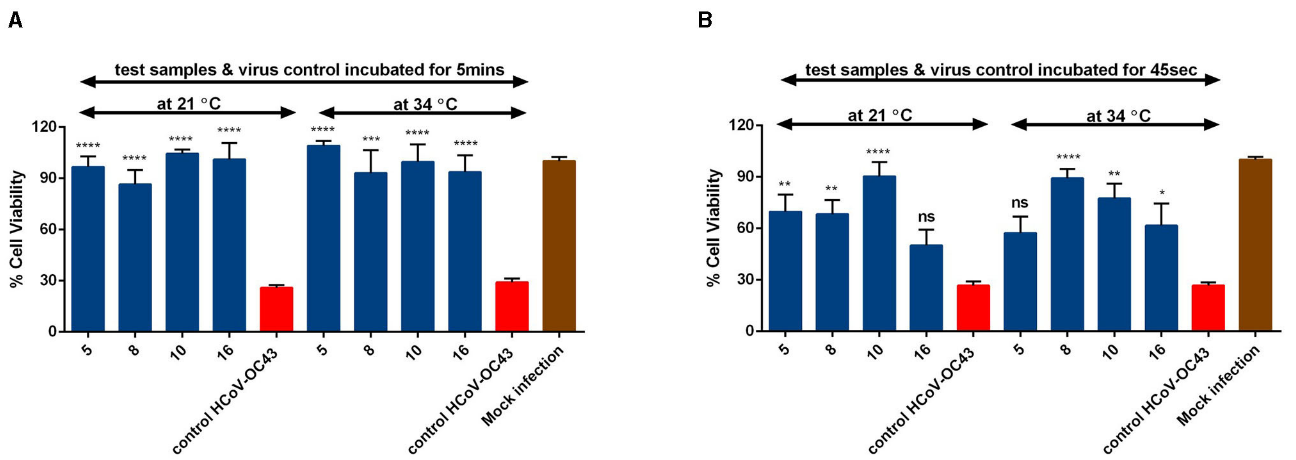


FIGURE 4
Effect of time and temperature on the antiviral activity of *Salix* bark extracts of selected willow clones (1% v/v) on coronaviruses using CPE inhibition assay. *Salix* bark extract–virus mix was incubated at 34°C and 21°C for **(A)** 5 min and **(B)** 45 s. Virus control and test samples are normalized against the mock infection. The results are the mean of two independent experiments and are shown as average values + standard errors of mean (SEM). $^{*}p_{\text{ANOVA}} < 0.05$, $^{**}p_{\text{ANOVA}} < 0.01$, $^{***}p_{\text{ANOVA}} < 0.001$, and $^{****}p_{\text{ANOVA}} < 0.0001$. ns, not significant.

peak, which gives us information about the Tm of the virus capsid (data not shown). Interestingly, the Tm results did not differ much between the untreated and treated viruses (for the control virus, the Tm was 51.5°C, and for the virus treated with clones 5, 10, or 16, it was 50°C, 49°C, or 51°C). However, the extent of opening was greatly diminished in the presence of *Salix* bark extracts.

3.6. *Salix* extracts do not interfere with the coronavirus binding to the cell surface

The binding assay was designed to study the first step of the coronavirus life cycle, i.e., receptor binding. The experiment was performed by pre-treating the virus with clones 5 or 10 and then adding them to the cells on ice. Ice binding ensures uniform binding and synchronized entry. RNA from the virus bound to its receptor was then isolated and quantified using qPCR. As seen in Figure 8A, the quantification results demonstrated that the extract-treated virus was able to bind to its host cell receptors in a similar fashion as the untreated (control) virus. This indicated that the spikes of coronavirus were intact enough to promote binding to the cell surface.

Confocal studies were also performed to track the movement of the extract-treated coronavirus inside the host cells. For this, spike protein and RNA of the virus were used as markers. The pre-treated virus was added to the cells at RT for 1 h. After removing the unbound virus, cells were incubated for another 1 h at 34°C before washing and fixing them. The aim was to synchronize the virus binding and only provide enough time for its entry. Cells infected with the untreated virus (control HCoV-OC43) showed signal only from the spike protein and not from the RNA (Figure 8B). This confirmed that the virus was able to successfully enter the cells but had not started with its replication step. In addition, the virus was predominantly present in the cellular periphery, where it was accumulated in small vesicles. For cells infected with extract-treated virus, even though the S-protein signal was very faint, the virus appeared in cellular endosomes, similar to that of untreated virus. This gives insight into the fact that the extract-treated virus, after its entry, ends up in endosome vesicles because of which it is not able to continue with its infection.

3.7. Chemical composition of hot water extracts

Willow bark extracts were screened by LC-DAD-Orbitrap-MS to characterize their chemical composition. The ion molecular formula of 54 components was defined from the mass spectra, and 37 compounds were tentatively classified (Supplementary Tables S1, S2). Tentative identification was based on measured accurate mass, calculated mass error (<2 ppm), MS^2 fragmentation, retention time, and wavelength of the UV absorption of the components. Compounds identified were hydroxycinnamic acids (Zhou et al., 2021), salicylates (Pizzato et al., 2022), flavonoids (Panche et al., 2016), flavan-3-ols (Karia et al., 2020), and proanthocyanidin oligomers (Galabov, 2007). Compounds differed qualitatively both between willow species and among genotypes within species.

3.8. Stem extract and its fractions

To examine the active components of willow extracts, a pilot-scale willow stem extract was subjected to fractioning in column chromatography over Sephadex LH-20 resin. Fractioning yielded eight fractions. The composition of the stem extract and the fractions was screened by LC-DAD-Orbitrap-MS (Supplementary Figure S1, Table S2), and their efficacy against enterovirus was measured.

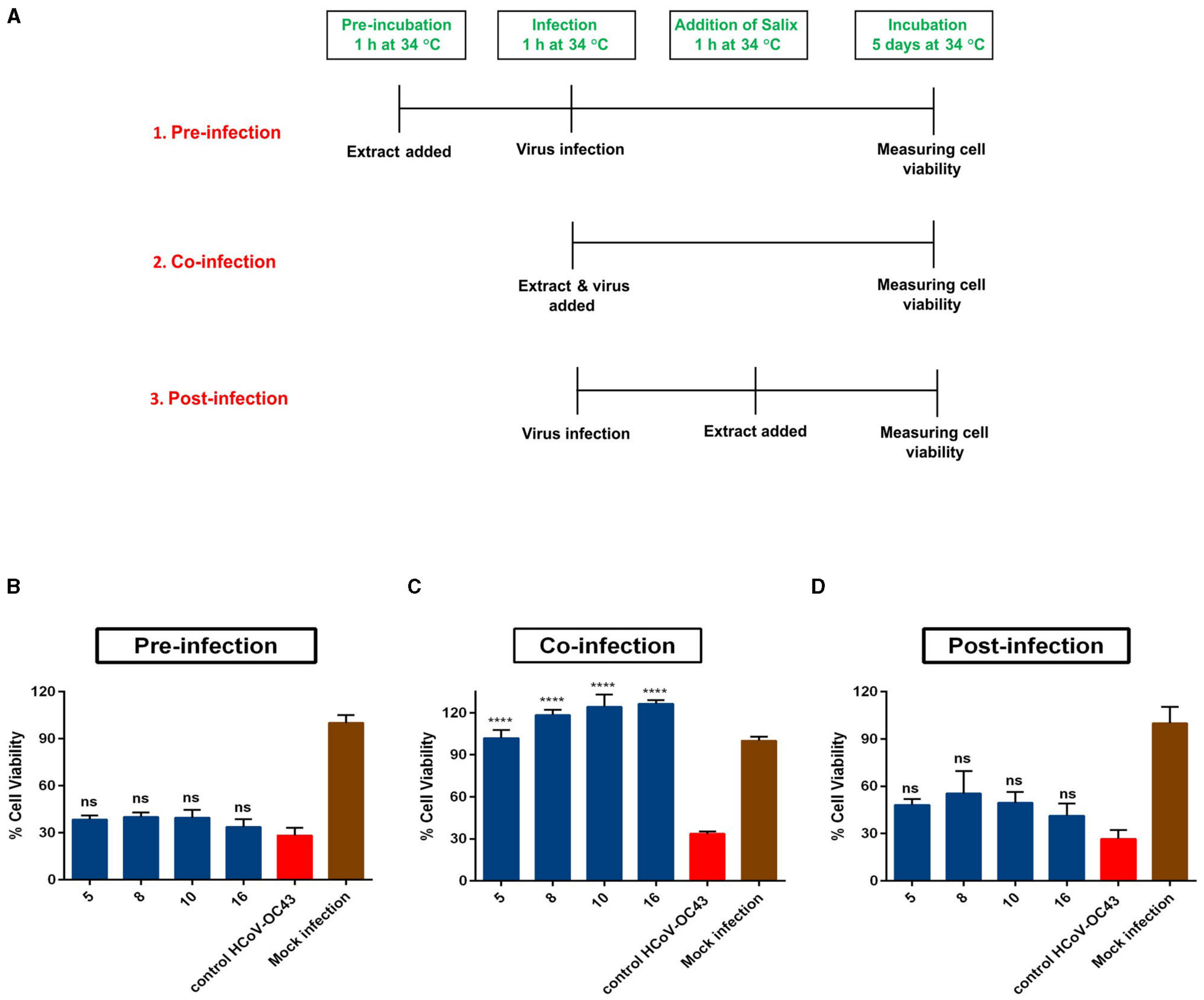


FIGURE 5
Time-of-addition assay. **(A)** Schematic of three modes of infection studied in the *Salix* bark extract of willow clones (1% v/v) against HCoV-OC43. **(B)** In pre-infection mode, the bark extracts were added 1 h before virus infection. **(C)** Co-infection, where virus and *Salix* clones were added together; **(D)** post-infection mode, where the clones were added 1 h after viral infection. The results are the mean of two independent experiments and are shown as average values + standard error of the mean (SEM). ****$p_{\text{ANOVA}} < 0.0001$. ns, not significant.

Some minor qualitative differences in the content between stem extract and bark extracts of the same species (samples 16 and P-16) were observed (Supplementary Tables S2, S3). Observed differences may be attributable to intraspecific variation among genotypes and to changes in extraction method, extracted plant part (bark vs. whole shoots), harvesting season, growing conditions, and stage of growth.

The stem extract and all the fractions, except fraction 1, were able to protect the A549 cells from CVA9 infection at a concentration of 3 µg/ml (Supplementary Figure S2). A lower concentration of 1 µg/ml had already lost antiviral activity, whereas 5 µg/ml showed very similar results with 3 µg/ml (Supplementary Figure S2). Of the active fractions, fraction 2 contained hydroxycinnamic acid derivatives, such as caffeoyl and coumaroyl quinic acids, and some unidentified compounds. Fractions 3 and 4 consisted of mainly flavonoids, such as quercetin and isorhamnetin glycosides. Fraction 5 contained flavonoids, flavan-3-ols, and procyanidin and prodelphinidin dimers. Fraction 6 had dimeric and trimeric procyanidins and prodelphinidins, together with some unidentified compounds. Fractions 7 and 8 consisted mainly of proanthocyanidins: The total ion chromatogram of fraction 7 showed peaks of trimeric and tetrameric proanthocyanidins together with a hump of higher degree of polymerization proanthocyanidins, while the chromatogram of fraction 8 showed only the unresolved hump of oligomeric proanthocyanidins (Supplementary Figure S1).

All the fractions that contained hydroxycinnamic acids (fraction 2) or polyphenolic flavonoids and proanthocyanidins (fractions 3–8) were active against CVA9 at a concentration of 3 µg/ml. Thus, the fractioning data imply that the polyphenolic structure in general might contribute to the observed activity. However, the active fractions obtained were mixtures of several

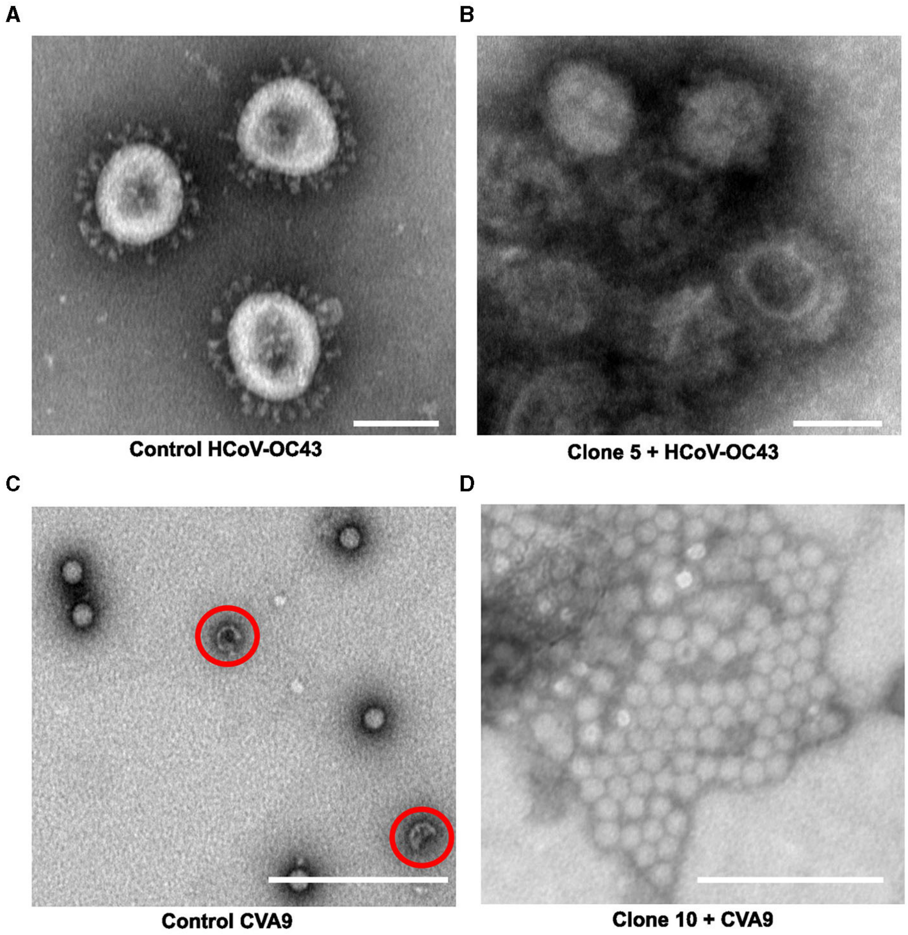


FIGURE 6
Studying the effect of *Salix* bark extract of clone 5 or 10 on enteroviruses or coronaviruses using transmission electron microscopy. **(A)** control HCoV-OC43, **(B)** HCOV-OC43 treated with clone 5, **(C)** control CVA9, and **(D)** CVA9 treated with clone 10. Using negative staining, empty enterovirus capsids can be identified with a darker center (red circle), whereas intact capsids appear brighter. Negatively stained control coronavirus appeared mostly spherical to elliptical with finger-like projections sticking out from the envelope. A scale bar of 100 nm has been added using Fiji.

components, some of them unidentified, and it is possible that the unknown compounds or the synergistic effects between the compounds have an impact on the detected efficacies. The next steps for our studies will be further fractioning of the active preparations and a more detailed investigation of their composition and activities.

4. Discussion

Viral outbreaks causing pandemics and yearly epidemics not only affect public health worldwide but also put a strain on the global economy due to the high costs associated with managing these outbreaks. The recent pandemic highlighted the limited resources the world had for fighting such outbreaks. Hence, there is an urgent need for developing broad-spectrum antivirals that can effectively reduce the viral load in the environment and on surfaces. Until 2021, there have been no previous studies exploring the antiviral potential of the *Salix* bark hot water extracts. Our study was the first to report the antiviral properties of these extracts against the highly stable, non-enveloped enteroviruses (CVA9) (Tienaho et al., 2021). In this study, we expanded our research to study the antiviral properties of these extracts against the enveloped human coronaviruses and investigated their mechanism of action against both coronaviruses and enteroviruses.

Our results showed that bark extracts of most of the willow clones were able to protect the MRC-5 cells from HCoV-OC43 infection when the virus was pre-treated with extracts before infecting the cells. Virucidal assay revealed a 3–4 log reduction in the virus infectivity of the extract-treated virus as compared to the untreated virus. The extracts showed efficacy both at high and low temperatures and even after a very short period of incubation (less than a minute). We demonstrated the antiviral nature of the extract against the clinically isolated SARS-CoV-2. All the clones studied were successful in inhibiting the SARS-CoV-2 infection and showed at least a 6-log reduction in the viral RNA.

Interestingly, none of the reference compounds (triandrin, salicin, salicylic acid, and picein) tested showed any antiviral activity against OC43. None of these reference compounds were effective against CVA9 in our previous study (Tienaho et al., 2021). This suggests that any of the commercial reference compounds do not individually contain high enough antiviral activity. This

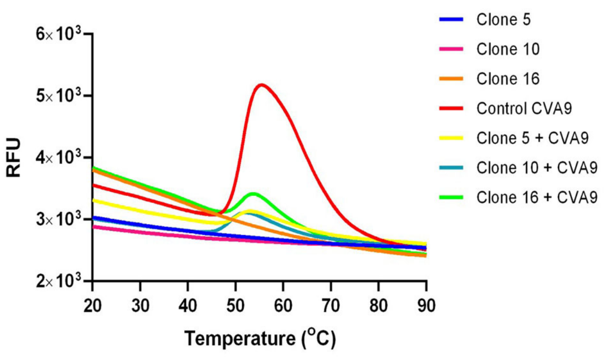


FIGURE 7
Capsid stability of *Salix* bark extract-treated enteroviruses was studied using the PaSTRy assay. The melt curve is a representation of the increase in fluorescence when the SG II gets access to the viral genome once the virus capsid opens. The red curve is a typical bell-shaped melt curve for enteroviruses. The indigo, pink, and orange curves represent the background fluorescence contributed by the willow clones with SG II. The yellow, blue, and green stunted curves are the melt curves from the virus treated with bark extract from different willow clones. This experiment was performed three times independently, and this is a representative result.

became more evident after fractionation of the bark extract when we observed that all fractions except fraction number 1 contained very high virucidal activities. Those active fractions contained various interesting chemical groups, of which many are known to contain biological activities. The bioactive properties of these bark extracts and broad-spectrum antiviral activity are thus likely to be due to the synergistic effects of the different detected flavonoids, hydroxycinnamic acid derivatives, and procyanidins. This hypothesis is further affirmed by the finding that there seemed to be a strong relationship between the values obtained from the Folin–Ciocalteu test for total phenolics and virucidal results against enteroviruses (Tienaho et al., 2021). Tannins isolated from *Hamamelis virginiana* bark extract have been reported to show antiviral properties against Influenza A virus and Human Papillomavirus (Theisen et al., 2014). Before the outbreak of SARS-CoV-2, researchers had already reported the effectiveness of some polyphenols such as luteolin and resveratrol against SARS-CoV (Yi et al., 2004) and MERS-CoV, respectively (Lin et al., 2017).

Interestingly, significant differences were detected between different willow clones in terms of the antiviral activity of their bark extract. Clones 11 and 12 showed significantly poorer antiviral activity than the rest of the clones. These clones are half-siblings and hybrids between clone 8 as a female parent and the species *S. gmelinii* as a male parent. The poor antiviral activity of these two clones seems to be connected to *S. gmelinii* in the ancestry, but the primary reason remains unclear without more detailed studies on the chemistry of the clones. However, the finding of existing differences between clones is important. The production of willow biomass for any possible antiviral applications would be carried out in plantations of selected willow clones.

Time-of-addition studies demonstrated here that the extracts do not primarily target the viral replication life cycle in the host cells, nor does it have any major intracellular effects. Instead, the studies suggested that the extracts act by interacting directly with the HCoV-OC43 surface. Previously, polyphenols isolated from *Eupatorium perfoliatum* were shown to inhibit influenza virus infection when the virus was treated with the polyphenols before adding it to the cells (Derksen et al., 2016). In another study, tannic acid was reported to inhibit the hepatitis C virus only when added to cells, thus also suggesting that there would be direct action on the virions (Liu et al., 2015). Here, negative staining for TEM showed opposite outcomes for coronaviruses and enteroviruses. While both virions were clustered due to *Salix* extract, the images suggested increased stability for enteroviruses but structural disintegration for coronaviruses. However, the binding assay for coronaviruses pointed out that the extract treatment did not interfere with the binding of the virus. Confocal studies revealed that treated coronaviruses could enter the cells but could not start replication/translation in the cells. Instead, the treated virus ended up inside the cellular endosomes.

Thermal assay clearly demonstrated that the willow bark extract had a stabilizing effect on enteroviruses. Normally, enteroviruses readily release their genome when heated to temperatures between 50°C and 60°C. Strikingly, after treatment with *Salix* bark extract, the genome release was almost totally blocked. An increase in stabilization has been earlier reported for enteroviruses when an antiviral molecule replaces the fatty acid in the hydrophobic pocket of the capsid (Tsang et al., 2000). Enterovirus capsid-binding drugs such as pleconaril often target this hydrophobic pocket, which is normally occupied by an aliphatic fatty acid (Pevear et al., 1999). The hydrophobic pocket is linked with virus stability, and the expulsion of lipid moiety from this pocket is associated with its genome release (Smyth et al., 2003). Also, we showed previously that polyphenols epigallocatechin gallate (EGCG) and resveratrol (RES) effectively inhibit the enterovirus infection by causing clustering and stabilization of the virions and prevent its genome release (Reshamwala et al., 2021). Docking simulations showed that polyphenols can actually bind several sites on the virion, not just the hydrophobic pocket. 4–6 sites were identified depending on the serotype and the compound, which resulted in strong stabilization (Reshamwala et al., 2021). Here, our confocal microscopy studies also confirmed a strong inhibition of infection after 1 h treatment of the virions before adding cells. Even after long periods of incubation, there was no apparent accumulation of enteroviruses in endosomes, suggesting that binding and entry to the cells had been compromised. Analogous to our previous results with EGCG and RES, it is likely that the effective compounds from *Salix* extracts bind strongly directly onto enterovirus capsid, interfere with receptor binding on cells, and inhibit entry to endosomes (Reshamwala et al., 2021).

Future prospects for this study would be a more detailed fractionation and characterization of the *Salix* bark extracts to identify the several bioactive compounds contributing to the antiviral nature of the willow bark. Once their chemical structure has been determined, *in silico* docking studies could be performed to understand the binding sites between the bioactive compounds and the viruses. A computational study to identify different polyphenols isolated from pomegranate peel extract as potential inhibitors for SARS-CoV-2 reported punicalin and punicalagin (two forms of tannins) to interact with the S-protein and to bind with

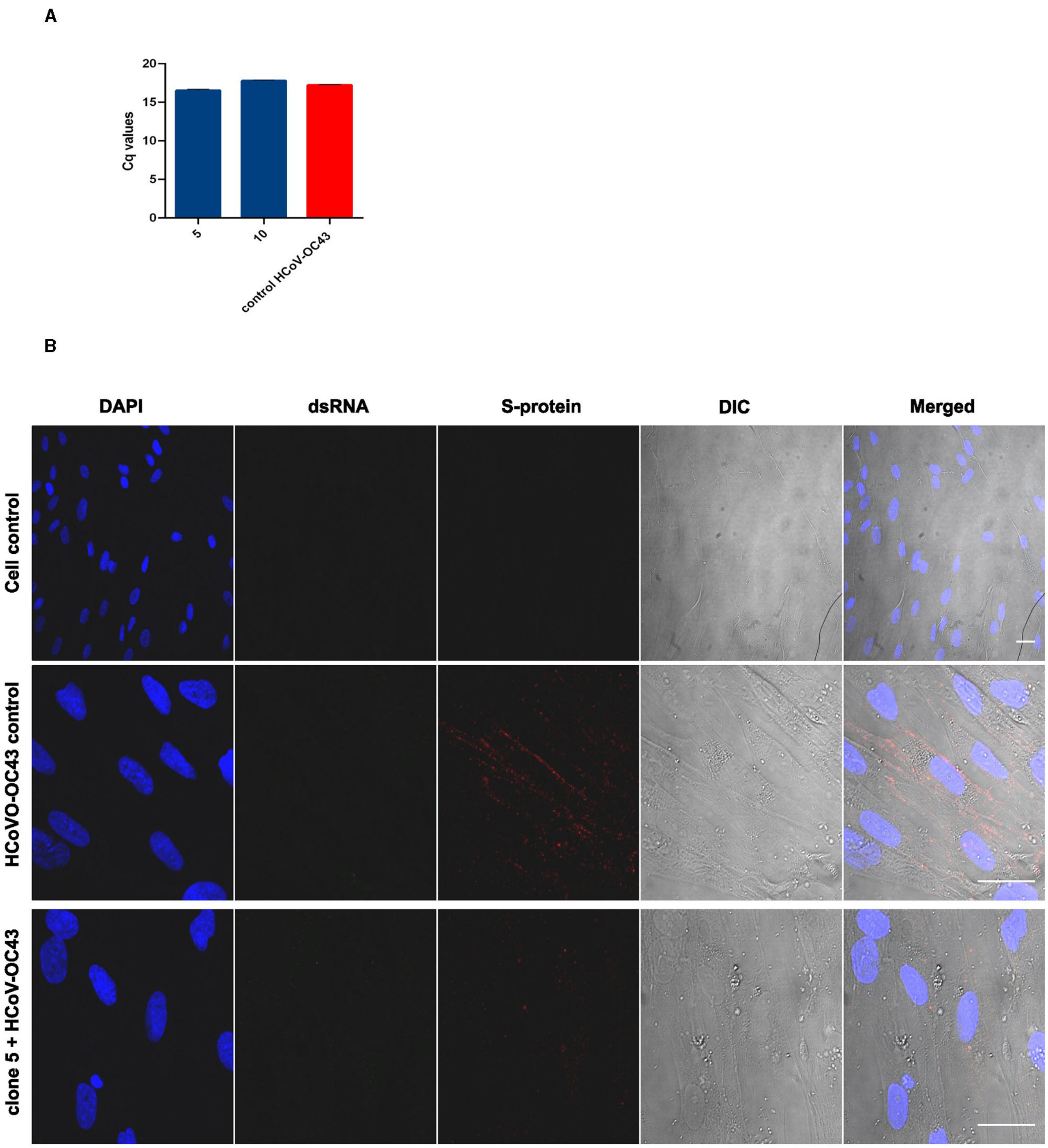


FIGURE 8
Effect of extract treatment on coronaviruses. **(A)** Studying the effect of *Salix* bark extract of clones 5 and 10 on the binding of HCOV-OC43 to MRC-5 cells. The *Salix*-treated virus was allowed to bind on ice on MRC-5 cells, and the RNA from the virus bound to the cells was isolated, transcribed, and quantified using RT-qPCR. The quantified RNA can be represented by the Cq values from the qPCR. **(B)** Studying the effect of *Salix* bark extract of clone 5 on HCoV-OC43 using immunofluorescence microscopy. Untreated and treated viruses were added to cells for 1 h at RT. After removing the unbound virus, cells were incubated for 1 h at 34°C before being fixed. The virus was labeled with the spike antibody and the replication intermediate (dsRNA) antibody. Scale bar: 30 µm.

higher affinity to inhibit the viral infection (Suručić et al., 2021). Whether similar or analogous tannins or other antiviral compounds would be found in the *Salix* bark extracts remains to be seen.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

DR: conceptualization, methodology, data curation, formal analysis, investigation, and writing—original draft. SS: methodology, formal analysis, investigation, and writing—reviewing and editing. JL and MK: data curation, formal analysis, investigation, and writing—reviewing and editing. JT: formal analysis, methodology, and writing—reviewing and editing. ML: methodology and writing—reviewing and editing. PK: writing—reviewing and editing. AV-A: formal analysis and methodology. TJ: conceptualization, project administration, funding acquisition, validation, and writing—reviewing and editing. VM: conceptualization, methodology, investigation, supervision, project administration, funding acquisition, validation, and writing—reviewing and editing. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by grants from the Jane & Aatos Erkko Foundation (VM) and the Academy of Finland grant Forest Antivirals (VM and TJ).

Conflict of interest

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1249794/full#supplementary-material

References

Álvarez, Á. L., Dalton, K. P., Nicieza, I., Abade Dos Santos, F. A., de la Peña, P., Domínguez, P., et al. (2022). Virucidal properties of photocatalytic coating on glass against a model human coronavirus. *Microbiol. Spectr.* 10, e00269–e00222. doi: 10.1128/spectrum.00269-22

Cassidy, H., Poelman, R., Knoester, M., Van Leer-Buter, C. C., and Niesters, H. G. M. (2018). Enterovirus D68 – the new polio? *Front. Microbiol.* 9, 2677. doi: 10.3389/fmicb.2018.02677

Chan, Y. F., and Abu Bakar, S. (2005). Virucidal activity of Virkon S on human enterovirus. *Med. J. Malaysia* 60, 246–248.

Derksen, A., Kühn, J., Hafezi, W., Sendker, J., Ehrhardt, C., Ludwig, S., et al. (2016). Antiviral activity of hydroalcoholic extract from *Eupatorium perfoliatum* L. against the attachment of influenza A virus. *J. Ethnopharmacol.* 188, 144–152. doi: 10.1016/j.jep.2016.05.016

Dou, J., Xu, W., Koivisto, J. J., Mobley, J. K., Padmakshan, D., Kögler, M., et al. (2018). Characteristics of hot water extracts from the bark of cultivated willow (*Salix* sp.). *ACS Sust. Chem. Eng.* 6, 5566–5573. doi: 10.1021/acssuschemeng.8b00498

Galabov, A. S. (2007). Virucidal agents in the eve of manorapid synergy®. *GMS Krankenhaushyg. Interdiszip.* 2.

Gaunt, E. R., Hardie, A., Claas, E. C. J., Simmonds, P., and Templeton, K. E. (2010). Epidemiology and clinical presentations of the four human coronaviruses 229E, HKU1, NL63, and OC43 detected over 3 years using a novel multiplex real-time PCR method. *J. Clin. Microbiol.* 48, 2940–2947. doi: 10.1128/JCM.00636-10

Goh, V. S. L., Mok, C. K., and Chu, J. J. H. (2020). Antiviral natural products for arbovirus infections. *Molecules* 25, 2796. doi: 10.3390/molecules25122796

Haveri, A., Smura, T., Kuivanen, S., Österlund, P., and Hepojoki, J., Ikonen, N., et al. (2020). Serological and molecular findings during SARS-CoV-2 infection: the first case study in Finland, January to February 2020. *Eurosurveillance* 25, 2000266. doi: 10.2807/1560-7917.ES.2020.25.11.2000266

Julkunen-Tiitto, R. (1985). Chemotaxonomical screening of phenolic glycosides in northern willow twigs by capillary gas chromatography. *J. Chromatograph. A* 324, 129–139. doi: 10.1016/S0021-9673(01)81312-1

Kammerer, B., Kahlich, R., Biegert, C., Gleiter, C. H., and Heide, L. (2005). HPLC-MS/MS analysis of willow bark extracts contained in pharmaceutical preparations. *Phytochem. Anal.* 16, 470–478. doi: 10.1002/pca.873

Karia, R., Gupta, I., Khandait, H., Yadav, A., and Yadav, A. (2020). COVID-19 and its modes of transmission. *SN Compr. Clin. Med.* 2, 1798–1801. doi: 10.1007/s42399-020-00498-4

Karonen, M., Imran, I. B., Engström, M. T., and Salminen, J. P. (2021). Characterization of natural and alkaline-oxidized proanthocyanidins in plant extracts by ultrahigh-resolution UHPLC-MS/MS. *Molecules* 26, 1873. doi: 10.3390/molecules26071873

Kilpeläinen, P. O., Hautala, S. S., Byman, O. O., Tanner, L. J., Korpinen, R. I., Lillandt, M. K. J., et al. (2014). Pressurized hot water flow-through extraction system scale up from the laboratory to the pilot scale. *Green Chem.* 16, 3186–3194. doi: 10.1039/C4GC00274A

Kumar, S., and Pandey, A. K. (2013). Chemistry and biological activities of flavonoids: an overview. *Sci. World J.* 2013, 162750. doi: 10.1155/2013/162750

Kurai, D., Saraya, T., Ishii, H., and Takizawa, H. (2013). Virus-induced exacerbations in asthma and COPD. *Front. Microbiol.* 4, 293. doi: 10.3389/fmicb.2013.00293

Kutter, J. S., Spronken, M. I., Fraaij, P. L., Fouchier, R. A., and Herfst, S. (2018). Transmission routes of respiratory viruses among humans. *Curr. Opin. Virol.* 28, 142–151. doi: 10.1016/j.coviro.2018.01.001

Lin, L. T., Hsu, W. C., and Lin, C. C. (2014). Antiviral natural products and herbal medicines. *J. Tradit. Complement. Med.* 4, 24–35. doi: 10.4103/2225-4110.124335

Lin, S. C., Ho, C. T., Chuo, W. H., Li, S., Wang, T. T., Lin, C. C., et al. (2017). Effective inhibition of MERS-CoV infection by resveratrol. *BMC Infect Dis.* 17, 144. doi: 10.1186/s12879-017-2253-8

Linnakoski, R., Reshamwala, D., Veteli, P., Cortina-Escribano, M., Vanhanen, H., Marjomäki, V., et al. (2018). Antiviral agents from fungi: diversity, mechanisms and potential applications. *Front. Microbiol.* 9, 2325. doi: 10.3389/fmicb.2018.02325

Liu, S., Chen, R., and Hagedorn, C. H. (2015). Tannic acid inhibits hepatitis C virus entry into huh7.5 cells. *PLoS ONE* 10, e0131358. doi: 10.1371/journal.pone.0131358

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395, 565–574. doi: 10.1016/S0140-6736(20)30251-8

Martikainen, M., Salorinne, K., Lahtinen, T., Malola, S., Permi, P., Häkkinen, H., et al. (2015). Hydrophobic pocket targeting probes for enteroviruses. *Nanoscale* 7, 17457–17467. doi: 10.1039/C5NR04139B

Myllynen, M., Kazmertsuk, A., and Marjomäki, V. A. (2016). Novel open and infectious form of echovirus 1. *J. Virol.* 90, 6759–6770. doi: 10.1128/JVI.00342-16

Nekoua, M. P., Alidjinou, E. K., and Hober, D. (2022). Persistent coxsackievirus B infection and pathogenesis of type 1 diabetes mellitus. *Nat. Rev. Endocrinol.* 18, 503–516. doi: 10.1038/s41574-022-00688-1

Newman, D. J., and Cragg, G. M. (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* 83, 770–803. doi: 10.1021/acs.jnatprod.9b01285

Panche, A. N., Diwan, A. D., and Chandra, S. R. (2016). Flavonoids: an overview. *J. Nutr. Sci.* 5, e47. doi: 10.1017/jns.2016.41

Pevear, D. C., Tull, T. M., Seipel, M. E., and Groarke, J. M. (1999). Activity of pleconaril against enteroviruses. *Antimicrob. Agents Chemother.* 43, 2109–2115. doi: 10.1128/AAC.43.9.2109

Pizzato, M., Baraldi, C., Boscato Sopetto, G., Finozzi, D., Gentile, C., Gentile, M. D., et al. (2022). SARS-CoV-2 and the host cell: a tale of interactions. *Front. Virol.* 1, 815388. doi: 10.3389/fviro.2021.815388

Reed, L. J., and Muench, H. A. (1938). Simple method of estimating fifty per cent endpoints12. *Am. J. Epidemiol.* 27, 493–497. doi: 10.1093/oxfordjournals.aje.a118408

Reshamwala, D., Shroff, S., Sheik Amamuddy, O., Laquintana, V., Denora, N., Zacheo, A., et al. (2021). Polyphenols epigallocatechin gallate and resveratrol, and polyphenol-functionalized nanoparticles prevent enterovirus infection through clustering and stabilization of the viruses. *Pharmaceutics* 13, 1182. doi: 10.3390/pharmaceutics13081182

Ruokolainen, V., Domanska, A., Laajala, M., Pelliccia, M., Butcher, S. J., Marjomäki, V., et al. (2019). Extracellular albumin and endosomal ions prime enterovirus particles for uncoating that can be prevented by fatty acid saturation. *J. Virol.* 93, e00599–e00519. doi: 10.1128/JVI.00599-19

Salminen, J. P., and Karonen, M. (2011). Chemical ecology of tannins and other phenolics: we need a change in approach. *Funct. Ecol.* 25, 325–338. doi: 10.1111/j.1365-2435.2010.01826.x

Sauerbrei, A., and Wutzler, P. (2010). Virucidal efficacy of povidone-iodine-containing disinfectants. *Lett. Appl. Microbiol.* 51, 158–163. doi: 10.1111/j.1472-765X.2010.02871.x

Smyth, M., Pettitt, T., Symonds, A., and Martin, J. (2003). Identification of the pocket factors in a picornavirus. *Arch. Virol.* 148, 1225–1233. doi: 10.1007/s00705-002-0974-4

Suručić, R., Tubić, B., Stojiljković, M. P., Djuric, D. M., Travar, M., GrabeŽ, M., et al. (2021). Computational study of pomegranate peel extract polyphenols as potential inhibitors of SARS-CoV-2 virus internalization. *Mol. Cell Biochem.* 476, 1179–1193. doi: 10.1007/s11010-020-03981-7

Theisen, L. L., Erdelmeier, C. A. J., Spoden, G. A., Boukhallouk, F., Sausy, A., Florin, L., et al. (2014). Tannins from hamamelis virginiana bark extract: characterization and improvement of the antiviral efficacy against influenza a virus and human papillomavirus. *PLoS ONE* 9, e88062. doi: 10.1371/journal.pone.0088062

Tian, Y., Liimatainen, J., Puganen, A., Alakomi, H. L., Sinkkonen, J., Yang, B., et al. (2018). Sephadex LH-20 fractionation and bioactivities of phenolic compounds from extracts of Finnish berry plants. *Food Res. Int.* 113, 115–130. doi: 10.1016/j.foodres.2018.06.041

Tienaho, J., Reshamwala, D., Sarjala, T., Kilpeläinen, P., Liimatainen, J., Dou, J., et al. (2021). *Salix* spp. bark hot water extracts show antiviral, antibacterial, and antioxidant activities—the bioactive properties of 16 clones. *Front. Bioeng. Biotechnol.* 9, 797939. doi: 10.3389/fbioe.2021.797939

Tsang, S. K., Danthi, P., Chow, M., and Hogle, J. M. (2000). Stabilization of poliovirus by capsid-binding antiviral drugs is due to entropic effects. *J. Mol. Biol.* 296, 335–340. doi: 10.1006/jmbi.1999.3483

Väre, H., Saarinen, J., Kurtto, A., and Hämet-Ahti, L. (2021). Suomen puu- ja pensaskasvio. Woody Flora of Finland. 3. fully revised edition. *Publ. Finnish Dendrol. Soc.* 10, 1552.

WHO (2022). *(COVID-19) Dashboard.* Available online at: https://covid19.who.int (accessed November 10, 2022).

Yi, L., Li, Z., Yuan, K., Qu, X., Chen, J., Wang, G., et al. (2004). Small molecules blocking the entry of severe acute respiratory syndrome coronavirus into host cells. *J. Virol.* 78, 11334–11339. doi: 10.1128/JVI.78.20.11334-11339.2004

Zhou, L., Ayeh, S. K., Chidambaram, V., and Karakousis, P. C. (2021). Modes of transmission of SARS-CoV-2 and evidence for preventive behavioral interventions. *BMC Infect. Dis.* 21, 496. doi: 10.1186/s12879-021-06222-4

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 07 December 2023
DOI 10.3389/fmicb.2023.1301861

EDITED BY
Shailendra Saxena,
King George's Medical University, India

REVIEWED BY
Liu Sidang,
Shandong Agricultural University, China
Qi Wang,
Chinese Academy of Agricultural Sciences,
China

*CORRESPONDENCE
Anchun Cheng
✉ chenganchun@vip.163.com
Shun Chen
✉ shunchen@sicau.edu.cn

RECEIVED 25 September 2023
ACCEPTED 10 November 2023
PUBLISHED 07 December 2023

CITATION
Xu L, Jiang B, Cheng Y, Gao Z, He Y, Wu Z, Wang M, Jia R, Zhu D, Liu M, Zhao X, Yang Q, Wu Y, Zhang S, Huang J, Ou X, Gao Q, Sun D, Cheng A and Chen S (2023) Molecular epidemiology and virulence of goose astroviruses genotype-2 with different internal gene sequences.
Front. Microbiol. 14:1301861.
doi: 10.3389/fmicb.2023.1301861

Molecular epidemiology and virulence of goose astroviruses genotype-2 with different internal gene sequences

Linhua Xu[1,2,3], Bowen Jiang[1,2,3], Yao Cheng[1,2,3], Zhenjie Gao[1,2,3], Yu He[1,2,3], Zhen Wu[1,2,3], Mingshu Wang[1,2,3], Renyong Jia[1,2,3], Dekang Zhu[1,2,3], Mafeng Liu[1,2,3], Xinxin Zhao[1,2,3], Qiao Yang[1,2,3], Ying Wu[1,2,3], Shaqiu Zhang[1,2,3], Juan Huang[1,2,3], Xumin Ou[1,2,3], Qun Gao[1,2,3], Di Sun[1,2,3], Anchun Cheng[1,2,3]* and Shun Chen[1,2,3]*

[1]Research Center of Avian Disease, College of Veterinary Medicine, Sichuan Agricultural University, Chengdu, Sichuan, China, [2]Key Laboratory of Animal Disease and Human Health of Sichuan Province, Sichuan Agricultural University, Chengdu, Sichuan, China, [3]Engineering Research Center of Southwest Animal Disease Prevention and Control Technology, Ministry of Education of the People's Republic of China, Chengdu, China

Goose astrovirus (GAstV) is a small, non-enveloped, single-stranded, positive-sense RNA virus. GAstV has rapidly spread across various regions in China since 2016. In Sichuan, out of 113 samples were collected from goose diseases between 2019 and 2022, 97 were positive for GAstV through PCR testing. Remarkably, over the past three years, GAstV outbreak in Sichuan has accounted for an astonishing 85.8% of all goose-origin viruses. Among these cases, 63.9% had single GAstV infections, 29.9% had dual infections, and 6.2% had quadruple infections. To comprehend the variations in virulence among distinct strains of GAstV. 12 representative strains of single GAstV infections were isolated. These strains exhibited distinct characteristics, such as prominent white urate depositions in organs and joints, as well as extensive tissues phagocytosis in major target organs' tissues. The conserved ORF1b genes and the variable ORF2 genes of these representative GAstV strains were sequenced, enabling the establishment of phylogenetic trees for GAstV. All GAstV strains were identified as belonging to genotype-2 with varying internal gene sequences. Experiments were conducted on GAstV genotype-2, both *in vivo* and *in vitro*, revealed significant variations in pathogenicity and virulence across susceptible cells, embryos, and goslings. This comprehensive study enhances researchers' understanding of the transmission characteristics and virulence of GAstV genotype-2, aiding in a better comprehension of their molecular epidemiology and pathogenic mechanism.

KEYWORDS

goose astrovirus, genotype-2, internal gene, molecular epidemiology, virulence

1 Introduction

Goose astrovirus is a small, non-enveloped, positive-sense RNA virus (Zhu and Sun, 2022; Zhang F. et al., 2022; Zhang X. et al., 2022). Its genome consists of a 5′-untranslated region (UTR), three open reading frames (ORF1a, ORF1b, and ORF2), a 3’-UTR, and a poly(A) tail (Yang et al., 2018; Ding et al., 2021). The ORF1a genes and ORF1b genes encode non-structural viral proteins (NSPs) that contain elements such as transmembrane domains (TMs), serine protease motifs, zinc protein models, nuclear localization signals, and RNA-dependent RNA polymerases (RdRps) (Wei et al., 2020b). Conversely, the ORF2-encoded viral capsid protein (CP) exhibits substantial genomic diversity, with conserved N-terminal segments and variable C-terminal regions (Chen Q. et al., 2020). The core antigenic components of GAstV are the capsid proteins (Ren et al., 2020). Adhering to the classification guidelines set by the International Committee on Taxonomy of Viruses (ICTV), classification is often based on the p-distance (p-dist) calculated between the full amino acid sequences of the hosts and the structural protein (Zhang M. et al., 2022). Members of the same genotypes species typically exhibit p-dist values >75% within their complete capsid protein sequences (Wang H. et al., 2022). Significantly, the p-dists between GAstV genotype-1 and GAstV genotype-2 stand at 0.609 ± 0.021. This value is akin to the p-dist observed between distinct avian astrovirus (ranging from 0.576 to 0.741), highlighting the presence of two distinct astrovirus variants in goslings (Fei et al., 2022). In accordance with this viral classification criterion, GAstV can be divided into two distinct genotypes: GAstV genotype-1 and GAstV genotype-2. Moreover, there are significant differences in strains characteristics between these two genotypes (Niu et al., 2018; Wang et al., 2021; Fu et al., 2022).

GAstV has erupted and spread extensively across various regions of China, emerging as a grave threat to the goose breeding industry in recent years (Xu et al., 2019; Yin et al., 2021; Zhu et al., 2022). After their initial outbreaks, GAstV rapidly spreads throughout various regions in China (Ji et al., 2020). In the most recent survey of viruses originating from geese in China, a staggering 50.84% of these viruses were identified as GAstV (Yuan et al., 2019; Yin et al., 2020; He et al., 2022; Yi et al., 2022). The infection rates of GAstV in clinical samples from six northern provinces of China reached 81.5%, resulting in substantial economic losses for the goose industry (Liu et al., 2020). Infected goslings typically exhibit symptoms such as depression, anorexia, accompanied by gray and cloudy eyelids (Zhang Q. et al., 2018; Liu et al., 2019; Yang et al., 2021; Wang A. et al., 2022). Notably, key pathological characteristics of GAstV infection are the extensive depositions of urate within the viscera and joints of affected goslings, leading to inhibited growth (Wu et al., 2020). The fatality rate associated with GAstV infection generally ranges between 20 and 50%. Moreover, GAstV has the potential for vertical transmission from geese to goslings (Wei et al., 2020a). The remarkable genetic diversity and recombination capabilities of GAstV underscore its ability to trigger diverse diseases across various hosts, even crossing species barriers to infect Pekin ducks (Wei et al., 2020b), Cherry Valley ducks (Chen H. et al., 2020), Moscow ducks (Chen et al., 2021), and other hosts. GAstV poses a significant threat to the waterfowl industry's development. The susceptibility of GAstV to changes in virulence is intricately linked to the structure of its single-stranded genome (Xu et al., 2019; He et al., 2020), with the highly variable nature of this genome being a probable reason for its efficacy in host-to-host transmission (Cui et al., 2022).

In a concerted effort to delve deeper into transmission characteristics and pathogenicity of distinct GAstV genotype-2, this study systematically collected 113 samples from infected goslings across various farms in Sichuan and isolated 12 single-infected GAstV strains from them. By analyzing the strains' pathological characteristics, phylogenetics, distributions, host adaptability, virulence, as well as pathogenicity. A comprehensive understanding was achieved regarding molecular epidemiology and characteristics of GAstV genotype-2. As far as our knowledge extends, this research represents the inaugural exploration into the virulence and host adaptability of GAstV genotype-2. It stands as the first systematic documentation, furnishing critical insights into the myriad aspects of GAstV genotype-2 within this specific locale. The findings from this research contribute to the enhancement of researchers' comprehension on epidemiology and pathogenesis of GAstV genotype-2.

2 Materials and methods

2.1 Cell lines, embryos, and viruses

10-days-old goose embryos and duck embryos were sourced from the Ya'an Animal Breeding Base of Sichuan Agricultural University. 3-day-old Tianfu meat goslings were also obtained from the same breeding base. LMH cells were cultured in Dulbecco's modified Eagle's medium Nutrient Mixture F-12 (DMEM-F12) supplemented with 10% fetal bovine serum (FBS), both of which were provided by Gibco (Shanghai, China). Twelve single-infected strains of GAstV were successfully isolated from deceased goslings within the Sichuan region in Southwest China.

2.2 Samples collection

Several goose farms in Sichuan have reported a recurring issue of goslings infected with GAstV displaying symptoms of depression and anorexia in recent years. Significantly, some of these infected goslings have developed gray and cloudy eyelids, coupled with inhibited growth. GAstV's capacity to replicate within various tissues has been observed, encompassing organs like the hearts, livers, spleens, kidneys, and intestines. Upon autopsy, significant urate depositions have been noted within both the viscera and joints of the afflicted goslings. To investigate further, PCR or RT-PCR analyses were conducted on these diseased goose samples, all of which had succumbed to non-environmental factors. For this study, the 12 selected strains of GAstV were thoroughly screened for the presence of other viruses, including but not limited to gosling plague viruses, tembusu viruses, goose reoviruses, adenoviruses, and avian influenza viruses. The results indicated the absence of these viruses in the examined GAstV strains, highlighting their specificity and isolation from other potential pathogens.

2.3 Viruses DNA/RNA genomes extraction

To investigate the potential causative pathogens or pathogens behind this disease outbreaks, viruses obtained from an intestinal and liver homogenate were subjected to blind passages in 10-day-old goose embryos via allantoic cavity injection. Following infection, the allantoic fluids from the affected embryos were harvested and subsequently centrifuged at 7,500 g/min for a duration of 20 min to eliminate cells debris. The resulting supernatant was collected and subjected to further ultracentrifugation at 9,000 g/min for 2 h. The resultant pellet was resuspended in PBS and utilized for the extraction of viral DNA/RNA. Distinct viral DNA and RNA components were extracted separately utilizing the TIANamp Virus DNA/RNA Fast Kit (TIANGEN, Beijing, China), following the manufacturer's stipulated protocols. The concentration of the extracted RNA was determined utilizing the Nanodrop 2000.

2.4 Virus identification and phylogenetic analysis

The isolation of total RNA from cells lysates, allantoic fluids, and tissue homogenates was executed using the Trizol extraction method (Takara, Dalian, China). Reverse transcription fluorescence quantitative PCR (RT-qPCR) was conducted on the Bio-Rad (Hurcules, CA). For the detection of viral RNA in infected samples, complementary DNA (cDNA) was first synthesized using the HiScript QRT SuperMix (Vazyme) as per the manufacturer's instructions for RT-PCR. Subsequently, real-time RT-PCR was conducted using the 2 × Taq SYBR Green qPCR mix (Innovagene, Changsha, China). The PCR products originating from the GAstV ORF1b genes were employed to establish a standard curve, thereby facilitating the absolute quantification of GAstV genome copies via the obtained Ct values. The quantified viral genome values were normalized against the absolute genome copies of the GAstV ORF1b genes, and the outcomes were expressed as GAstV genome copies per milligram (mg) of total RNA. To quantify viral load within infected tissues and allantoic fluids, the viral load was reported in terms of absolute genome copies per microliter (μL) of viral fluids. This quantification was derived from the standard curve generated using the product of the GAstV ORF1b genes. Primers used for RT-PCR or RT-qPCR in this study (Table 1). Primers: RT-PCR-GAstV-F/R 1–4 stand for the primers used for segmented sequencing of different strains of GAstV. Phylogenetic analysis was carried out using ORF1b and ORF2 internal sequences of GAstV, multiple alignments were constructed using the ClustalW method available within the MEGA 7.0 software. For the construction of phylogenetic trees of each gene segment for the twelve goose astroviruses, MEGA 7.0 with the distance-based neighbor-joining method was applied.

2.5 Virus growth kinetics in susceptible cells and embryos

In brief, the experimental process involved the following steps. LMH cells were introduced into a 12-well plate, with 1 mL of culture medium added to each well. The LMH cells were divided into two groups (the infected group and the control group). Each group comprised three separate cells wells, serving as biological replicates. The infected group contained LMH cells inoculated with each of the 12 GAstV genotype-2, while the control group remained untreated. After a 48-h period of GAstV infection, molecular techniques like RT-PCR and RT-qPCR were utilized. RT-PCR and RT-qPCR were applied to detect GAstV infection and quantify the respective GAstV genome copies. For the *in vivo* and *in vitro* aspect of the study, 12 GAstV genotype-2 were injected separately into goose embryos and duck embryos. Each strain was introduced into three embryos, serving as biological replicates. A control group consisted of embryos injected with an equal volume of phosphate-buffered saline (PBS). The viral dose for each strain was set at 2×10^4 $TCID_{50}$. Throughout the experiment, daily checks were conducted on embryos survival. In instance of embryos mortality, allantoic fluids were promptly collected for subsequent RT-PCR and RT-qPCR analysis. Alternatively, if embryos remained viable, RT-qPCR was employed to track viral load for a period of seven days post infection. In summary, these procedures enabled the comprehensive assessment of GAstV responses within LMH cells and embryos, facilitating insights into infection dynamics, replication, and its impact on host organisms.

TABLE 1 Primers used for RT-PCR or RT-qPCR in this study.

Primer name	Primer sequence (5′ → 3′)	Product size (bp)
RT-PCR-GAstV-F1	GAGAATAAGAAGAACATTTT	1815
RT-PCR-GAstV-R1	AGAAGTCGGGCCCGACCTC	
RT-PCR-GAstV-F2	TCTGGGGTAAATTGGTTTC	844
RT-PCR-GAstV-R2	TCACGTAAATGACAAAAGTT	
RT-PCR-GAstV-F3	AGTGCATTTACTGTTTTCAA	1,500
RT-PCR-GAstV-R3	TCGGCGTGGCCGCGGCTGCT	
RT-PCR-GAstV-F4	TGGTGGTGTTTTCTCAAAAATGA	788
RT-PCR-GAstV-R4	ACATTGGGAACTCCAACAAA	
RT-PCR-GPV-F	GGGTGCCGATGGAGTGGG	661
RT-PCR-GPV-R	GAGCCTGTCTAAGTCCTGTG	
RT-PCR-ADV-F	TGCGACAACTACCTGTGGAC	235
RT-PCR-ADV-R	GCGTACGGAAGTAAGCCAT	
RT-PCR-ARV-F	GTTCCATTCTGCTCCCCGG	634
RT-PCR-ARV-R	CGTCGAACACCATGTCAACC	
RT-PCR-TMUV-F	GCCACGGAATTAGCGGTTGT	401
RT-PCR-TMUV-R	TAATCCTCCATCTCAGCGGTGTAG	
RT-PCR-AIV-F	TTCTAACCGAGGTCGAAAC	229
RT-PCR-AIV-R	AAGCGTCTACGCTGCAGTCC	
PCA-GAstV-ORF1b-F	CATCATTTTGGCAAAGAATTCGCCACCATGGGCAGGATGATATTATTGAGTG	4,880
PCA-GAstV-ORF1b-R	TTGGCAGAGGGAAAAAGATCTCTAGGAGCATATTCATCTTGTTG	
RT-qPCR-GAstV-F	GGCAGGATGATATTATTGAGTG	160
RT-qPCR-GAstV-R	GGAGCATATTCATCTTGTTG	

2.6 Animal infection experiments

The animal infection experiments aimed to assess the impact of GAstV infection on goslings, focusing on factors such as mortality rate, viral load, and changes in body weight. Newly 3-day-old goslings, sourced from the Ya'an breeding base of Sichuan Agricultural University, were used for the experiment. Each gosling was injected with 5×10^4 $TCID_{50}$ of different GAstV strains. Three goslings were injected with each strain, serving as biological replicates. A control group was established, with three goslings injected with the same volume of phosphate-buffered saline (PBS). Goslings' health and behavior were closely monitored post-infection. Goslings showing signs of not eating or drinking were euthanized. Organ tissue samples, including infected livers, spleens, kidneys, and others, were collected daily from deceased goslings. This animal infection experiments sought to provide insights into the effects of GAstV infection on goslings, offering a comprehensive assessment of mortality rates, viral load, and physiological changes in response to different GAstV strains.

2.7 Histology HE staining and immunohistochemical staining

A portion of the spleen, liver, heart, brain, kidney and intestine samples was fixed in buffered 10% formalin for 24 h, dehydrated in graded alcohol, embedded in paraffin wax and cut into 5-μm-thick sections. Some sections were stained with haematoxylin and eosin (H&E) using a conventional protocol. Immunohistochemical (IHC) staining was performed at the same time. Briefly, a mouse monoclonal antibody against GAstV Cap protein was diluted at 1:500. After an overnight incubation with the primary antibody at 4°C and three washes with PBST, the sections were incubated with the goat anti-mouse secondary antibody (Biotin-Streptavidin HRP Detection Systems, ZSGB-BIO, Beijing, China) for 30 min at 37°C. Finally, the sections were observed under an optical microscope (Nikon, Tokyo, Japan).

2.8 Viral titers detection

Viral titers were determined by the median tissue culture infectious dose 50 ($TCID_{50}$) method in LMH cells. Viral samples were serially diluted 10-fold in DMEM, and then 100 μL dilutions of the viral sample were distributed to each of 8 wells of a 96-well plate seeded with a monolayer of LMH cells. After 120 h incubation at 37°C with 5% CO2, the presence of viruses was detected by assaying CPE using microscopy, and the viral titers were calculated according to the Reed-Muench method.

2.9 Quantification and statistical analysis

Data of the RT-qPCR is presented as means ± Standard Error (SEM). Student's t-test was used to assess statistical significance, with significance defined by p value <0.05 (*) in GraphPad Prism 9.0 software. Statistical significance of survival was analyzed using survival curve, Log-rank (Mantel-Cox) test in GraphPad Prism 9.0 software.

2.10 Ethical statements

All animal experimental procedures were approved by the Institutional Animal Care and Use Committee of Sichuan Agriculture University in Sichuan, China (Protocol Permit Number: SYXK (川)2019-187).

3 Results

3.1 Pathological characteristics of the goose astrovirus isolates

A significant number of goslings in Sichuan within Southwest China experienced mortality between 2019 and 2022, exhibiting distinct clinical manifestations including blackened eyelids, unsteady posture, reduced appetite, and substantial weight loss. Investigation into these fatalities involved the examination of liver tissues obtained from the deceased goslings. Through the application of RT-PCR, these liver tissues were confirmed to be infected with GAstV. Upon conducting dissections on the affected goslings, a range of notable observations emerged. An examination of the infected livers revealed marked congestion on the posterior side (Figure 1A). Urate deposits were conspicuously present on both the intestines (Figure 1B) and the hearts (Figure 1C) of the infected goslings. The gastric region of the deceased goslings exhibited evident redness and swelling (Figure 1D). The kidneys of the affected goslings were characterized by significant congestion (Figure 1E). Particularly noteworthy was the presence of substantial urate salt depositions on the surface of the liver (Figure 1F). These findings collectively offer valuable insights into the pathological impact of GAstV infection on the goslings. The presence of urate deposits and organ abnormalities underscores the systemic effects and varied organs involvements associated with GAstV infection in the affected goslings.

3.2 Molecular epidemiology and infection characteristics of the goose astrovirus isolates

To gain a more comprehensive understanding of the outbreaks and prevalence of GAstV in Sichuan, a series of measures were undertaken. RT-PCR was employed to detect and differentiate various disease types, encompassing notable candidates like goose astroviruses, gosling plague viruses, tembusu viruses, goose reoviruses, adenoviruses, and avian influenza viruses. The comprehensive analysis involved 97 samples from Sichuan, China. Key observations and findings from this investigation as follows: the regions in Sichuan with the highest incidence of GAstV were identified as Deyang, Chengdu, Meishan, Zigong, and Mianyang (Figure 2A). The presence of mixed infection, particularly GAstV along with other goose-origin viruses, intensified the epidemic situation of goose-origin viruses. This complexity further highlighted the interconnectedness of various viral agents in contributing to disease outbreaks. Among the analyzed samples, 63.9% (62 cases) exhibited exclusive GAstV infection, making it the most common infection type. Meanwhile, 16.5% of cases were co-infected with gosling plague viruses, 13.4% with tembusu viruses, and 6.2% with a combination of goose reoviruses, adenoviruses, and avian influenza viruses (Figure 2B). Significantly, GAstV infection constituted a significant portion, accounting for 85.8% of all goose-origin virus samples analyzed (Table 2). This information provided valuable insights into the overarching influence of GAstV among the array of goose-origin viruses. The intricate interactions between different viral agents underscore the complex nature of the epidemic situation, thereby contributing essential reference points for effective management and control strategies These findings offer a deep comprehension on the regional dynamics of GAstV strains and their characteristics during outbreaks.

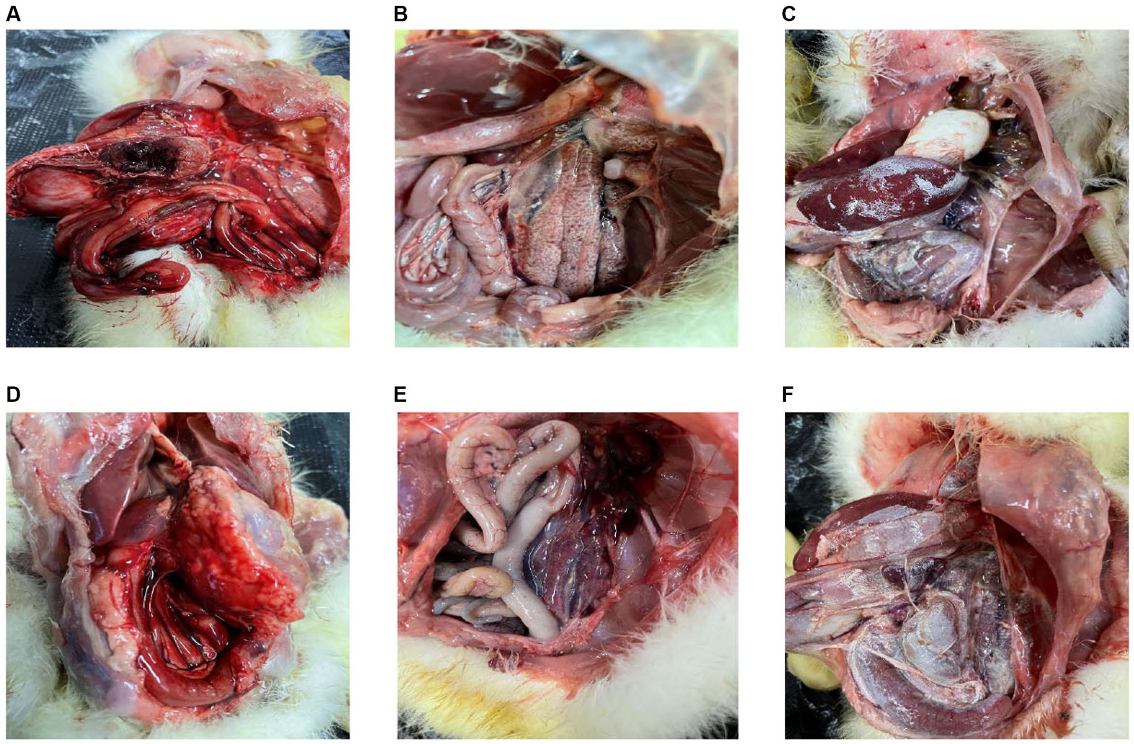


FIGURE 1
Pathologic changes in goslings infected with goose astroviruses genotype-2. **(A)** Congestion on the back of the livers. **(B)** Deposit of urate on the intestines. **(C)** Deposit of urate on the hearts. **(D)** Gastric redness and swelling. **(E)** Congestion in the kidneys. **(F)** Deposit of urate on the surface of livers.

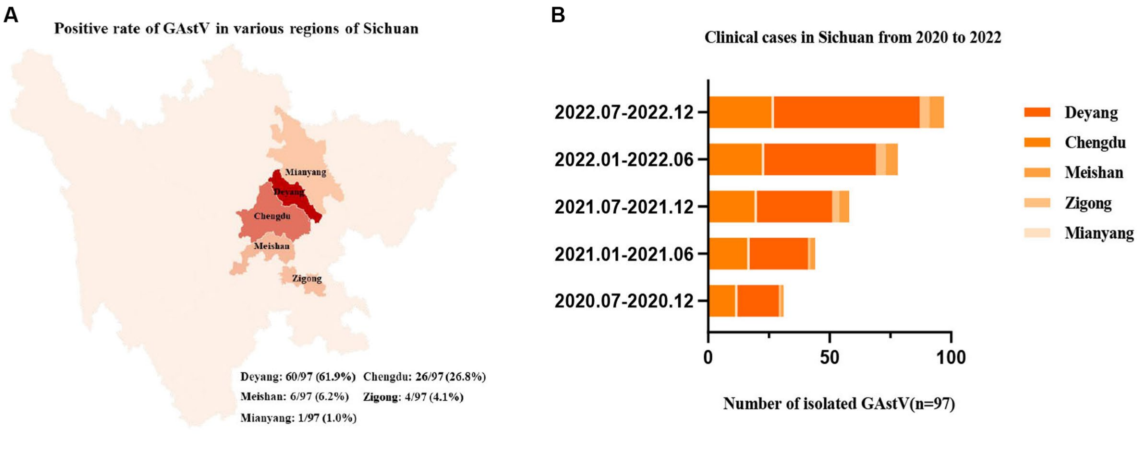


FIGURE 2
Epidemiological analysis of goose astroviruses in Sichuan within Southwest China. **(A)** Positive rates of goose astroviruses in various regions of Sichuan (*n* = 97). **(B)** Clinical case statistics of goose astroviruses from 2019 to 2022 (*n* = 97).

3.3 Genetic evolution analysis of the twelve goose astrovirus isolates

In the effort to analyze 12 strains of GAstV, four primer pairs were designed to sequence both the ORF1b and ORF2 genes. These sequences were subsequently aligned and compared to gather insights into their genetic characteristics. Key observations from the genetic analysis are as follows: the sequence alignment of the ORF1b genes from the 12 GAstV genotype-2, alongside four representative strains of GAstV genotype-1 and four representative strains of GAstV genotype-2, indicated the highest homology with GAstV genotype-2 (Figure 3A). Similarly, the sequence alignment of the ORF2 genes showcased the closest genetic affinity to GAstV genotype-2 (Figure 3B). In order to better analyze the homology between different strains of goose astroviruses, we compared the nucleotide sequences of different GAstV strains with the virulent strain (GAstV SCG3), and

TABLE 2 Infection statistics of goose astroviruses within samples collected from farms in Sichuan.

Region	Sample	Monoinfection	Coinfection with		
			GPV	TMUV	ARV, AdV and AIV
Deyang	60	37	10	13	0
Chengdu	24	18	0	0	6
Meishan	6	0	6	0	0
Zigong	4	4	0	0	0
Mianyang	3	3	0	0	0
Sum	97	62	16	13	6
Infection rate	97/113 (85.8%)	62/97 (63.9%)	16/97 (16.5%)	13/97 (13.4%)	6/97 (2.1%)

GAstV, goose astrovirus; GPV, gosling plague virus; TMUV, tembusu virus; GRV, goose reovirus; ADV, adenovirus; AIV, avian influenza virus.

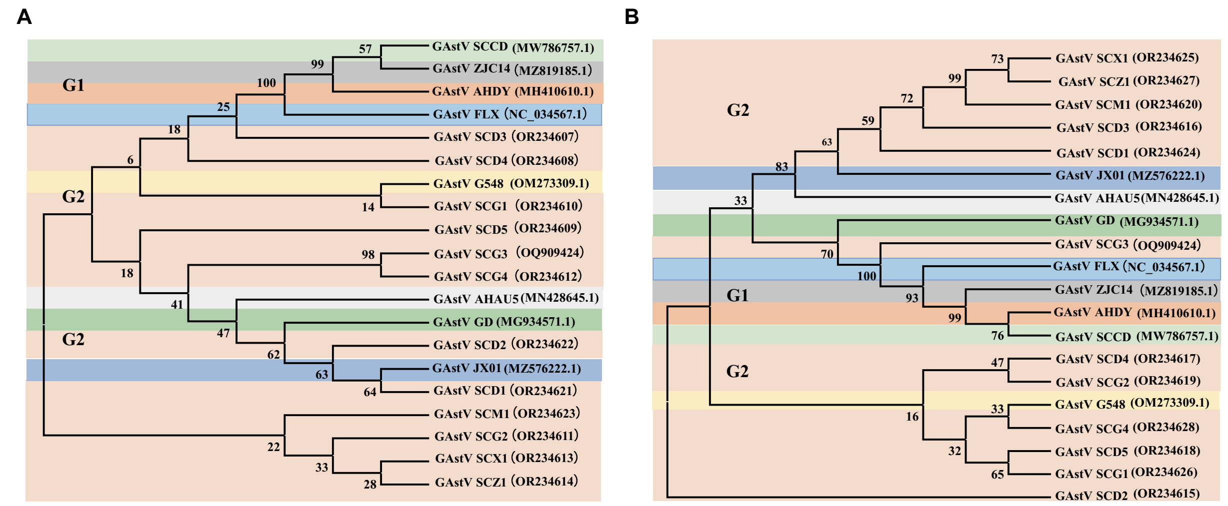


FIGURE 3
Phylogenetic analysis of goose astroviruses genotype-2. **(A)** Phylogenetic analysis of goose astroviruses genotype-2 based on ORF1b genes. **(B)** Phylogenetic analysis of goose astroviruses genotype-2 based on ORF2 genes. The trees were generated using MEGA 7.0 software and the Neighbor-joining method with 1,000 bootstrap replicates.

presented the detailed information of the strains in Table 3. These findings collectively underscore that there exists a single genotype of GAstV, specifically genotype-2, that prevails in the Sichuan region. This genetic uniformity provides valuable insight into the genetic makeup of the circulating GAstV strains, further informing our understanding of the viral strains prevalent in the area.

3.4 Growth kinetics of the goose astroviruses in susceptible cells and embryos

The isolation and analysis of 12 GAstV genotype-2 from Sichuan yielded interesting findings about their behavior and pathogenicity. The 12 isolated GAstV strains exhibited varying viral load. For instance, the SCG3 strain showed a notably higher viral load, while the SCM1 strain displayed lower viral load (Figure 4A). Different GAstV strains exhibited differing sensitivities to LMH cells. Those strains with higher viral load demonstrated successful cultivation in susceptible LMH cells (Figure 4B). The sensitivity of GAstV strains to goose embryos and duck embryos varied. Among the strains, only SCX1 and SCG3 effectively replicated and propagated in goose embryos (Figure 4C). Overall, the adaptability of all GAstV strains to goose embryos exceeded their adaptability to duck embryos. Significantly, strains with higher viral load displayed enhanced adaptability to both goose embryos and duck embryos (Figure 4D). These observations collectively shed light on the diverse responses and adaptability of different GAstV strains. The variations in viral load, cells culture sensitivity, and embryos adaptability contribute to understanding of how these strains interact with different hosts.

3.5 Virulence of the goose astroviruses in susceptible embryos and goslings

The investigation into GAstV strains in relation to their virulence and impact on different hosts led to noteworthy observations. The virulence of GAstV strains differed, affecting the fatality rate of goose embryos. Strains with higher viral load demonstrated the capability to induce higher embryos mortality, with strains with elevated viral load causing easier fatalities (Figure 5A). Interestingly, none of the GAstV strains were found to be lethal to duck embryos (Figure 5B). The

TABLE 3 Alignments with the nucleotide identities of GAstV SCG3.

Goose astrovirus genotype-2	Percent Identify (%) of amino acids (aa) to Goose astrovirus SCG3 (OQ909424)			
	Isolate Date	Isolate Region	ORF1b (GenBank accession no.)	ORF2 (GenBank accession no.)
Goose astrovirus SCD1	2021.10.18	Chengdu	98.58 (OR234621)	98.35 (OR234624)
Goose astrovirus SCD2	2021.12.14	Chengdu	98.52 (OR234622)	98.49 (OR234615)
Goose astrovirus SCD3	2021.12.28	Chengdu	97.88 (OR234607)	97.64 (OR234616)
Goose astrovirus SCG1	2021.12.30	Deyang	98.14 (OR234610)	98.11 (OR234626)
Goose astrovirus SCD4	2022.02.15	Chengdu	97.69 (OR234608)	98.58 (OR234617)
Goose astrovirus SCG2	2022.05.07	Deyang	98.90 (OR234611)	98.53 (OR234619)
Goose astrovirus SCX1	2022.04.20	Chengdu	98.71 (OR234613)	98.25 (OR234625)
Goose astrovirus SCM1	2022.05.01	Meishan	98.52 (OR234623)	98.49 (OR234620)
Goose astrovirus SCD5	2022.05.17	Chengdu	97.88 (OR234609)	98.25 (OR234618)
Goose astrovirus SCZ1	2022.05.18	Zigong	98.77 (OR234614)	98.30 (OR234627)
Goose astrovirus SCG4	2022.08.27	Deyang	98.20 (OR234612)	98.49 (OR234628)

majority of GAstV strains exhibited the ability to infect goslings effectively. Many strains resulted in high levels of viral load detected within the target organs of the goslings. Notably virulent strains, such as SCG3, were capable of causing the death of all three infected goslings (Figure 5C). Each GAstV strain displayed varying degrees of impacts on gosling weight. Overall, all strains significantly inhibited the growth of gosling weight (Figure 5D). These findings illustrate the diverse impacts of GAstV strains on different hosts and their varying levels of virulence. The experimentally observed effects on embryos survival, gosling infection, and growth inhibition contribute to our comprehension on the range of outcomes associated with different GAstV strains.

3.6 Transmissibility and pathogenicity of the representative goose astrovirus in goslings

Through immunohistochemistry experiments, it can be found that the positive signal in the liver is the strongest, while the positive signal in intestine, heart, and brain is very weak (red arrow). indicating that the antigen of GAstV is mainly distributed in liver, followed by kidney, spleen, intestine, heart and brain (Figures 6A–F), According to the grading of positive cell rate, the average positive rate of liver, kidney and spleen reach over 51% (level 3), which can indicate that the antigen of GAstV is enriched in liver, kidney and spleen. Through pathological analysis, a small number of scattered lymphocytes show necrosis with fragmented nuclei, along with a mild infiltration of neutrophils (Figure 7A). In the liver tissue, there is diffuse hepatocellular steatosis with varying-sized round vacuoles in the cytoplasm; there is mild congestion and dilation in multifocal liver sinusoids; a significant amount of central vein congestion is observed; occasional presence of numerous white blood cells in central veins (Figure 7B). In the kidney tissue, there is widespread thickening of the glomerular basement membranes; extensive edema and fibroblast proliferation accompanied by a significant infiltration of lymphocytes; numerous renal tubules show indistinct structures, cellular vacuolar degeneration, and varying-sized round vacuoles within the cytoplasm; there is a small amount of lymphocyte infiltration in the interstitium; scattered areas exhibit slight cell necrosis and nuclear fragmentation; small areas of hemorrhage are visible (Figure 7C). Cardiac muscle fibers are uniformly stained, with clear cell boundaries, consistent striations, distinct cross-striations in cardiac muscle cells, and no abnormalities in the interstitium (Figure 7D). The brain did not exhibit significant lesions (Figure 7E). The small intestine features villi on its surface and is covered by a single layer of columnar epithelial cells; there is extensive necrosis and shedding of mucosal epithelial cells; goblet cells are distributed among epithelial cells; the mucosal muscle layer, consisting of double-layered smooth muscle cells, separates the crypts from the submucosa; the submucosa is composed of connective tissue; the remaining layers of the intestinal wall include a muscle layer composed of smooth muscle cells and a serosal layer; there is significant hypertrophy of the muscle layer with an increased number of muscle cells (Figure 7F). These strains showcased distinct pathological changes including extensive tissues phagocytosis in major target organs.

4 Discussion

Except for a virulent strain of GAstV that can cause atypical jaundice symptoms and gosling gout, the majority of isolated GAstV strains in Sichuan primarily exhibit the pathological characteristic of inducing severe gout and fatality in goslings. These strains belong to the typical features of GAstV genotype-2. Gout in poultry, also referred to as urate depositions or poultry urolithiasis, denotes the accumulation of excessive urate in the bloodstream, resulting in hyperuricemia that cannot be efficiently eliminated from the body (Xi et al., 2020; Wang Z. et al., 2022). There are two primary factors contributing to gout: excessive presence of uric acid in the body and disrupted host uric acid excretion (Bidin et al., 2012a,b; Wang et al., 2020; Dalbeth et al., 2021). The key driver of heightened uric acid production is the consumption of protein-rich feed containing nucleoproteins and purine bases, such as animal viscera, fish meals, soybeans, and other ingredients (Liao et al., 2015; Wan et al., 2019; Ren et al., 2022). Elevated protein intake, particularly nucleoproteins,

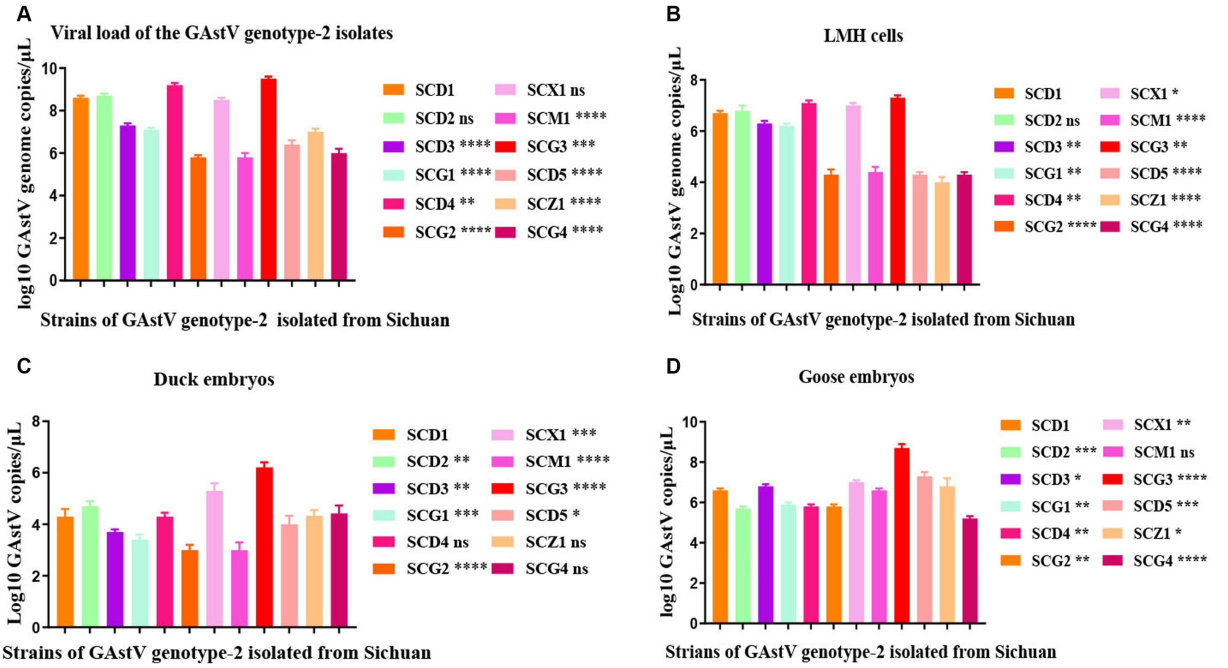


FIGURE 4
Replication kinetics of goose astroviruses genotype-2 *in vivo* and *in vitro*. **(A)** Viral load of goose astroviruses genotype-2 isolated from Sichuan. LMH cells from the experimental group (*n*=36) were infected with 2×10^4 $TCID_{50}$ of goose astroviruses genotype-2 (each strain *n*=3), while embryos from the control group (*n*=3) were mock infected with equal volume of PBS. **(B)** Viral load change in LMH cells infected with goose astroviruses genotype-2. 10-day-old goose embryos and duck embryos from the experimental group (*n*=36) were infected intravenously with 2×10^4 $TCID_{50}$ of goose astroviruses (each strain *n*=3), while embryos from the control group (*n*=3) were mock infected with equal volume of PBS. Statistical results one generation post infection. **(C)** Viral load change in duck embryos infected with goose astroviruses genotype-2. Statistical results one generation post infection. **(D)** Viral load change in goose embryos infected with goose astroviruses genotype-2. Statistical results one generation post infection. * Indicates statistically significant differences (ns indicates not significant, * indicates $p \leq 0.05$, ** indicates $p \leq 0.01$, *** indicates $p \leq 0.001$, **** indicates $p \leq 0.0001$).

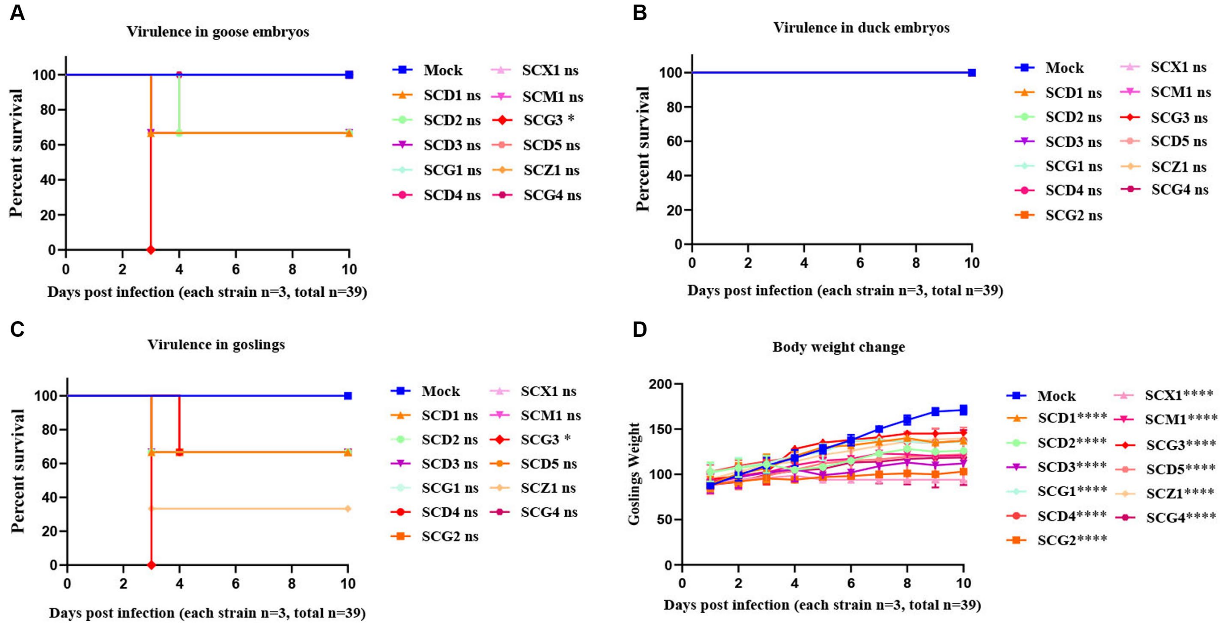


FIGURE 5
Virulence evaluation of goose astroviruses genotype-2 isolated from Sichuan. 10-day-old goose embryos and duck embryos from the experimental group (*n*=36) were infected intravenously with 2×10^4 $TCID_{50}$ of goose astroviruses genotype-2 (each strain *n*=3), while embryos from the control group (*n*=3) were mock infected with equal volume of PBS. **(A)** Virulence of goose astroviruses genotype-2 in goose embryos. **(B)** Virulence of goose astroviruses genotype-2 in duck embryos. 3-day-old goslings from the experimental group (*n*=36) were infected intravenously with goose astroviruses genotype-2 at 5×10^4 $TCID_{50}$ (each strain *n*=3), while goslings from the control group (*n*=3) were mock infected with equal volume of PBS. **(C)** Virulence of goose astroviruses genotype-2 in goslings. **(D)** Goslings were weighed at the indicated days post infection (dpi 1, 2, 3, 4, 5, 6, 7, 8, 9, 10) with goose astroviruses genotype-2. * Indicates statistically significant differences (ns indicates not significant, * indicates $p \leq 0.05$, **** indicates $p \leq 0.0001$).

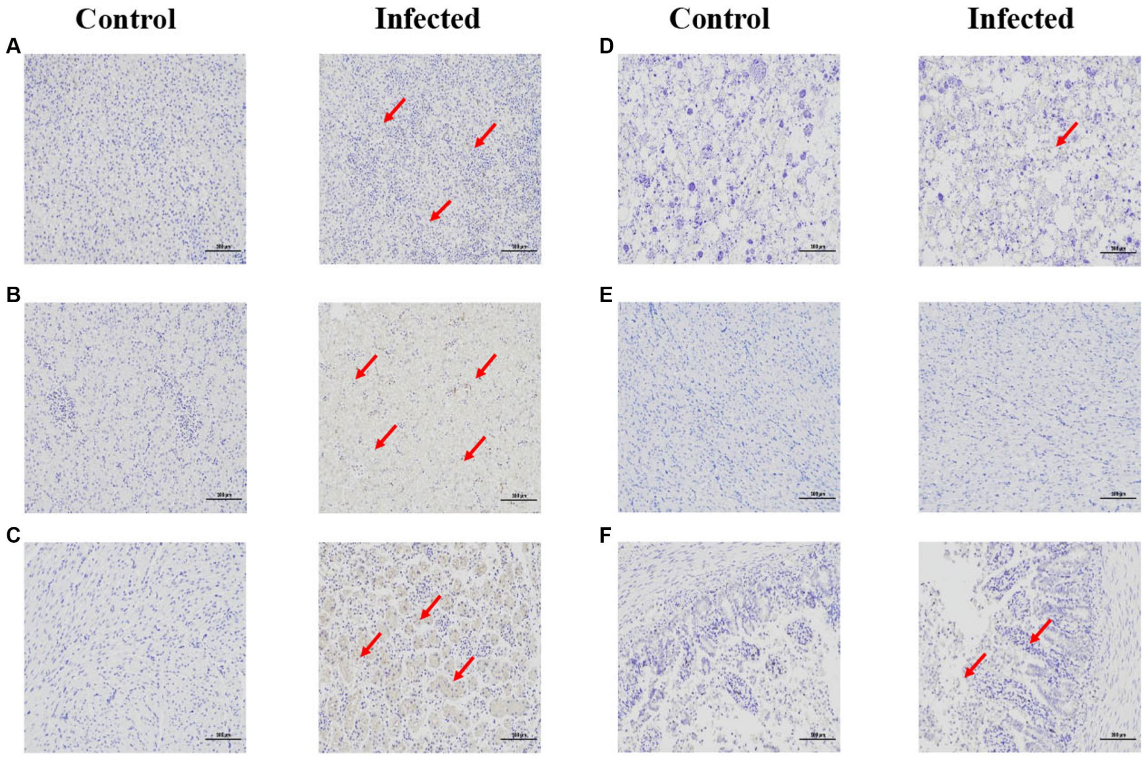


FIGURE 6
Immunohistochemical staining of infected and control tissues. 3-day-old goslings from the experimental group (*n* = 36) were infected intravenously with goose astroviruses genotype-2 at 5×10^4 $TCID_{50}$ (each strain *n* = 3), while goslings from the control group (*n* = 3) were mock infected with equal volume of PBS. **(A)** Spleen. **(B)** Liver. **(C)** Kidney. **(D)** Heart. **(E)** Brain. **(F)** Intestine. Goose astroviruses distribution in the dead gosling tissues post infection with goose astroviruses genotype-2 evaluated via an IHC analysis. Red arrows indicate positive signals. Scale bar, 100 μm.

contributes to the gradual formation of increased ammonia levels within the body (Xi et al., 2022). If the rate of urate production surpasses the excretion capacity of the urinary organs, it can result in uric acidemia (Xu et al., 2023). Some studies have demonstrated that a highly protein-rich diet for goslings can lead to crystal formation and gout (An et al., 2020). In cases where inflammation and obstruction occur in the kidney or ureter, uric acid excretion becomes obstructed, causing uric acid to accumulate in the bloodstream (Yuan et al., 2018). Uric acid salts precipitate on the surfaces of various organs, including the pleuras, pericardiums, peritoneums, mesenteries livers, kidneys, spleens, and intestines (Zhang X. et al., 2018). A shared characteristic of infectious factors is their potential to induce nephritis, kidney damage (Huang et al., 2021) and hindered uric acid excretion (Wu et al., 2021). These infectious factors encompass the infectious bronchitis virus, avian adenovirus inflammation, egg drop syndrome-76 (EDS-76), renal cryptosporidiosis, and most recently, GAstV.

Among the various viruses originating in geese that have emerged in Sichuan, GAstV comprises the largest proportion, accounting for 85.8% of cases. The isolated strains exhibit the closest genetic similarity to typical strains like G418, AHAU5, JX01, and others. This suggests that the GAstV outbreak in Sichuan likely originated from southeastern coastal areas of China, possibly spreading through horizontal transmission among goslings into the densely populated goose areas in Sichuan. In accordance with the International Committee on Taxonomy of Viruses (ICTV) classification, the categorization of genera is typically based on the p-distance between the complete amino acid sequences of the host and the structural protein. Viruses with a p-distance greater than 75% in the complete ORF2 protein sequence are typically classified as members of the same genotype. Through phylogenetic analysis, 12 isolated GAstV strains are classified as GAstV genotype-2 (G-2), significantly distant from strains belonging to GAstV genotype-1 (G-1). A thorough alignment of multiple sequences revealed substantial amino acid mutations in specific regions of the proteins encoded by these strains (Li et al., 2022). These mutations predominantly occur in the ORF2 genes, resulting in a significant elevation in blood uric acid levels (Liu et al., 2018). Bayesian analysis indicates that GAstV likely emerged in April 1985. The average phylogenetic rate of GAstV has been estimated to be approximately 1.42×10^{-3} nucleotide substitutions per site per year, suggesting a relatively slow evolutionary pace (Li et al., 2021).

In summary, the research on GAstV genotype-2 offers valuable insights into the pathological characteristics, epidemiology, genetic evolution, growth kinetics, virulence, transmissibility and pathogenicity. (1) Our research found that GAstV genotype-2 consistently exhibit the distinctive trait of depositing significant amounts of white urate in the viscera and joints. (2) The occurrence of GAstV outbreaks is concentrated in specific cities, notably, Deyang (61.9%), Chengdu (26.8%), followed by Meishan (6.2%), Zigong (4.1%), and Mianyang (1.0%). (3) GAstV infection in Sichuan primarily manifest as single GAstV infection, followed by double and

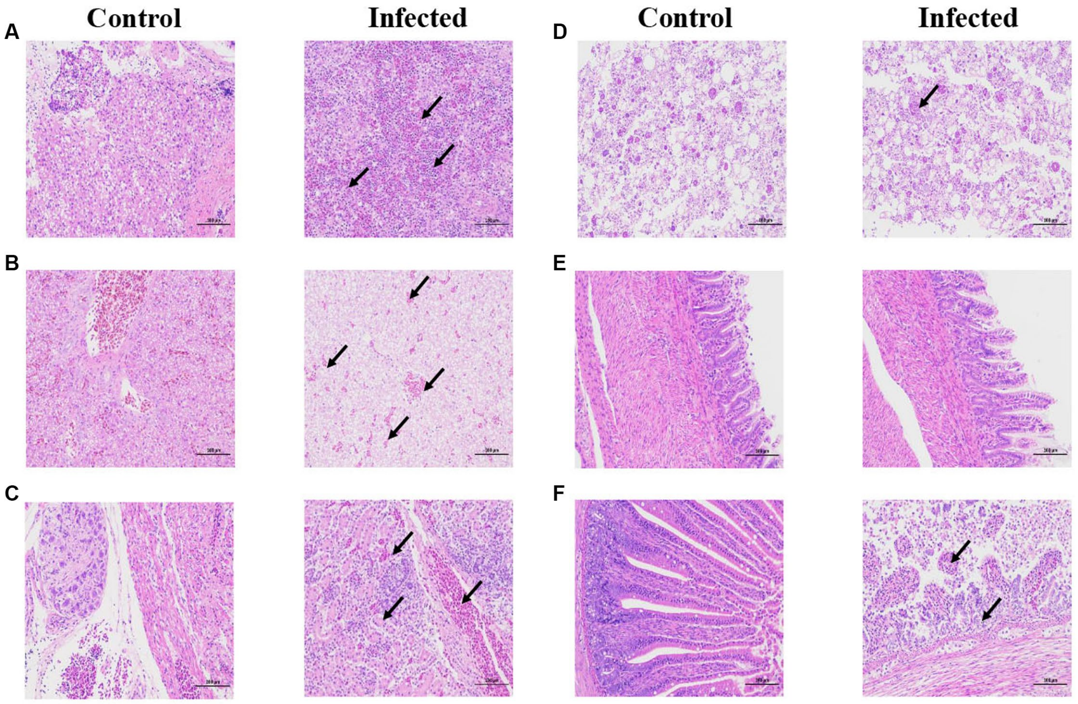


FIGURE 7
Histopathological staining of infected and control tissues. 3-day-old goslings from the experimental group ($n = 36$) were infected intravenously with goose astroviruses genotype-2 at 5×10^4 TCID$_{50}$ (each strain $n = 3$), while goslings from the control group ($n = 3$) were mock infected with equal volume of PBS. **(A)** Spleen. **(B)** Liver. **(C)** Kidney. **(D)** Heart. **(E)** Brain. **(F)** Intestine. Black arrows indicate different pathological characteristics. Scale bar, 100 μm.

(4) quadruple infection. GAstV typically coinfects with gosling plague viruses, goose reoviruses, adenoviruses, tembusu viruses, and avian influenza viruses. (5) Through analysis of conserved ORF1b genes and highly variable ORF2 genes, it has been established that all isolated GAstV strains prevalent in Sichuan belong to GAstV genotype-2, the earliest identified GAstV strains in Sichuan dates back to 2019. (6) Moreover, it is estimated that these prevalent GAstV strains have undergone evolutionary changes, characterized by an average substitution rate of 4.97×10^{-2} nucleotide substitutions. (7) Within this context, GAstV genotype-2 demonstrate varying degrees of virulence, ranging from highly virulent to weak. Each strain also showcases diverse replication situation to susceptible LMH cells, goose embryos, duck embryos, and goslings. Animal experiments underscore the varied virulence of different strains. Highly virulent strains can lead to significant mortality among goslings and goose embryos, whereas weaker strains exhibit limited lethality in these hosts. However, none of the strains were found to cause fatalities in duck embryos. (8) Despite the varying degrees of virulence, a shared characteristic across all strains is their ability to substantially hinder the growth of goslings. GAstV genotype-2 concentrated distributed in liver, kidney and spleen, followed by intestine, heart and brain. All GAstV genotype-2 exhibit different adaptations to different hosts and varying degrees of damage to tissues and organs. Overall, the strains with higher viral load have stronger pathogenicity and exhibit higher lethality against waterfowl in animal infection experiments. This comprehensive study can help researchers to better understand the molecular epidemiology and pathogenic mechanism of GAstV genotype-2.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request. Internal gene sequences of goose astroviruses genotype-2 SCD1 ORF1b, SCD1 ORF2, SCD2 ORF1b, SCD2 ORF2, SCD3 ORF1b, SCD3 ORF2, SCG1 ORF1b, SCG1 ORF2, SCD4 ORF1b, SCD4 ORF2, SCG2 ORF1b, SCG2 ORF2, SCG3 ORF1b, SCG3 ORF2, SCX1 ORF1b, SCX1 ORF2, SCM1 ORF1b, SCM1 ORF2, SCD5 ORF1b, SCD5 ORF2, SCZ1 ORF1b, SCZ1 ORF2, SCG4 ORF1b, SCG4 ORF2 can be found under the NCBI GenBank accession No. OR234621, No. OR234624, No. OR234622, No. OR234615, No. OR234607, No. OR234616, No. OR234610, No. OR234626, No. OR234608, No. OR234617, No. OR OR234611, No. OR234619, No. OQ909424, No. OQ909424, No. OR234613, No. OR234625, No. OR234623, No. OR234620, No. OR234609, No. OR234618, No. OR234614, No. OR234627, No. OR234612 and No. OR234628 respectively.

Ethics statement

The animal studies were approved by the Institutional Animal Care and Use Committee of Sichuan Agriculture University in Sichuan, China (Protocol Permit Number: SYXK(川)2019-187). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

LX: Conceptualization, Formal analysis, Investigation, Methodology, Software, Writing – original draft. BJ: Conceptualization, Methodology, Software, Validation, Writing – review & editing. YC: Conceptualization, Writing – review & editing. ZG: Conceptualization, Writing – review & editing. YH: Conceptualization, Validation, Writing – review & editing. ZW: Conceptualization, Writing – review & editing. MW: Conceptualization, Validation, Writing – review & editing. RJ: Conceptualization, Writing – review & editing. DZ: Conceptualization, Writing – review & editing. ML: Conceptualization, Validation, Writing – review & editing. XZ: Conceptualization, Writing – review & editing. QY: Conceptualization, Writing – review & editing. YW: Conceptualization, Writing – review & editing. SZ: Conceptualization, Writing – review & editing. JH: Writing – review & editing. XO: Conceptualization, Writing – review & editing. QG: Conceptualization, Writing – review & editing. DS: Conceptualization, Writing – review & editing. AC: Conceptualization, Writing – review & editing. SC: Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Key Research and Development Program of China (grant number 2022YFD1801900), National Natural Science Foundation of China (grant number 32272976), Sichuan Provincial Department of science and technology international scientific and technological innovation cooperation (grant number 2022YFH0026), the earmarked fund for China Agriculture Research System (grant number CARS-42-17), and the Program Sichuan Veterinary Medicine and Drug Innovation Group of China Agricultural Research System (grant number SCCXTD-2021-18).

Acknowledgments

The authors want to thank the reviewers for their very careful reading and serious suggestions to this manuscript. We appreciate SC for her rigorous advice on this manuscript and BJ for his instructions on the drawing of this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

An, D., Zhang, J., Yang, J., Tang, Y., and Diao, Y. (2020). Novel goose-origin astrovirus infection in geese: the effect of age at infection. *Poult. Sci.* 99, 4323–4333. doi: 10.1016/j.psj.2020.05.041

Bidin, M., Bidin, Z., Majnaric, D., Tisljar, M., and Lojkic, I. (2012a). Circulation and phylogenetic relationship of chicken and Turkey-origin astroviruses detected in domestic ducks (*Anas platyrhynchos* domesticus). *Avian Pathol.* 41, 555–562. doi: 10.1080/03079457.2012.733340

Bidin, M., Lojkic, I., Tisljar, M., Bidin, Z., and Majnaric, D. (2012b). Astroviruses associated with stunting and pre-hatching mortality in duck and goose embryos. *Avian Pathol.* 41, 91–97. doi: 10.1080/03079457.2011.642796

Chen, Q., Xu, X., Yu, Z., Sui, C., Zuo, K., Zhi, G., et al. (2020). Characterization and genomic analysis of emerging astroviruses causing fatal gout in goslings. *Transbound. Emerg. Dis.* 67, 865–876. doi: 10.1111/tbed.13410

Chen, Q., Yu, Z., Xu, X., Ji, J., Yao, L., Kan, Y., et al. (2021). First report of a novel goose astrovirus outbreak in Muscovy ducklings in China. *Poult. Sci.* 100:101407. doi: 10.1016/j.psj.2021.101407

Chen, H., Zhang, B., Yan, M., Diao, Y., and Tang, Y. (2020). First report of a novel goose astrovirus outbreak in Cherry Valley ducklings in China. *Transbound. Emerg. Dis.* 67, 1019–1024. doi: 10.1111/tbed.13418

Cui, H., Mu, X., Xu, X., Ji, J., Ma, K., Leng, C., et al. (2022). Extensive genetic heterogeneity and molecular characteristics of emerging astroviruses causing fatal gout in goslings. *Poult. Sci.* 101:101888. doi: 10.1016/j.psj.2022.101888

Dalbeth, N., Gosling, A. L., Gaffo, A., and Abhishek, A. (2021). Gout. *Lancet* 397, 1843–1855. doi: 10.1016/S0140-6736(21)00569-9

Ding, R., Huang, H., Wang, H., Yi, Z., Qiu, S., Lv, Y., et al. (2021). Goose nephritic Astrovirus infection of goslings induces lymphocyte apoptosis, reticular Fiber destruction, and CD8 T-cell depletion in spleen tissue. *Viruses* 13:1108. doi: 10.3390/v13061108

Fei, Z., Jiao, A., Xu, M., Wu, J., Wang, Y., Yu, J., et al. (2022). Genetic diversity and evolution of goose astrovirus in the east of China. *Transbound. Emerg. Dis.* 69, e2059–e2072. doi: 10.1111/tbed.14542

Fu, X., Hou, Z., Liu, W., Cao, N., Liang, Y., Li, B., et al. (2022). Insight into the epidemiology and evolutionary history of novel goose Astrovirus-associated gout in goslings in southern China. *Viruses* 14:1306. doi: 10.3390/v14061306

He, D., Wang, F., Zhao, L., Jiang, X., Zhang, S., Wei, F., et al. (2022). Epidemiological investigation of infectious diseases in geese on mainland China during 2018-2021. *Transbound. Emerg. Dis.* 69, 3419–3432. doi: 10.1111/tbed.14699

He, D., Yang, J., Jiang, X., Lin, Y., Chen, H., Tang, Y., et al. (2020). A quantitative loop-mediated isothermal amplification assay for detecting a novel goose astrovirus. *Poult. Sci.* 99, 6586–6592. doi: 10.1016/j.psj.2020.09.077

Huang, H., Ding, R., Chen, Z., Yi, Z., Wang, H., Lv, Y., et al. (2021). Goose nephritic astrovirus infection increases autophagy, destroys intercellular junctions in renal tubular epithelial cells, and damages podocytes in the kidneys of infected goslings. *Vet. Microbiol.* 263:109244. doi: 10.1016/j.vetmic.2021.109244

Ji, J., Chen, Q., Sui, C., Hu, W., Yu, Z., Zhang, Z., et al. (2020). Rapid and visual detection of novel astroviruses causing fatal gout in goslings using one-step reverse transcription loop-mediated isothermal amplification. *Poult. Sci.* 99, 4259–4264. doi: 10.1016/j.psj.2020.05.024

Li, J. Y., Hu, W. Q., Liu, T. N., Zhang, H. H., Opriessnig, T., and Xiao, C. T. (2021). Isolation and evolutionary analyses of gout-associated goose astrovirus causing disease in experimentally infected chickens. *Poult. Sci.* 100, 543–552. doi: 10.1016/j.psj.2020.11.003

Li, L., Sun, M., Zhang, Y., and Liao, M. (2022). A review of the emerging poultry visceral gout disease linked to avian Astrovirus infection. *Int. J. Mol. Sci.* 23:429. doi: 10.3390/ijms231810429

Liao, Q., Liu, N., Wang, X., Wang, F., and Zhang, D. (2015). Genetic characterization of a novel astrovirus in Pekin ducks. *Infect. Genet. Evol.* 32, 60–67. doi: 10.1016/j.meegid.2015.02.025

Liu, H., Hu, D., Zhu, Y., Xiong, H., Lv, X., Wei, C., et al. (2020). Coinfection of parvovirus and astrovirus in gout-affected goslings. *Transbound. Emerg. Dis.* 67, 2830–2838. doi: 10.1111/tbed.13652

Liu, N., Jiang, M., Dong, Y., Wang, X., and Zhang, D. (2018). Genetic characterization of a novel group of avastroviruses in geese. *Transbound. Emerg. Dis.* 65, 927–932. doi: 10.1111/tbed.12873

Liu, M., Zhao, Y., Hu, D., Huang, X., Xiong, H., Qi, K., et al. (2019). Clinical and histologic characterization of co-infection with Astrovirus and goose parvovirus in goslings. *Avian Dis.* 63, 731–736. doi: 10.1637/aviandiseases-D-19-00110

Niu, X., Tian, J., Yang, J., Jiang, X., Wang, H., Chen, H., et al. (2018). Novel goose Astrovirus associated gout in Gosling, China. *Vet. Microbiol.* 220, 53–56. doi: 10.1016/j.vetmic.2018.05.006

Ren, D., Li, T., Zhang, X., Yao, X., Gao, W., Xie, Q., et al. (2020). OASL triggered by novel goose Astrovirus via ORF2 restricts its replication. *J. Virol.* 94. doi: 10.1128/JVI.01767-20

Ren, D., Li, T., Zhang, W., Zhang, X., Zhang, X., Xie, Q., et al. (2022). Identification of three novel B cell epitopes in ORF2 protein of the emerging goose astrovirus and their application. *Appl. Microbiol. Biotechnol.* 106, 855–863. doi: 10.1007/s00253-021-11711-2

Wan, C., Chen, C., Cheng, L., Fu, G., Shi, S., Liu, R., et al. (2019). Specific detection of the novel goose astrovirus using a TaqMan real-time RT-PCR technology. *Microb. Pathog.* 137:103766. doi: 10.1016/j.micpath.2019.103766

Wang, Y., Bai, C., Zhang, D., Yang, K., Yu, Z., Jiang, S., et al. (2020). Genomic and phylogenetic characteristics of a novel goose astrovirus in Anhui Province, Central-Eastern China. *Gene* 756:144898. doi: 10.1016/j.gene.2020.144898

Wang, Z., Chen, H., Gao, S., Song, M., Shi, Z., Peng, Z., et al. (2022). Core antigenic advantage domain-based ELISA to detect antibody against novel goose astrovirus in breeding geese. *Appl. Microbiol. Biotechnol.* 106, 2053–2062. doi: 10.1007/s00253-022-11852-y

Wang, A., Liu, L., Zhang, S., Ye, W., Zheng, T., Xie, J., et al. (2022). Development of a duplex real-time reverse transcription-polymerase chain reaction assay for the simultaneous detection of goose astrovirus genotypes 1 and 2. *J. Virol. Methods* 310:114612. doi: 10.1016/j.jviromet.2022.114612

Wang, A. P., Zhang, S., Xie, J., Gu, L. L., Wu, S., Wu, Z., et al. (2021). Isolation and characterization of a goose astrovirus 1 strain causing fatal gout in goslings, China. *Poult Sci.* 100:101432. doi: 10.1016/j.psj.2021.101432

Wang, H., Zhu, Y., Ye, W., Hua, J., Chen, L., Ni, Z., et al. (2022). Genomic and epidemiological characteristics provide insights into the phylogeographic spread of goose astrovirus in China. *Transbound. Emerg. Dis.* 69, e1865–e1876. doi: 10.1111/tbed.14522

Wei, F., Yang, J., He, D., Diao, Y., and Tang, Y. (2020a). Evidence of vertical transmission of novel astrovirus virus in goose. *Vet. Microbiol.* 244:108657. doi: 10.1016/j.vetmic.2020.108657

Wei, F., Yang, J., Wang, Y., Chen, H., Diao, Y., and Tang, Y. (2020b). Isolation and characterization of a duck-origin goose astrovirus in China. *Emerg. Microbes Infect.* 9, 1046–1054. doi: 10.1080/22221751.2020.1765704

Wu, W., Qiu, S., Huang, H., Xu, R., Bao, E., and Lv, Y. (2021). Immune-related gene expression in the kidneys and spleens of goslings infected with goose nephritic astrovirus. *Poult. Sci.* 100:100990. doi: 10.1016/j.psj.2021.01.013

Wu, W., Xu, R., Lv, Y., and Bao, E. (2020). Goose astrovirus infection affects uric acid production and excretion in goslings. *Poult. Sci.* 99, 1967–1974. doi: 10.1016/j.psj.2019.11.064

Xi, Y., Huang, Y., Li, Y., Huang, Y., Yan, J., and Shi, Z. (2022). The effects of dietary protein and fiber levels on growth performance, gout occurrence, intestinal microbial communities, and immunoregulation in the gut-kidney axis of goslings. *Poult. Sci.* 101:101780. doi: 10.1016/j.psj.2022.101780

Xi, Y., Huang, Y., Li, Y., Yan, J., and Shi, Z. (2020). Fermented feed supplement relieves Caecal microbiota Dysbiosis and kidney injury caused by high-protein diet in the development of Gosling gout. *Animals (Basel)* 10:2139. doi: 10.3390/ani10112139

Xu, L., Jiang, B., Cheng, Y., He, Y., Wu, Z., Wang, M., et al. (2023). Infection and innate immune mechanism of goose astrovirus. *Front. Microbiol.* 14:1121763. doi: 10.3389/fmicb.2023.1121763

Xu, D., Li, C., Liu, G., Chen, Z., and Jia, R. (2019). Generation and evaluation of a recombinant goose origin Newcastle disease virus expressing cap protein of goose origin avastrovirus as a bivalent vaccine in goslings. *Poult. Sci.* 98, 4426–4432. doi: 10.3382/ps/pez255

Yang, J., Tian, J., Tang, Y., and Diao, Y. (2018). Isolation and genomic characterization of gosling gout caused by a novel goose astrovirus. *Transbound. Emerg. Dis.* 65, 1689–1696. doi: 10.1111/tbed.12928

Yang, X., Wei, F., Tang, Y., and Diao, Y. (2021). Development of immunochromatographic strip assay for rapid detection of novel goose astrovirus. *J. Virol. Methods* 297:114263. doi: 10.1016/j.jviromet.2021.114263

Yi, Z., Ding, R., Cao, R., Sun, W., Sun, M., Dong, Y., et al. (2022). Development of a duplex TaqMan real-time RT-PCR assay for simultaneous detection of goose astrovirus genotypes 1 and 2. *J. Virol. Methods* 306:114542. doi: 10.1016/j.jviromet.2022.114542

Yin, D., Tian, J., Yang, J., Tang, Y., and Diao, Y. (2021). Pathogenicity of novel goose-origin astrovirus causing gout in goslings. *BMC Vet. Res.* 17:40. doi: 10.1186/s12917-020-02739-z

Yin, D., Yang, J., Tian, J., He, D., Tang, Y., and Diao, Y. (2020). Establishment and application of a TaqMan-based one-step real-time RT-PCR for the detection of novel goose-origin astrovirus. *J. Virol. Methods* 275:113757. doi: 10.1016/j.jviromet.2019.113757

Yuan, X., Meng, K., Zhang, Y., Qi, L., Ai, W., and Wang, Y. (2018). Establishment and application of rapid diagnosis for reverse transcription-quantitative PCR of newly emerging goose-origin nephrotic Astrovirus in China. *mSphere* 3:380. doi: 10.1128/mSphere.00380-18

Yuan, X., Meng, K., Zhang, Y., Yu, Z., Ai, W., and Wang, Y. (2019). Genome analysis of newly emerging goose-origin nephrotic astrovirus in China reveals it belongs to a novel genetically distinct astrovirus. *Infect. Genet. Evol.* 67, 1–6. doi: 10.1016/j.meegid.2018.10.014

Zhang, Q., Cao, Y., Wang, J., Fu, G., Sun, M., Zhang, L., et al. (2018). Isolation and characterization of an astrovirus causing fatal visceral gout in domestic goslings. *Emerg. Microbes Infect.* 7, 71–11. doi: 10.1038/s41426-018-0074-5

Zhang, X., Deng, T., Song, Y., Liu, J., Jiang, Z., Peng, Z., et al. (2022). Identification and genomic characterization of emerging goose astrovirus in Central China, 2020. *Transbound. Emerg. Dis.* 69, 1046–1055. doi: 10.1111/tbed.14060

Zhang, F., Li, H., Wei, Q., Xie, Q., Zeng, Y., Wu, C., et al. (2022). Isolation and phylogenetic analysis of goose astrovirus type 1 from goslings with gout in Jiangxi province, China. *Poult. Sci.* 101:101800. doi: 10.1016/j.psj.2022.101800

Zhang, M., Lv, X., Wang, B., Yu, S., Lu, Q., Kan, Y., et al. (2022). Development of a potential diagnostic monoclonal antibody against capsid spike protein VP27 of the novel goose astrovirus. *Poult. Sci.* 101:101680. doi: 10.1016/j.psj.2021.101680

Zhang, X., Ren, D., Li, T., Zhou, H., Liu, X., Wang, X., et al. (2018). An emerging novel goose astrovirus associated with gosling gout disease, China. *Emerg. Microbes Infect.* 7:152. doi: 10.1038/s41426-018-0153-7

Zhu, Q., Miao, Y., Wang, J., Bai, W., Yang, X., Yu, S., et al. (2022). Isolation, identification, and pathogenicity of a goose astrovirus causing fatal gout in goslings. *Vet. Microbiol.* 274:109570. doi: 10.1016/j.vetmic.2022.109570

Zhu, Q., and Sun, D. (2022). Goose astrovirus in China: a comprehensive review. *Viruses* 14:1759. doi: 10.3390/v14081759

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 19 December 2023
DOI 10.3389/fmicb.2023.1309156

EDITED BY
Mohamad S. Hakim,
Gadjah Mada University, Indonesia

REVIEWED BY
Nongluk Sriwilaijaroen,
Thammasat University, Thailand
Yanbing Li,
Chinese Academy of Agricultural Sciences,
China

*CORRESPONDENCE
Mikhail Matrosovich
✉ m.matrosovich@gmail.com
Stefano Elli
✉ elli@ronzoni.it
Sergei Kosakovsky Pond
✉ spond@temple.edu

RECEIVED 07 October 2023
ACCEPTED 29 November 2023
PUBLISHED 19 December 2023

CITATION
Elli S, Raffaini G, Guerrini M, Kosakovsky Pond S and Matrosovich M (2023) Molecular modeling and phylogenetic analyses highlight the role of amino acid 347 of the N1 subtype neuraminidase in influenza virus host range and interspecies adaptation.
Front. Microbiol. 14:1309156.
doi: 10.3389/fmicb.2023.1309156

Molecular modeling and phylogenetic analyses highlight the role of amino acid 347 of the N1 subtype neuraminidase in influenza virus host range and interspecies adaptation

Stefano Elli[1]*, Giuseppina Raffaini[2], Marco Guerrini[1], Sergei Kosakovsky Pond[3]* and Mikhail Matrosovich[4]*

[1]Istituto di Ricerche Chimiche e Biochimiche 'G. Ronzoni', Milan, Italy, [2]Department of Chemistry, Materials, and Chemical Engineering "Giulio Natta", Politecnico di Milano, Milan, Italy, [3]Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA, United States, [4]Institute of Virology, Philipps University, Marburg, Germany

The N1 neuraminidases (NAs) of avian and pandemic human influenza viruses contain tyrosine and asparagine, respectively, at position 347 on the rim of the catalytic site; the biological significance of this difference is not clear. Here, we used molecular dynamics simulation to model the effects of amino acid 347 on N1 NA interactions with sialyllacto-N-tetraoses 6′SLN-LC and 3′SLN-LC, which represent NA substrates in humans and birds, respectively. Our analysis predicted that Y347 plays an important role in the NA preference for the avian-type substrates. The Y347N substitution facilitates hydrolysis of human-type substrates by resolving steric conflicts of the Neu5Ac2−6Gal moiety with the bulky side chain of Y347, decreasing the free energy of substrate binding, and increasing the solvation of the Neu5Ac2−6Gal bond. Y347 was conserved in all N1 NA sequences of avian influenza viruses in the GISAID EpiFlu database with two exceptions. First, the Y347F substitution was present in the NA of a specific H6N1 poultry virus lineage and was associated with the substitutions G228S and/or E190V/L in the receptor-binding site (RBS) of the hemagglutinin (HA). Second, the highly pathogenic avian H5N1 viruses of the Gs/Gd lineage contained sporadic variants with the NA substitutions Y347H/D, which were frequently associated with substitutions in the HA RBS. The Y347N substitution occurred following the introductions of avian precursors into humans and pigs with N/D347 conserved during virus circulation in these hosts. Comparative evolutionary analysis of site 347 revealed episodic positive selection across the entire tree and negative selection within most host-specific groups of viruses, suggesting that substitutions at NA position 347 occurred during host switches and remained under pervasive purifying selection thereafter. Our results elucidate the role of amino acid 347 in NA recognition of sialoglycan substrates and emphasize the significance of substitutions at position 347 as a marker of host range and adaptive evolution of influenza viruses.

KEYWORDS

influenza, neuraminidase, substrate specificity, MD simulation, natural selection, H5N1

1 Introduction

Wild aquatic birds are a primary natural reservoir of influenza A viruses (IVs). Transmission and occasional adaptation of aquatic bird viruses to other avian and mammalian species, with or without gene reassortment and/or additional host-switching events, led to the formation of the known variety of HxNy host-specific lineages of IVs, such as H5N1, H7N9 and H9N2 poultry IVs, H1N1 "classical" and "avian-like" swine IVs, H3N8 equine IVs, H3N8 and H3N2 canine IVs. Infrequent transmissions of animal IVs to humans result in isolated zoonotic infections; on very rare occasions, zoonotic IVs can adapt for efficient human-to-human transmission, initiate global pandemics, and continue to circulate in humans causing seasonal influenza disease [for recent reviews, see Krammer et al. (2018) and Liu et al. (2022)].

Interspecies transmission of IVs is usually accompanied by adaptive changes in their genomes required for improved viral fitness in the new host species. Adaptive changes in the receptor-binding properties of IVs and underlying amino acid substitutions in the attachment protein HA have been relatively well studied [reviewed by Matrosovich et al. (2006), Shi et al. (2014), Thompson and Paulson (2020), and Liu et al. (2023)]. Avian-to-human and avian-to-swine adaptation was shown to require a change in HA receptor-binding specificity from preferential recognition of Neu5Ac2–3Gal-terminated glycans ("avian-type" receptors) expressed on avian intestinal target cells to preferential recognition of Neu5Ac2–6Gal-terminated glycans ("human-type" receptors) expressed on target cells of airway epithelium in humans and pigs. Changes in receptor specificity were mediated by amino acid substitutions at the conserved positions of the avian HA, such as G228S and/or Q226L (the H3 numbering system is used throughout the text) in the case of H2N2/1957 and H3N2/1968 pandemic IVs and G225D/E and/or E190D (H1N1/1918 pandemic IV, H1N1 classical and avian-like swine IVs). Substitutions at some of these four "canonical" positions and/or several other HA positions of poultry-adapted IVs and sporadic mammalian isolates with H4, H5, H6, H7, H9, and H10 HAs have been found to facilitate binding to human-type receptors (Paulson and de Vries, 2013; Thompson and Paulson, 2020; Liu et al., 2022, 2023).

The primary function of IV neuraminidase is to attenuate HA interactions with decoy receptors [for reviews on structure and functions of IV NA, see von Itzstein (2007), Shtyrya et al. (2009), and McAuley et al. (2019)]. In the early stages of infection, NA removes sialic acid receptors from soluble sialylglycoproteins, mucous blanket, and the cellular glycocalyx, thereby facilitating virus motility and access to functional receptors on the cell membrane. In the late stages of infection, NA desialylates viral progeny and the surface of infected cells, preventing virus aggregation and promoting its release. As the receptor-destroying activity of NA counteracts the receptor-binding activity of HA, a balance of these activities with respect to the spectrum of sialoglycans present in the target host tissues is essential for IV fitness (Wagner et al., 2002; de Vries et al., 2020). However, in contrast to HA, there is much less understanding of NA specificity for Neu5Ac2–3Gal- and Neu5Ac2–6Gal-containing sialoglycans and of alterations in its substrate specificity during avian-to-mammalian adaptation of IVs.

Studies on the desialylation of soluble monovalent sialyloligosaccharides showed that N1 and N2 NAs of avian IVs predominantly hydrolyze Neu5Ac2–3Gal-terminated substrates, whereas NAs of human and porcine viruses have dual specificity with less efficient cleavage of Neu5Ac2–6Gal compared to Neu5Ac2–3Gal (Baum and Paulson, 1991; Couceiro and Baum, 1994; Kobasa et al., 1999; Mochalova et al., 2007; Gerlach et al., 2012; Garcia et al., 2013). The catalytic activity correlated with the binding avidity of the substrates to the NA, suggesting that the NA specificity for the avian-type and human-type sialoglycans depends on differences in their binding to the catalytic site (Mochalova et al., 2007; Garcia et al., 2013). Avian-origin N2 NA was introduced into the human population with the H2N2/1957 pandemic IV. Pandemic and early post-pandemic H2N2 virus strains had avian-type substrate specificity; the first detectable increase in their ability to cleave Neu5Ac2–6Gal was observed with A/England/12/1962 and correlated with the acquisition of the I275V amino acid substitution in the NA (Kobasa et al., 1999). The molecular mechanism behind the effect of this substitution on NA specificity remains undefined.

The NA of the H1N1/1918 pandemic IV differed from the closest avian N1 NAs by about 30 amino acid residues, including the Y347N substitution (the N2 numbering system is used throughout the text; position 347 corresponds to codon 344 of the N1 NA gene). This substitution was predicted to be a marker for mammalian adaptation (Reid et al., 2000) because it was also present in the N1 NA of the swine-adapted IV lineage that emerged from the independent introduction of an avian H1N1 virus into European swine in the late 1970s. In the crystal structures of N1 NA, the side chain of amino acid 347 protrudes into the solvent at the rim of the catalytic site; its main chain carbonyl oxygen forms a part of the Ca^{2+}-binding site (Russell et al., 2006; Xu et al., 2008). Several research groups have found that the identity of amino acid at position 347 affects N1 NA avidity for the substrate, catalytic activity, and sensitivity to the NA inhibitors oseltamivir and zanamivir (Collins et al., 2009; Rameix-Welti et al., 2011; Yongkiettrakul et al., 2013; Dai et al., 2016; Su et al., 2021). Although these results demonstrated a contribution of amino acid 347 to substrate binding and catalysis, they were obtained using the non-natural substrate MUNANA and did not provide information on the effect of amino acid 347 on NA specificity toward Neu5Ac2–3Gal- and Neu5Ac2–6Gal-terminated sialoglycans.

X-ray analysis of NA binding to sialoglycans could not be performed for technical reasons. Therefore, molecular dynamics (MD) simulation was used to study the interactions of Neu5Ac2–3Gal- and Neu5Ac2–6Gal-containing trisaccharides with N1 NAs from H5N1 avian and H1N1 human IVs (Raab and Tvaroska, 2011; Jongkon and Sangma, 2012; Phanich et al., 2019). Collectively, these studies predicted that the catalytic sites of avian-type NAs favor binding to Neu5Ac2–3Gal- over Neu5Ac2–6Gal-containing substrates, whereas human-type NAs bind both substrates. Binding specificity was related, in part, to the poor fit of the Neu5Ac2–6Gal-containing glycan, in its bended solution-dominant conformation, into the catalytic site of the avian-type NAs. In contrast, the nearly linear conformers of the Neu5Ac2–3Gal-containing glycan were accommodated by both avian-type and human-type NAs. The amino acid at position 347 was predicted to contribute to binding specificity by steric hindrance and by interactions with the Neu5Ac and Gal moieties of the substrate. Because these modeling experiments used wild-type N1 NAs separated by multiple amino acid substitutions, further studies are needed to confirm and specify the predictions regarding the role of amino acid 347 in substrate specificity of N1 NA.

Here, we used molecular modeling to characterize complexes of two N1 NAs differing by the single substitution N347Y with two pentasaccharides representing models of the human-type and avian-type NA substrates. The NA-ligand complexes were built by molecular docking, and their geometry was refined by hundred-nanosecond MD simulation in explicit solvent. Using these models, we determined the effects of amino acid 347 on the conformation and atomic interactions of the ligand in the catalytic site, the free energy of binding, and the solvation of the glycosidic oxygen. In addition, we performed analysis of the host- and lineage-specific variation of the amino acid 347 among N1 NAs in the GISAID EpiFlu sequence database and assessed selective pressures at this position using comparable phylogenetic dN/dS techniques.

2 Materials and methods

2.1 Molecular modeling

2.1.1 Models of sialoglycans and NAs

Two pentasaccharides, α-D-Neu5Ac(2 → 6)β-D-Gal(1 → 4)β-D-GlcNAc(1 → 3)β-D-Gal(1 → 4)β-D-Glc-OH, referred to in this study as 6S, and α-D-Neu5Ac(2 → 3)β-D-Gal(1 → 4)β-D-GlcNAc(1 → 3)β-D-Gal(1 → 4)β-D-Glc-OH (referred to as 3S) were used as models of the human-type and avian-type glycan receptors, respectively (Supplementary Figure S1). The pyranose ring conformations of 6S and 3S were set as chair $^2C_5$ for Neu5Ac and as $^4C_1$ for the remaining saccharides. The glycosidic backbone conformation of these glycans was defined by the set of dihedral angles ϕ_i/ψ_i, i = 1–4. The Neu5Ac2–3Gal linkage of 3S was defined by ϕ_1 (C1-C2-O2-C3) and ψ_1 (C2-O3-C3-H3). The Neu5Ac2–6Gal linkage of 6S was defined by ϕ_1 (C1-C2-O6-C6), ψ_1 (C2-O6-C6-C5), and ω (O6-C6-C5-H5). The other pairs of glycosidic dihedral angles were defined by four consecutive atoms H_i-C_i-O_{i+1}-C_{i+1} and C_i-O_{i+1}-C_{i+1}-H_{i+1}. The conformations of the glycosidic bonds were set according to the previously determined dominant unbound state conformations of the glycans in solution (Sassaki et al., 2013; Elli et al., 2021). The coordinate files of the two glycans were generated using the tleap application included in Ambertools 14.0 (Case et al., 2014). The crystal structure of the NA of the H1N1/1918 pandemic influenza virus A/Brevig Mission/1/1918 (3B7E) was obtained from the RCSB Protein Databank and designated NA/N in accordance with the nature of the amino acid at position 347. The N347Y point mutant of NA/N (designated NA/Y) was generated using PyMOL 2.4.0 (Schrödinger, LLC) with the default settings for the side chain rotational states.

2.1.2 Molecular docking

The ligands, 3S and 6S, were docked into the catalytic sites of NA/Y and NA/N using Autodock 4.2 (Morris et al., 2009) to generate four complexes, 3S-NA/Y, 6S-NA/Y, 3S-NA/N, and 6S-NA/N. To perform the docking, we determined the Gasteiger charges (Gasteiger and Marsili, 1978) for the atoms of the glycans (ligands) and NAs (acceptors). The conformational sampling of 6S and 3S during docking was set to allow a full conformational exploration of the Neu5Ac-Gal linkages. In contrast, the inter-glycosidic dihedrals ϕ_2/ψ_2, ϕ_3/ψ_3 and the conformation of the whole disaccharide Gal1-4Glc were fixed in the docking simulations. Each docking simulation sampled 21 free rotational dihedral angles. The NAs were set as rigid macromolecules. The grid box dimensions were Lx = Ly = 80 points, Lz = 100 points (spacing Δx = Δy = Δz = 0.375 Å); the center of the grid box was set at the hydroxyl oxygen of Y406, the key catalytic residue occupying the central position at the base of the NA active site (Taylor and von Itzstein, 1994). In each simulation, the Lamarckian genetic algorithm search was performed with the following parameters: number of runs, 100; population size, 4,000; maximum number of energy evaluations, $3{\cdot}10^7$; maximum number of generations, 300,000. The ligand-NA poses were selected for further geometry refinement by MD simulation based on the following criteria: (1) the position of the Neu5Ac residue in the center of the NA active site among the conserved residues R118, R292, R371, W178, D151, R152, and Y406; (2) the lowest predicted (Autodock 4.2) binding energy.

2.1.3 MD simulation of the complexes 3S-NA/Y, 6S-NA/Y, 3S-NA/N, and 6S-NA/N

The four glycan-NA complexes obtained by docking were submitted to MD simulation in explicit solvent. Each complex was surrounded by a 15 Å-wide layer of TIP3P water molecules (Jorgensen et al., 1983) to form an orthogonal box (simulation cell) with edges of approximately 90 Å. The molecular mechanics force fields used, GLYCAM06 (Kirschner et al., 2008) and Amber (ff14SB) (Case et al., 2014), represented the "state of the art" force fields for glycans and proteins, respectively. Tleap (Ambertools 14.0) was used to build the topology and coordinate files of the complexes. The standard cut-off (12 Å) was applied to describe non-bonded electrostatic and dispersive interactions. Each simulation cell was minimized by running 100 K steps of the default minimization algorithm included in the NAMD 2.14 (Phillips et al., 2005). The software VMD 1.9.3 (Humphrey et al., 1996) was used for the MD simulation visualization and analysis. The number of particles (N), pressure (P), and temperature (T) were kept constant during the MD simulations. The constant simulation temperature (300 K) was maintained with a Lowe–Andersen thermostat; the Nosé–Hoover–Langevin piston algorithm controlled the pressure (1.01325 bar) applied to the cell walls. During the cell density equilibration steps (duration approximately 10–15 ns), a harmonic potential energy constraint (harmonic constant of 2.0 kcal mol^{-1}) was applied to all atoms of the complexes, while water molecules were allowed to move freely. No additional constraints were applied at the equilibration and production stages of the MD simulation. For each simulated complex the equilibration period was estimated following the time evolution of two different properties, (1) the glycosidic dihedral angles $\phi_i(t)/\psi_i(t)$ (i = 1,4) of the glycan backbone; (2) the distance RMSD(t) of the glycan from its initial position ($t = 0$ ns); a stationary oscillatory behavior of these properties over time indicated the end of the equilibration and the beginning of the production MD simulation. During the MD simulation, all complexes were sampled every 10 ps for a period of 195 ns. The equilibration time was 110 ns for 6S-NA/N, 3S-NA/N and 3S-NA/Y, and 120 ns for 6S-NA/Y.

2.1.4 Analysis of conformations of 6S and 3S

The conformational analysis of the 6S and 3S in the unbound state was performed previously (Elli et al., 2021), and the results are presented in this study with permission from the Biochemical Journal via the Copyright Clearance Center. The conformations of 6S and 3S bound to the catalytic site of NA/N and NA/Y were sampled by MD simulation, and the Ramachandran plots $\phi_i(t)/\psi_i(t)$ with corresponding density color maps for the four glycosidic linkages connecting the five sugar residues of the ligands were calculated. In these plots, the color gradient from blue to red is proportional to the density of states that were sampled by the MD simulations (from 110 or 120 ns to 195 ns). The Ramachandran plots, the 2D binning procedure and the density color maps were generated using the statistical software R (R Core Team, 2013), as described previously (Elli et al., 2021). The most populated states of the ω angle of 6S and of the internal dihedral angle C1-C2-C3-C4 of Neu5Ac (see Supplementary Figure S1) were determined at the production stage of the MD simulation using the 1D binning procedure implemented in the OriginPro 8.0 software (OriginLab Corp.). The internal dihedral angle C1-C2-C3-C4 was used to monitor the conformation of the Neu5Ac moiety. Values of this angle around 60° correspond to the chair $^2C_5$ conformation typical for the terminal Neu5Ac residue in sialoglycans; values around ±180° characterize the distorted boat conformation observed in the crystal structures of NA complexes with the free Neu5Ac molecule (Varghese et al., 1992).

2.1.5 Structure of the glycan-NA complexes

The structures of the NA complexes with 3S and 6S were determined by MD simulation as follows. First, the average structure of each glycan-NA complex was calculated for the production stage of the MD simulation using the Wordom 0.22-rc3 software (Seeber et al., 2007). Next, the MD snapshot having the smallest root mean square distance (RMSD) from the average structure of the complex was extracted from the MD simulation trajectory using VMD 1.9.3 software. Thus, the MD simulation snapshots taken at times 142.03, 150.44, 147.95, and 141.59 ns were selected to represent the complexes 6S-NA/N, 6S-NA/Y, 3S-NA/N, and 3S-NA/Y, respectively. These snapshots had RMSDs of 0.494, 0.511, 0.495, and 0.507 Å, respectively, from the corresponding average structures.

2.1.6 Free energy of binding determined using MMPBSA approximation

To estimate the Poisson Boltzmann free energy of binding (ΔG_{PB}^{bind}) from the MD simulation trajectories of the glycan-NA complexes, we used Molecular Mechanics Poisson Boltzmann Surface Area (MMPBSA) method described previously (Weis et al., 2006; Elli et al., 2021). For each glycan-NA complex, the ΔG_{PB}^{bind} was determined as average on the production phase of MD simulation (between 130 to 195 ns) with a sampling frequency of 200 ps and a sample size of 325 poses covering a range of 65 ns. The standard error of the mean (SEM) of ΔG_{PB}^{bind} was calculated as $\sigma/(N)^{1/2}$, where σ represents the estimated standard deviation, and N represents the number of samples. The contributions to ΔG_{PB}^{bind} of individual binding counterparts (saccharide and amino acid residues) ($\Delta G_{PB}^{bind}(i)$) were calculated using the MMPBSA.py application (Miller et al., 2012) included in Ambertools 14.0 package. The chosen set of residues included all glycan residues (Neu5Ac, Gal, GlcNAc, Gal2, and Glc), and all amino acids of NA characterized by an energy cut-off $|\Delta G_{PB}^{bind}(i)| \geq 0.3$ Kcal mol^{-1}. Negative and positive values of $\Delta G_{PB}^{bind}(i)$ reflect, respectively, favorable and unfavorable contributions of the residue to the binding energy.

2.1.7 Distribution of the water molecules in proximity of the Neu5Ac-Gal glycosidic bond

The MD simulation in explicit solvent reproduces the distribution of water molecules at the interface between the complex (or the macromolecular surface) and the surrounding water (Raffaini et al., 2006; Raffaini and Ganazzoli, 2006). Here, we focused on the glycosidic oxygen of the Neu5Ac-Gal linkage, which is the target of hydrolysis catalyzed by NA (Taylor and von Itzstein, 1994; von Itzstein, 2007). The profile of the local concentration of the water molecules surrounding the oxygen atom was investigated using the radial pair distribution function *rdf(r)* (Levine et al., 2011) as implemented in VMD 1.9.3:

$$rdf(r) = \lim_{dr \to 0} \left[\frac{V}{4\pi N_{pair}} \frac{p(r)}{r^2\, dr} \right] \quad (1)$$

where the variable *r* represents the distance between the glycosidic oxygen of the Neu5Ac-Gal bond (reference atom) and the oxygen atom of each H_2O molecule, N_{pair} denotes the number of potential atom pairs in the system, and *V* is the volume of the simulation cell. *rdf(r) dr* reflects the concentration of water molecules in a spherical layer of infinitesimal thickness (between *r* and *r* + *dr*) centered on the glycosidic oxygen. *rdf(r)* profiles were calculated for each of the four simulated complexes sampled every 100 ps at the production stage (130 to 195 ns) of the MD simulation.

2.1.8 Hydrogen bonds

Tight H-bonds were defined based on the following criteria: a distance less than 3 Å between the atoms of donor X and acceptor Y (X-H---:Y) and a corresponding angle (X-H---:Y) greater than 150° (Almond et al., 2006). These conditions were checked for all possible H-bonds between the glycan and the protein in four studied complexes at the production stage of MD simulation. The percentage of time H-bonds persisted during the simulation was estimated using Origin 8 software.

2.2 Analyses of NA sequences

2.2.1 NA sequences, phylogenetic analyses, and amino acid prevalence at position 347

The sequences were obtained from the GISAID EpiFlu database (Shu and McCauley, 2017) accessed on June 19, 2023. Host- and lineage-specific nucleotide sequences were downloaded using the EpiFlu search filter “N1 NA, complete length” combined with each of the following filters: (1) “HA subtypes, H1-H4, H6-H16; Avian,” (2) “H5 HA; Avian,” (3) “H5 HA; Human,” (4) “H5 HA; Mammalian,” (5) “H1 HA; Swine,” (6) “Human; isolated before 2009,” and (7) “Human; isolated from 2009 to 2023.” Sequences were aligned using the MAFFT program implemented in the Unipro UGENE 47.0 (Okonechnikov et al., 2012). The datasets were processed using Bio-Edit 7.1.11 (Hall, 1999) and Jalview 2.11.2.7 (Waterhouse et al., 2009). Sequences with gaps, ambiguities, incomplete sequences, and sequences of laboratory-derived IVs were excluded; only one sequence from each cluster of identical sequences was retained. Due to the large number of H1N1pdm sequences (exceeding 30,000), a subset of 2,568 representative sequences was selected for the analyses. Maximum likelihood trees were generated using either the FastTree method (Price et al., 2010) implemented in Unipro UGENE or IQ-TREE 2 (Minh et al., 2020) with ModelFinder (Kalyaanamoorthy et al.,

2017) and the ultrafast bootstrap approximation (Hoang et al., 2018). The trees were plotted using FigTree 1.4.4.[1] Based on the trees, the sequences of swine IVs were separated into the classical and the avian-like lineage; the sequences of human IVs were separated into the A/H1N1/1918-like seasonal lineage and the A/H1N1/2009-like H1N1pdm lineage. In some of the sequence datasets, a few sequences failed to cluster with the rest. Some of those sequences were discarded as laboratory artifacts; the remaining atypical sequences represented either swine-virus-like isolates from birds and humans or human-virus-like isolates from pigs and birds; they were combined and analyzed separately. To facilitate analyses of the trees, we annotated the sequences with the single-letter code for the amino acid at position 347 (corresponds to codon 344 of the N1 NA ORF). The number of sequences in each group is listed in Table 3. The group name, virus name, and accession number for each analyzed sequence are shown on the global tree (Supplementary material, Global NA tree.nex and Global NA tree.svg). Prevalence of amino acids at position 347 was determined using the positional numerical summary tool of Bio-Edit.

To annotate titles of NA and/or HA sequences with specific amino acids of both proteins, we downloaded the sequences using the search filter "HxN1; HA + NA; complete length." Curated HA and NA sequences were concatenated with MEGA11 (Tamura et al., 2021). Using Bio-Edit, we selected specific columns in the alignments of translated concatenated sequences and copied them into sequence titles.

2.2.2 Selection pressure analyses

We applied several codon-based phylogenetic tests of natural selection to the global alignment of all N1 NA sequences [for a review see Spielman et al. (2019)]. These tests estimate the ratio of rates of non-synonymous and synonymous substitution (dN/dS) and compare it to the neutral expectation, dN/dS = 1 (FEL and MEME tests) (Kosakovsky Pond and Frost, 2005; Murrell et al., 2012). For analyses investigating within-group selection (e.g., within "Classical swine" group, or "H1N1pdm" group), we partitioned all branches in the combined NA tree using "conjunctive" labeling: each leaf of the tree has a label based on its annotation, and internal nodes (whose labels are not known) are assigned the same label as their descendants, but only when all of the descendants have the same label. The remaining internal nodes receive no label ("Unlabeled" group). This approach is conservative and will not attempt to infer virus group assignments for internal/ancestral nodes unless such assignment is unambiguous. Given G groups in the tree, we inferred G dN/dS ratios using maximum likelihood. For each group, we tested whether its dN/dS was different from 1, using the likelihood ratio test.

We also ran a multi-group MEME analysis, which allowed dN/dS to vary among branches within a group, and can be used to detect episodic selection, i.e., selection affecting only a proportion of branches in the tree (Murrell et al., 2012). For all dN/dS analyses, we used the HyPhy package (v 2.5.54).

3 Results

3.1 Molecular modeling

To study recognition by the catalytic sites of N1 NAs of human-type and avian-type sialoglycans and to dissect the role of amino acid at position 347 in recognition, we modeled binding of the sialopentasaccharides 6S and 3S to the NA of human-adapted pandemic IV A/Brevig Mission/1/1918 (H1N1) and to its "avianized" N347Y mutant. Four complexes, 3S-NA/Y, 6S-NA/Y, 3S-NA/N, and 6S-NA/N, were built by docking, and their geometry was analyzed and refined by MD simulation in explicit solvent (Figure 1). The complexes 6S-NA/N and 3S-NA/Y were used to model interactions between human-type and avian-type NAs, respectively, with corresponding host-specific sialoglycans (referred to as "homologous" interactions throughout the text). The complexes 6S-NA/Y and 3S-NA/N modeled "heterologous" interactions.

3.1.1 Alteration of the conformations of 6S and 3S upon binding to NA

The Ramachandran plots and the most populated conformations of the glycosidic torsion angles ϕ_i/ψ_i for both free and bound sialoglycans sampled during the productive stage of MD simulations are presented in Figure 2, Supplementary Figure S2, and Supplementary Table S1. The ϕ_1/ψ_1 torsions of free 6S demonstrated significant conformational flexibility of the Neu5Ac2–6Gal linkage and populated four states, −62°/−163° (41%), −73°/155° (26%), −66°/110° (20%) and − 173°/170° (13%) (Elli et al., 2021) (see Figure 2A; Supplementary Table S1). In the homologous 6S-NA/N complex, the conformational space of the Neu5Ac2–6Gal linkage was reduced to one of these four states (−67°/122°). In contrast, none of the four states were observed in the heterologous 6S-NA/Y complex, in which case the Neu5Ac2–6Gal linkage adopted a new conformation ($\phi_1/\psi_1 = 52°/-130°$) not previously seen in free Neu5Ac2–6Gal-terminated glycans or in their complexes with IV HA (Xu et al., 2009; Elli et al., 2014; Macchi et al., 2016). Distinctions in the recognition of Neu5Ac2–6Gal linkage by NA/N and NA/Y were also evident at the level of the linkage dihedral angle ω (O6-C6-C5-H5). More significant changes in ω (from *gauche* (−) to *trans*) occurred in the case of 6S binding to avian-type NA/Y compared to the homologous NA/N (Figure 2C).

MD simulations with the avian-type glycan 3S revealed analogous pattern. The Neu5Ac2–3Gal linkage of free 3S populated three states, −67°/−3° (74%), −92°/−58° (15%) and − 168°/−23° (11%) (Elli et al., 2021). The second conformation was selected when 3S bound to the homologous NA/Y ($\phi_1/\psi_1 = -79°/-42°$), whereas binding to heterologous NA/N induced a new conformation ($\phi_1/\psi_1 = -55°/-100°$) (Figure 2B; Supplementary Table S1).

In contrast to significant changes in the conformational space of the Neu5Ac-Gal linkage upon binding, much weaker effects were observed at the level of the Gal-GlcNAc linkage (ϕ_2/ψ_2) (Figure 2), and no effects were detected at the level of more distant saccharide residues (ϕ_3/ψ_3 of GlcNAc-Gal2 linkage and ϕ_4/ψ_4 of Gal2-Glc linkage, Supplementary Figure S2; Supplementary Table S1). This pattern suggests that the recognition of glycans by NA was primarily determined by the structure of the terminal Neu5Ac-Gal moieties.

---

[1] https://tree.bio.ed.ac.uk

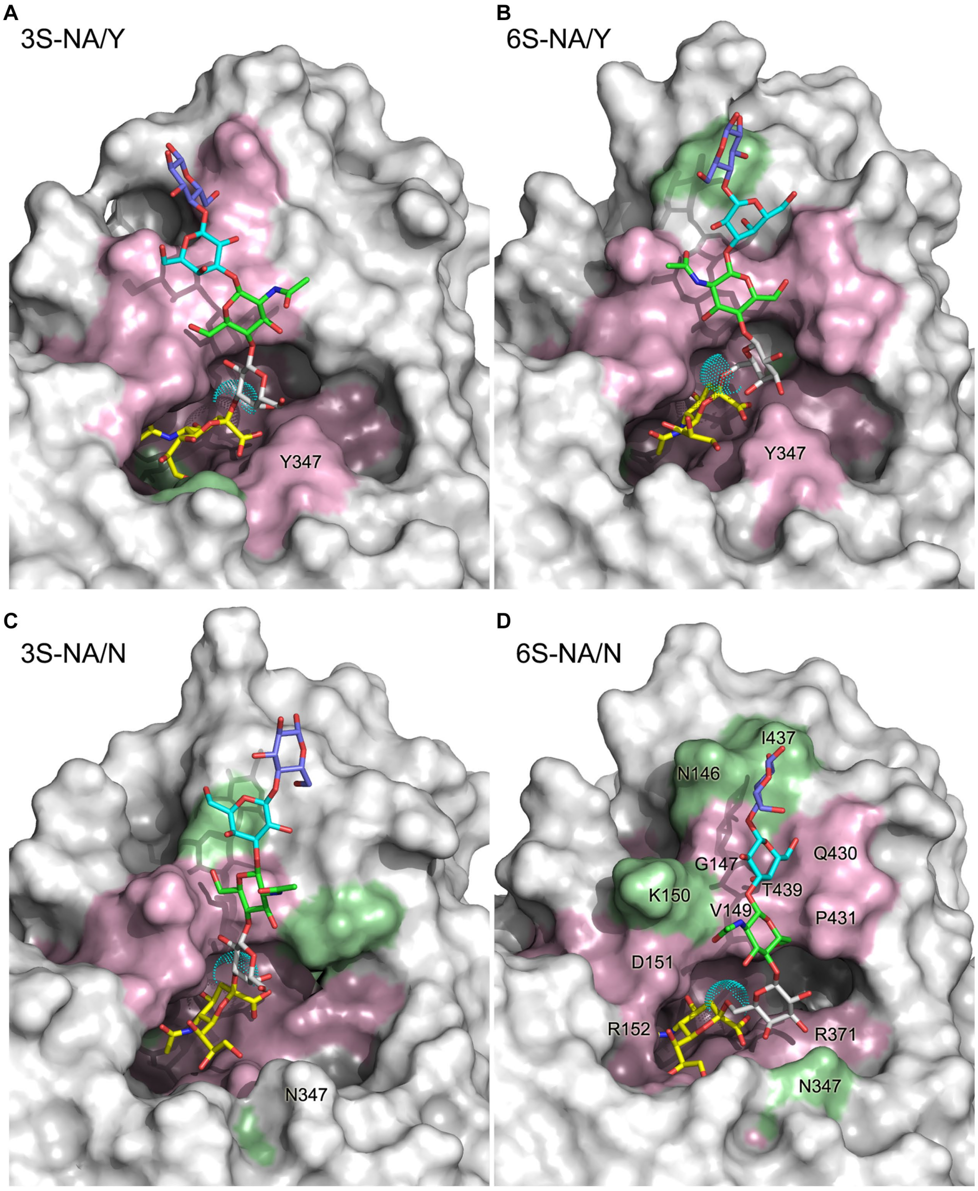


FIGURE 1
3D structure of the complexes 3S-NA/Y **(A)**, 6S-NA/Y **(B)**, 3S-NA/N **(C)**, and 6S-NA/N **(D)** predicted by MD simulation. The sialoglycans are shown as stick models with carbon atoms colored in yellow (Neu5Ac), white (Gal), green (GlcNAc), cyan (Gal2) and slate blue (Glc), and oxygen and nitrogen atoms colored in red and blue, respectively. Cyan dots show the van der Waals surface of the oxygen atom of the Neu5Ac-Gal linkage. The protein is presented as a molecular surface with atoms colored according to their contact distance to the sialoglycan (pink, < 4 Å; green, between 4 Å and 5 Å; white, >5 Å). Selected residues are labeled on the panel **(D)**.

These results demonstrated that the binding of sialoglycans to NA significantly reduced their conformational space. The extent of this effect (and the associated energy cost) depended on the interplay between the type of Neu5Ac-Gal linkage and the amino acid residue at position 347; the effect was less pronounced in the case of homologous glycan-NA interactions.

3.1.2 Atomic contacts in the NA complexes with sialoglycans

The contact distances between sialoglycans and the protein in the complexes are color coded in Figure 1. Figure 3 illustrates the interactions between the terminal Neu5Ac-Gal moiety and the catalytic site of the NA. Selected contact distances, H-bonds and their persistence during the production phase of the MD simulation are presented in Table 1. In all four complexes, the Neu5Ac residue was located between the conserved amino acid residues R118, E119, D151, R152, R156, W178, R292, E276, R371, and Y406. However, the position of Neu5Ac with respect to the arginine triad (R118, R292, R371), residues R152, R156, and Y406, as well as loops 150 and 430, and contact distances between Neu5Ac and the protein differed between homologous and heterologous complexes.

The homologous complexes 3S-NA/Y and 6S-NA/N showed considerable parallelism in the NA interactions with the terminal

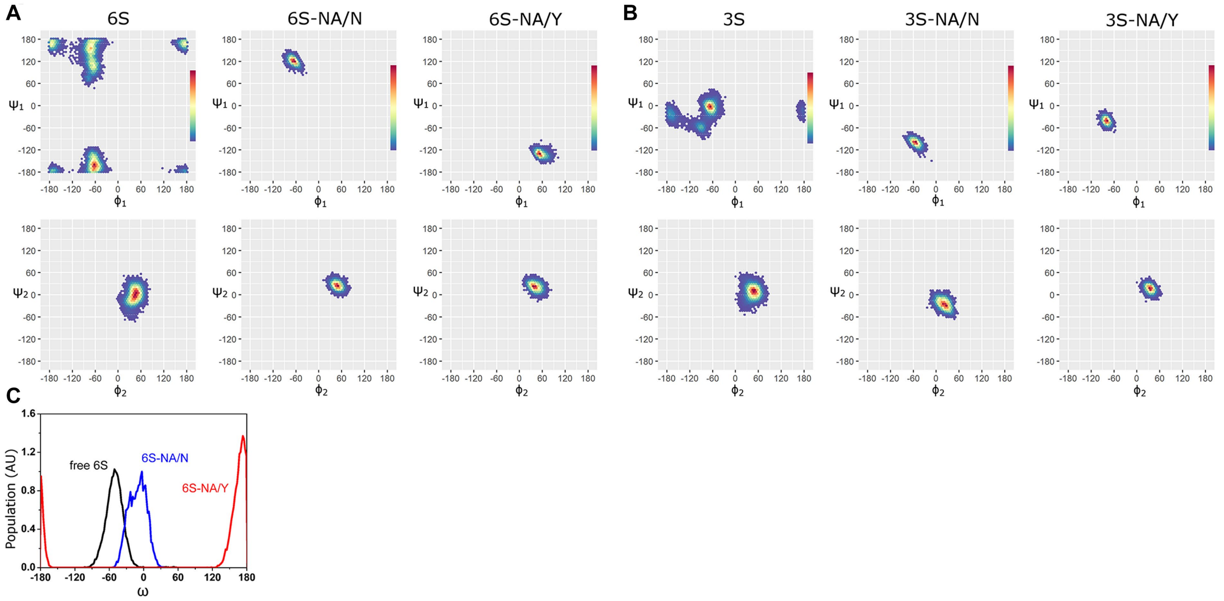


FIGURE 2
Ramachandran plots and color density maps of the glycosidic torsions ϕ_1/ψ_1 and ϕ_2/ψ_2 of the glycans 6S **(A)** and 3S **(B)** in unbound state [(Elli et al., 2021) with permission from Biochemical Journal] and in complex with NA/N and NA/Y. The color gradient from blue to red on each map is proportional to the increase in the population density of the states sampled by MD simulation. **(C)** Population (arbitrary units) of the ω angle of 6S in unbound state (black line) and in the complexes with NA/N (blue line) and NA/Y (red line).

Neu5Ac moiety of the glycan (Figures 3A,D; Table 1). Neu5Ac was co-planar with the arginine triad residues in both complexes. Tight H-bonds were formed by Neu5Ac with R292, R371, E119, E276 and either R152 (3S-NA/Y) or R156 (6S-NA/N). The distances between the hydroxyl group of the key nucleophilic residue Y406 and its target C2 atom of Neu5Ac (4.0 Å, 3S-NA/Y; 3.9 Å, 6S-NA/N) were compatible with the proposed role of Y406 in NA-mediated hydrolysis (von Itzstein, 2007; Vavricka et al., 2013). Comparison of 3S-NA/Y and 6S-NA/N with the published crystal structures of NA complexes with free Neu5Ac revealed a marked similarity in the contact distances of the Neu5Ac residue in these complexes (Table 1). This finding suggests that, in the homologous complexes, the asialic moieties of the bound glycan do not interfere with the optimal position of the terminal Neu5Ac moiety in the NA catalytic site. Importantly, the Gal residues of the glycan in the complexes 3S-NA/Y and 6S-NA/N facilitated binding through the formation of polar contacts and weak transient H-bonds with the side chain of amino acid 347. Thus, during the MD simulation, the 4-hydroxyl of Gal of 3S oscillated within 3.0–3.5 Å from the hydroxyl group of Y347, whereas the 4-hydroxyl of Gal of 6S oscillated within 4.7–5.9 Å from the side chain of the carboxamido group of N347. In addition to interacting with amino acid 347, the asialic residues of 3S and 6S also interacted with V149. Gal made polar contacts with V149 in the 3S-NA/Y complex, whereas GlcNAc was involved in hydrophobic interactions with this amino in the 6S-NA/N complex. Interestingly, in 6S-NA/N, the GlcNAc and Gal2 residues of the glycan approached loop 430 and were engaged in polar and hydrophobic interactions with Q430 and P431, respectively (Figure 1D; Table 1). This finding supports the previous hypothesis about potential role of amino acid 430 in the substrate specificity of human-type NAs (Jongkon and Sangma, 2012).

The heterologous complexes 6S-NA/Y and 3S-NA/N shared a few structural features that differentiated them from both the homologous complexes and the published Neu5Ac-NA co-crystal structures. Thus, 6S bound to NA/Y in a tilted orientation of the Neu5Ac residue with respect to the arginine triad. Compared to the 3S-NA/Y complex, the Neu5Ac moiety of 6S moved away from loop 340 toward loop 150, lost H-bonds with R292 (distance, 5.1 Å), R152 (6.1 Å) and E276 (4.4 Å), and acquired tight H-bonds with R118 (2.8 Å) and R156 (3.2 Å) (Figures 1B, 3B; Table 1). The distance between the hydroxyl group of Y406 and the C2 carbon atom of Neu5Ac has increased to 4.8 Å in the 6S-NA/Y complex, compared to 4.0 Å in the 3S-NA/Y complex. Observed differences in the binding of 3S and 6S to NA/Y were likely caused by a larger footprint of the Neu5Ac2–6Gal moiety, a steric conflict of the 6-linked Gal residue with Y347 (see the pink surface in Figure 1B), and resolution of this conflict by the shift of the sialoglycan toward loop 150.

Similar to the binding of 6S to NA/Y, 3S bound to NA/N in a tilted orientation characterized by the absence of H-bonds between Neu5Ac and R292, R152 and E276 and by the formation of H-bonds with R118, R371 and R156 (Table 1). Due to this orientation, the Neu5Ac-Gal moiety shifted toward loop 150, and the distance between Y406 and the target C2 atom of Neu5Ac increased to 5.6 Å. The tilted orientation of the Neu5Ac-Gal moiety in 3S-NA/N was stabilized by the H-bond and hydrophobic interactions of Gal with V149 (Figure 3C; Table 1). The close contact between the ligand and loop 150 in 3S-NA/N and 6S-NA/Y is illustrated in the Figures 1B,C by the dotted van der Waals spheres of glycosidic oxygen atoms approaching this loop.

In summary, these results suggested that N347 and Y347 bind the homologous sialoglycans 6S and 3S, respectively, with an almost

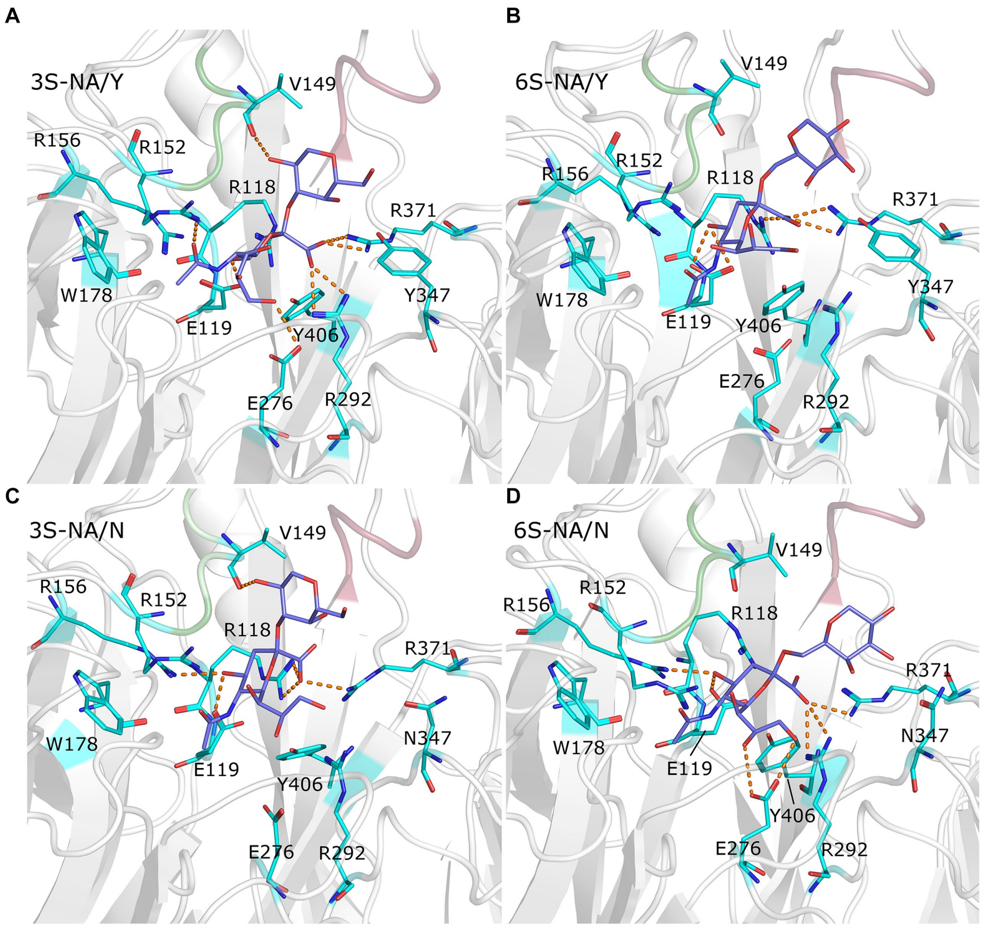


FIGURE 3
Close up view of the complexes 3S-NA/Y **(A)**, 6S-NA/Y **(B)**, 3S-NA/N **(C)**, and 6S-NA/N **(D)**. The white ribbon depicts the protein backbone, with loop 150 and loop 430 colored in green and pink, respectively. Stick models show terminal Neu5Ac-Gal moiety of the ligand (carbon atoms in slate blue) and selected contact residues (carbon atoms in cyan); oxygen and nitrogen atoms are colored in red and blue, respectively. The hydrogen bonds with the highest population are indicated by orange dashed lines.

identical optimal orientation of the sialic acid moiety with respect to the contact residues of the catalytic site. This orientation depends, at least in part, on van der Waals and polar interactions of amino acid 347 with the Gal residue of the glycan.

3.1.3 Poisson-Boltzmann free energy of binding

The Poisson Boltzmann free energy of binding ΔG_{PB}^{bind} (Weis et al., 2006) was calculated from the MD trajectories of glycan-NA complexes and decomposed into the contributions of individual saccharide residues of the glycan (Table 2) and amino acid residues of the protein (Figure 4). It should be noted that the MMPBSA procedure used to calculate ΔG_{PB}^{bind} does not account for the conformational changes (and associated energy costs) that the ligand and receptor undergo upon binding.

Based on the values of total ΔG_{PB}^{bind}, the human-type NA/N bound 6S with a higher avidity in comparison to 3S, whereas avianized NA/Y bound 3S more strongly than 6S. The N347Y substitution markedly reduced NA binding to 6S (ΔG_{PB}^{bind} difference, 19.2 Kcal mol^{-1}) and had weak unfavorable effect on binding to 3S (3.9 Kcal mol^{-1}). This pattern generally agreed with the results of MD simulation performed using sialotrisaccharides and NAs from wild type human and avian IVs (Phanich et al., 2019). Neu5Ac provided the major favorable contribution to the binding energy among the saccharide residues of the ligand, with the highest contribution in the case of 6S-NA/N complex (−16.8 Kcal mol^{-1}). The other residues, including penultimate Gal, made insignificant and frequently unfavorable contributions in all complexes (Table 2).

The amino acid residues with charged side chains made the highest favorable and unfavorable contributions to ΔG_{PB}^{bind} (Figure 4), highlighting the key role of electrostatic forces in the binding. The arginine triad R118, R292 and R371 together with R152, R156, and R224 had the most significant favorable effect on binding. A few amino acids without charged side chains, including S179, Y/N347 and Y406, also favorably contributed to binding avidity. In contrast, residues with the side chain carboxylic groups (E119, E227, E276, E277) destabilized the glycan-NA complexes. Interestingly, loop 150 containing positively charged R152 and R156 had a much higher favorable energy share compared to loop 430 (Q430, P431, K432). In most cases, both favorable and unfavorable effects were more pronounced in the homologous complex than in the heterologous

TABLE 1 Selected contact distances (Å) in NA complexes determined in this study and in published crystal structures 2BAT, 1MWE, and 1 W21.

Contacting atoms (bold)		Complex						
NA	Glycan	3S-NA/Y	6S-NA/Y	3S-NA/N	6S-NA/N	2BAT	1MWE	1 W21
H-bonds and polar contacts								
R118 NC**N**H$_2$	Neu5Ac C**OO**-	5.7	**2.8** (57)	**2.6** (86, 55)	4.8	**3.1**	**2.8**	**2.9**
R292 NC**N**H$_2$	Neu5Ac C**OO**-	**2.6** (26, 53)	5.1	5.8	**3.1** (62, 28)	**3.2**	**3.3**	**3.3**
R371 NC**N**H$_2$	Neu5Ac C**OO**-	**2.8** (21, 56)	**3.0** (67, 43)	**2.9** (86, 62)	**2.7** (94, 6)	**2.8**	**2.7**	**2.9**
E119 C**OO**-	Neu5Ac 4**O**H	**2.6** (67)	**3.0** (48, 15)	**2.6** (87)	**2.6** (92)	**3.3**	**3.3**	**3.1**
R152 NC**N**H$_2$	Neu5Ac NHC**O**Me	**3.1** (49)	6.1	5.6	4.0 (6)	4.0	**2.6**	**2.9**
R156 NC**N**H$_2$	Neu5Ac 4**O**H	3.9	**3.2** (10)	**2.9** (38)	3.8 (37)	4.8	4.7	4.7
E276 C**OO**-	Neu5Ac 8**O**H	4.7	5.5	4.9	**3.4** (95)	**2.6**	**2.7**	3.6
E276 C**OO**-	Neu5Ac 9**O**H	**3.0** (60)	4.4	7.0	**2.6** (89)	**3.1**	**3.4**	**2.7**
Y347 **O**H	Neu5Ac 9**O**H	7.4	4.2 (8)	-	-	-	-	-
N347 C**O**NH$_2$	Neu5Ac 9**O**H	-	-	6.1	7.5	10.8	8.5	-
V149 C**O**	Gal 2**O**H	**2.9** (69)	-	**2.8** (58)	-	-	-	-
Q430 C**O**NH$_2$	Gal2 6**O**H	6.3	6.1	7.6	3.8	-	-	-
K150 **NH$_3$**+	GlcNAc NHC**O**Me	12.4	3.6 (5)	7.8	6.2	-	-	-
Y347 **O**H	Gal 4**O**H	**3.0**	3.7	-	-	-	-	-
Y347 **O**H	Gal 3**O**H	-	**2.8**	-	-	-	-	-
N347 C**O**NH$_2$	Gal 4**O**H	-	-	9.4	5.5	-	-	-
N347 C**O**NH$_2$	Gal 3**O**H	-	-	10.9	5.9	-	-	-
van der Waals contacts								
Y406 **O**H	Neu5Ac **C2**	4.0	4.8	5.6	3.9	**3.2**	**3.2**	**3.1**
W178 **Cε3**	Neu5Ac NHCO**Me**	4.0	6.9	6.4	6.1	3.9	3.7	3.9
V149 CH**Me$_2$**	GlcNAc **C3,C5,C6**	4.4 (C3)	5.1 (C3)	3.9 (C5)	4.4 (C6)	-	-	-
P431 **Cδ**	GlcNAc **C6**	7.3	5.3	10.5	4.0	-	-	-

The closest contacts (<3.5 Å) are shown in bold. Values in brackets depict percent of H-bond persistence (if higher than 5%). 2BAT, 1MWE, and 1 W21, complexes of free Neu5Ac with N2, N9, and N6 NAs, respectively (Varghese et al., 1992, 1997; Rudiño-Piñera et al., 2006).

TABLE 2 Total ΔG_{PB}^{bind} (±SEM) of the complexes and contributions of saccharide residues (Kcal mol^{-1}).

ΔG_{PB}^{bind} component	Complex			
	3S-NA/Y	6S-NA/Y	3S-NA/N	6S-NA/N
Total	−26.8 (±0.3)	−18.7 (±0.3)	−30.7 (±0.4)	−37.9 (±0.4)
Neu5Ac	−11.5	−10.5	−9.5	−16.8
Gal	1.1	0.9	−0.5	1.7
GlcNAc	0.3	1.5	0.0	2.1
Gal2	−0.4	0.0	−0.4	2.1
Glc	0.9	−0.1	−0.3	0.2

complex of the same sialoglycan (as shown by the data in the same panels of Figure 4). However, R118 deviated from this pattern, suggesting a potential unique function for this residue in recognition. Of note, V149 contributed favorably to the binding in both 3S complexes as a result of its hydrophobic interactions with Gal (Table 1).

In summary, estimation of the free energy of binding reveals that avian-type NA/Y binds human-type 6S significantly weaker than the homologous 3S and that substitution Y347N markedly improves binding of 6S and slightly increases binding of 3S. Decomposition data show that this pattern does not correlate with the direct contribution of amino acid 347 to ΔG_{PB}^{bind} but primarily depends on the effect of this amino acid on favorable electrostatic interactions of 3S and 6S with the arginine triad, E119, E276, and other conserved residues.

3.1.4 Distortion of the Neu5Ac moiety in the complexes with NA

The initial stages of the proposed mechanism of NA-mediated hydrolysis involve the binding of sialoglycan with its Neu5Ac moiety in the chair $^2C_5$ conformation (with the carboxylate in the axial position),

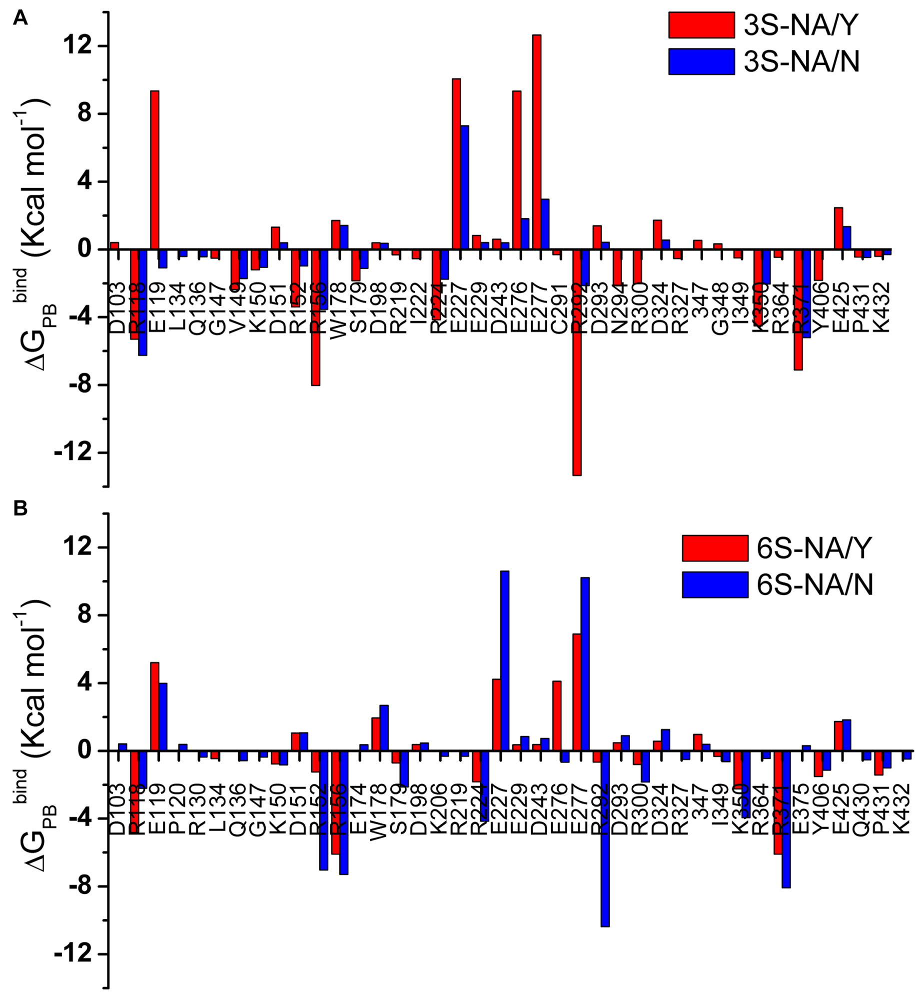


FIGURE 4
Per amino acid residue decomposition of ΔG_{PB}^{bind} for the complexes of 3S **(A)** and 6S **(B)** with NA/Y (red) and NA/N (blue). Only residues with absolute values of ΔG_{PB}^{bind} greater than 0.3 Kcal mol^{-1} are shown.

followed by the distortion of the Neu5Ac ring to a pseudo-boat conformation (with the carboxylate in the pseudo equatorial position). The distortion is mediated by the ionic, hydrogen bond, and steric interactions of Neu5Ac with active site residues; it induces formation of a strained oxocarbonium ion, ultimately resulting in the cleavage of the glycosidic bond (Janakiraman et al., 1994; Taylor and von Itzstein, 1994). Taking this mechanism into account, we decided to examine evolution of the Neu5Ac ring conformation during interaction of 3S and 6S with the NAs (Figure 5). MD simulation of the 3S-NA/Y complex revealed substantial alteration of the internal dihedral angle C1-C2-C3-C4 of Neu5Ac (mean value 93°) with respect to its conformation in the free ligand (mean value 72°) indicative of ring distortion toward pseudo-boat. Minimal distortion (if any) was observed in the case of the 3S-NA/N complex (Figures 5A,C). In the case of 6S, there was a slightly greater Neu5Ac distortion in the 6S-NA/N complex compared to 6S-NA/Y complex (Figures 5B,D), although the effect was smaller than that observed in 3S-NA/Y.

3.1.5 Solvation of the glycosidic oxygen of Neu5Ac-Gal linkage

According to the proposed solvent-assisted mechanism of hydrolysis catalyzed by NAs, water must be present in the catalytic pocket. In particular, the water molecule located near D151 and R152 may participate in the NA-mediated proton transfer to the glycosidic oxygen atom of the substrate (Chong et al., 1992; Taylor and von Itzstein, 1994). Thus, the efficiency of hydrolysis may be affected by the concentration of water molecules surrounding the Neu5Ac-Gal linkage. As the MD simulation in explicit solvent allows the analysis of water molecules at the interface between the complex and the solvent, we calculated the distribution profile of water around the glycosidic oxygen in both free 3S and 6S, as well as their complexes with NA/Y and NA/N as described in the Methods section (Figure 6). In the case of free glycans, local maxima of water concentration were observed at 3.2, 5.5, and 7.4 Å (3S) and 2.8, 5.3, and 7.5 Å (6S). These maxima represented the first, second and third hydration shells, respectively, surrounding the target glycosidic oxygen. The concentrations of water molecules in the second and the third shells decreased in all complexes as compared to free 3S and 6S. Moreover, the first three peaks were shifted in the 3S-NA/N complex toward longer distances indicating a significant depletion of water in this complex in comparison to the homologous 3S-NA/Y complex (Figure 6A). Similarly, a greater depletion of water was observed in the complex 6S-NA/Y compared to its homologous counterpart 6S-NA/N (Figure 6B). These effects correlate with the observed shift of the

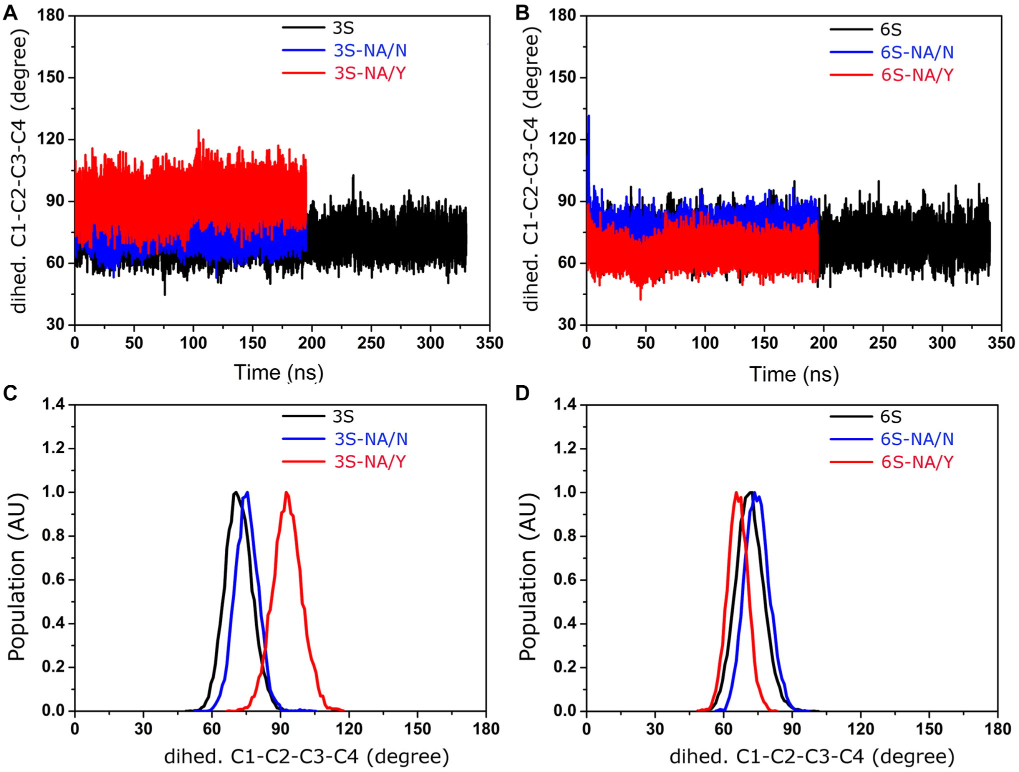


FIGURE 5
Conformation of the dihedral angle C1-C2-C3-C4 of the Neu5Ac residue of 3S and 6S in solution (black) and in complexes with NA/Y (red) and NA/N (blue). **(A,B)** MD simulation trajectories. **(C,D)** Population (arbitrary units).

Neu5Ac-Gal moieties in heterologous complexes 6S-NA/Y and 3S-NA/N toward loop 150 and the accompanying shielding of the glycosidic oxygen from the solvent (see models in the Figure 1).

3.2 Analysis of NA sequences

3.2.1 Prevalence of amino acid 347 in avian and mammalian IV lineages

To investigate the prevalence of amino acid 347 in N1 NAs in different host species, we analyzed sequences retrieved from the GISAID EpiFlu database (Figure 7). In agreement with previous reports (Xu et al., 2012; Hufnagel et al., 2023; Yang et al., 2023), the N1 NAs were represented by the following major groups and lineages. No HxN1 sequences with HA subtypes H13-H16 were found in the database. The NAs of avian IVs with HA subtypes H1-H4 and H6-H12 were located near the root of the global phylogenetic tree and shared a common ancestor with two H1N1 mammalian lineages. Both these lineages, the "seasonal" human IVs and the "classical" swine IVs, were derived from the 1918 influenza pandemic. The second circulating swine IV lineage, commonly called "avian-like swine" (ALS) lineage, originated from an H1N1 avian ancestor; the first ALS viruses were isolated in Belgium and Germany in 1979. The 2009 influenza pandemic was initiated by a reassortant swine virus (H1N1pdm) that contained the NA gene from the ALS lineage. The novel H1N1pdm lineage replaced the previous seasonal human IVs. The H1N1pdm viruses were repeatedly reintroduced from humans to pigs, and this phenomenon significantly influenced the evolution of the ALS lineage (Hufnagel et al., 2023). Therefore, we divided the ALS sequences into two groups as shown in Figure 7. The sequences of highly pathogenic and panzootic H5N1 avian IVs comprised the last major N1 NA group. Due to the importance of H5N1 viruses for animal and human health, their sequences were significantly overrepresented compared to other avian sequences. We divided the NAs of all H5N1 IVs into four groups containing sequences from human isolates (H5hum), sequences from other mammalian isolates (H5mam), and two groups of avian sequences (Figure 7). The H5av-1 group included NAs of the A/goose/Guangdong/1996-like viruses from 1997–2023 (Gs/Gd-like lineage). The H5av-2 group included NAs of clade 2.3.4.4b H5N1 IVs that emerged in 2020, spread globally, and have been causing large outbreaks of infection in wild birds, poultry and mammals in 2020–2023 (Kandeil et al., 2023; Tian et al., 2023). Both H5av groups also included a small number of non-zoonotic H5N1 IVs from wild birds and poultry.

The number of NA sequences in each group and the prevalence of amino acid 347 are shown in Table 3. Our initial analysis of avian NAs revealed an unusually high frequency of substitutions at position 347 in H6N1 samples. Therefore, we decided to analyze H6N1 viruses separately. Y347 was perfectly conserved in the NAs of avian viruses with HA subtypes H1-H4 and H7-H12. The H6N1 group contained 11.6% of NA variants with the Y347F substitution. The H5av-1 group contained 1.33% of mutant NA sequences, ten times more than the H5av-2 group. It remains unclear whether this difference reflects a shorter evolutionary time of the H5av-2 lineage, unique virus properties, or a combination both.

Among the human and porcine IV NAs, only one H1N1pdm sequence (0.09%) and 14 ALS-1 sequences (1.1%) contained the avian-type 347Y. NAs from classical swine, ALS, and H1N1pdm lineages contained almost exclusively N347, whereas NAs from seasonal human IVs contained either N347 or D347.

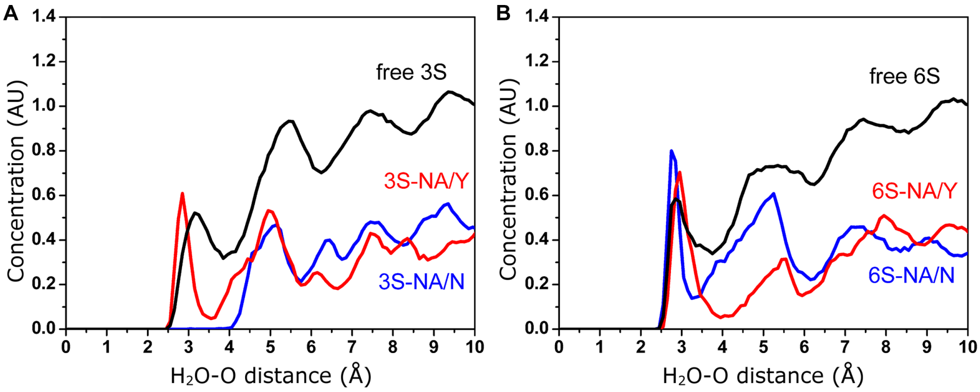


FIGURE 6
Concentration (arbitrary units) of water molecules in the vicinity of the glycosidic oxygen of the Neu5Ac-Gal linkage determined by MD simulation. **(A)** Free and bound 3S. **(B)** Free and bound 6S.

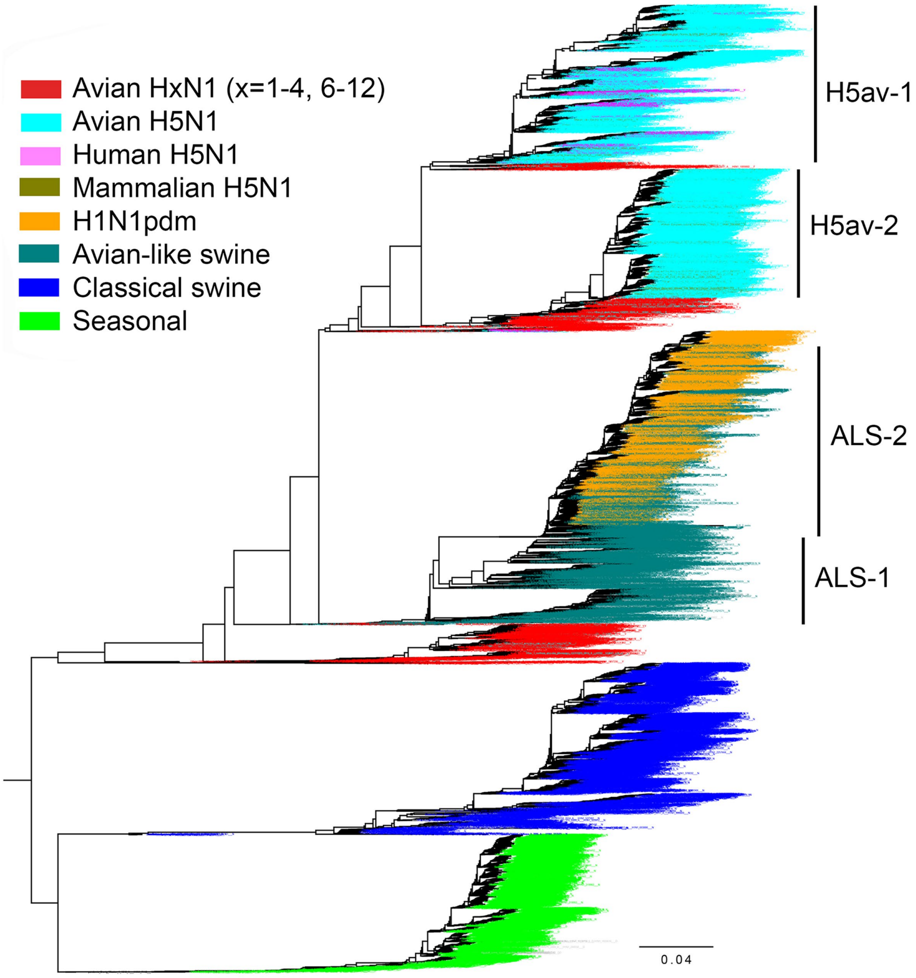


FIGURE 7
Phylogenetic relationship of 18,097 N1 NA gene sequences from the GISAID EpiFlu database used in this study. For the H1N1pdm lineage, only 2,568 representative sequences were included in the tree. The colors of the taxa names correspond to the virus group as indicated in the legend. Group separation of avian H5N1 IVs and avian-like swine IVs is illustrated. The tree was generated with FastTree. The fully annotated tree in nexus and svg formats can be found in the Supplementary information.

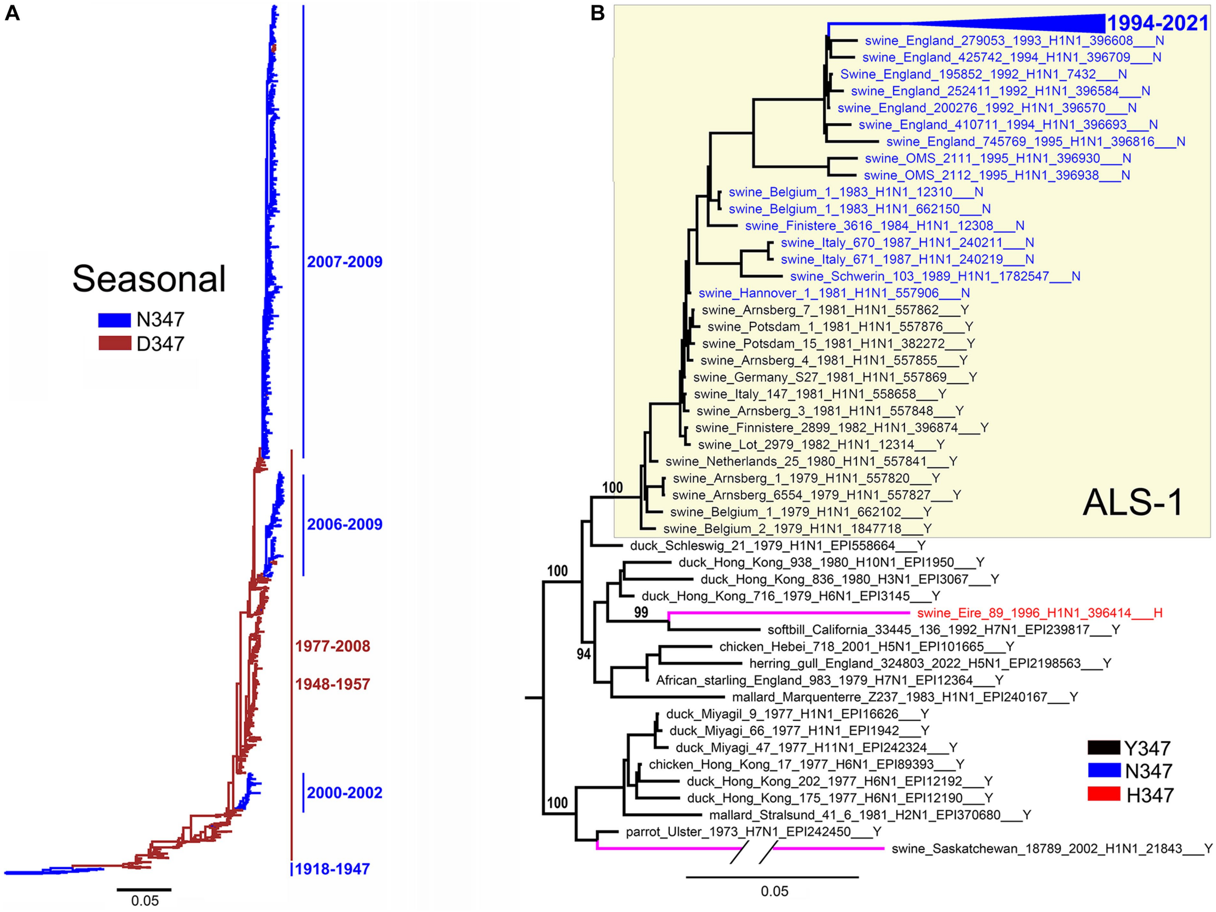


FIGURE 8
Substitutions at position 347 color-coded on NA phylogenetic trees as indicated in the panel legends. **(A)** Seasonal human lineage. Taxa names are not shown. Bars and numbers on the right present circulation periods of NA clades with N347 (blue) and D347 (brown). **(B)** The tree displays ALS-1 lineage (yellow box), phylogenetically close avian IVs, and two independent avian-like porcine isolates indicated by magenta branches. The branch containing sequences of IVs isolated after 1994 is collapsed for clarity (blue triangle). Numbers on specific branches represent bootstrap support values.

The NAs of 101 atypical sporadic swine-virus-like isolates from humans and birds and human-virus-like isolates from pigs as a rule contained the same residue 347 as their closely related original host species (see sequences labeled "Atypical" in the tree in Supplementary material).

3.2.2 Evolution of position 347 in seasonal and ALS-1 lineages

For the NA groups containing more than 1% of position-347 variants, we studied the location of these variants on the phylogenetic tree (Figures 8–10 and Global tree.nex/Global tree.svg in Supplementary material).

The H1N1/1918-derived seasonal human IVs circulated in 1918–1957 and 1977–2009 (Krammer et al., 2018). The NAs of the 1918 virus and its successors isolated between 1933 and 1947 contained N347 (Figure 8A). The N347D substitution occurred around 1947, and viruses with this substitution circulated in 1948–1957 and 1977–2008. During this time, the IVs with the D347N reversion emerged on three separate occasions and circulated together with the D347 variants in 2000–2002 and 2006–2009. These observations suggest that both N347 and D347 support the replication of seasonal viruses in humans and that occasional swapping of N and D occurs, conferring epidemiological advantage to the novel variant and leading to its expansion.

All 14 NAs of the ALS-1 lineage containing the avian-type Y347 were located at the root of the avian-like lineage (Figure 8B). The Y347N substitution occurred in 1981 and was strictly maintained thereafter. The only other substitution in the ALS-1 NAs, N347D (Table 3), was present in a cluster of two viruses isolated in 2013 (see Global tree in Supplementary material). In addition to the stable ALS lineage, two atypical H1N1 IVs, A/swine/Eire/1996 and A/swine/Saskatchewan/2002, represented independent introductions of avian IVs into swine (Figure 8B; Karasin et al., 2004; Lycett et al., 2012). Interestingly, these IVs contained canonical markers of HA adaptation to Neu5Ac2-6-Gal-terminated receptors, the E190D substitution (both viruses) and the Q226L substitution (A/swine/Eire/1996). The latter virus also had the Y347H substitution in NA.

3.2.3 Substitutions in the NA of H6N1 and H5N1 viruses are associated with substitutions in HA RBS

All H6N1 viruses with the Y347F NA substitution belonged to the specific clade of poultry IVs from Taiwan (Figure 9A). A case of human

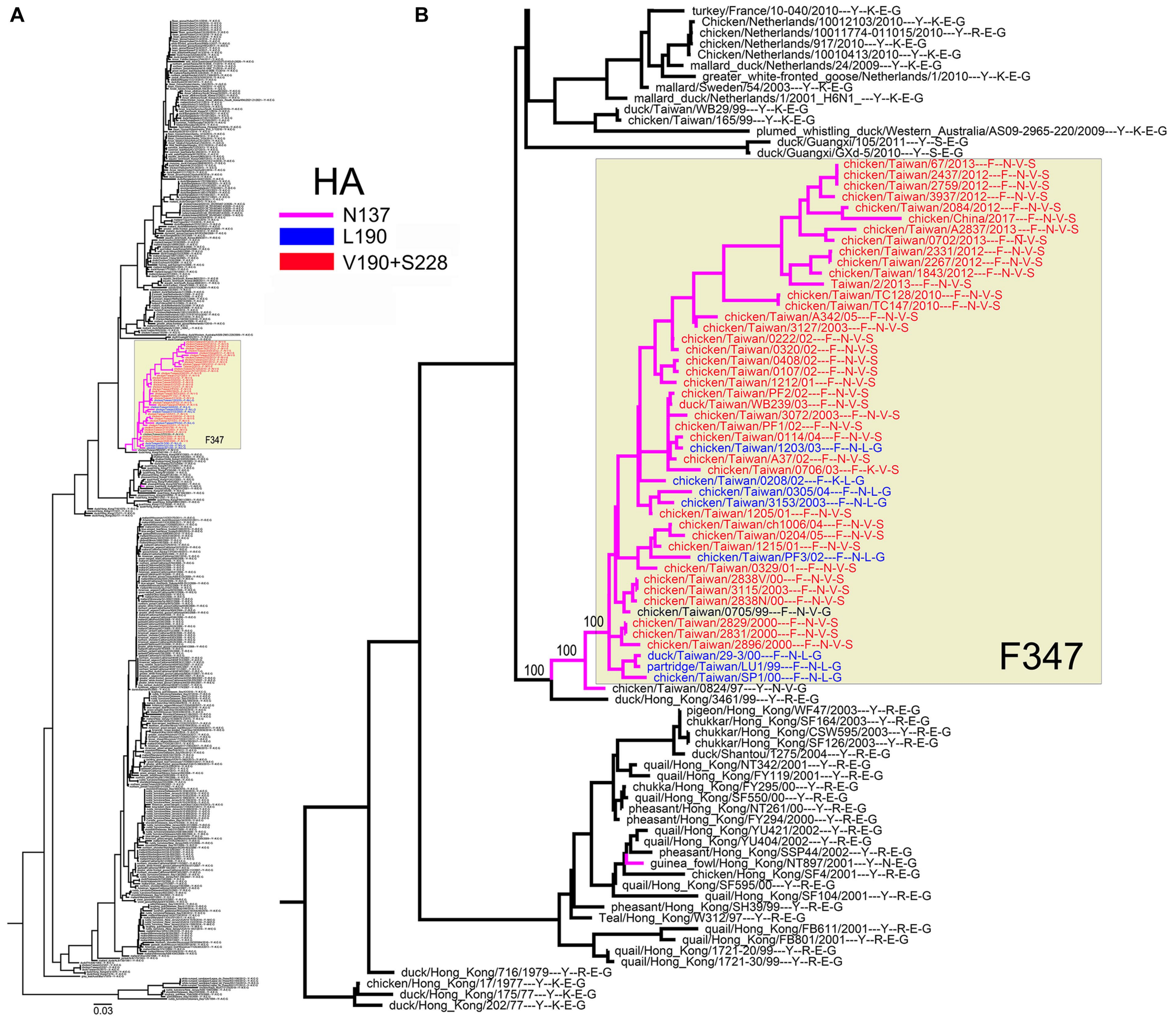


FIGURE 9

Association of the Y347F substitution in H6N1 NA with amino acids at HA positions 137, 190, and 228. **(A)** Phylogenetic relationships of NA. Colors denote the NA clade with F347 (yellow box) and the HAs with N137 (purple branches), 190 L (blue taxa names), and 190 V + 228S (red taxa names). **(B)** Enlarged partial view with specific amino acid residues listed after the strain name. The first character defines the NA residue 347, the next 3 characters define the HA residues 137, 190, and 228. Numbers on specific branches represent bootstrap support values.

infection with the virus from this clade in 2013 prompted research into the structure and receptor-binding properties of HA [for a review, see Everest et al. (2020)]. One of the reports predicted that the substitutions E190V and G228S in the HA RBS, in combination with N137, could allow the H6N1 viruses to bind to human-type receptors (Ni et al., 2015). In this view, we generated a phylogenetic tree of the H6N1 NA annotated with amino acids 137, 190, and 228 of the corresponding HAs (Figure 9B). According to the phylogeny, the Y347F substitution emerged and was fixed in parallel with unique substitutions in HA, R137N and either E190L or E190V + G228S.

In the H5av-1 group, the non-Y NA variants were scattered across different clades of Gs/Gd-like IVs (Figure 10A). Occasional clusters of the Y347H and Y347D mutants containing from 2 to 10 sequences were observed, suggesting limited expansion of these mutants. Variants having Q, F, and S were represented by a single sequence each. Two H5hum mutants (Y347H, Y347G) and one H5mam mutant (Y347H) clustered with avian-derived NAs that had Y347 (see the trees in Supplementary material). We next analyzed the phylogeny of H5 HA sequences that included annotations of NA residue 347 and groups of HA residues, that form the RBS. As an example, the part of the phylogeny shown in Figure 10B demonstrates that the cluster of viruses with D347 in the NA differs from the close Y347-containing IVs by 3 amino acid substitutions in HA, namely S137A, D187N, and K193D. Using this analysis, we found association between NA substitutions and HA substitutions in roughly half of the H5N1 virus clusters and single sequences examined (Figure 10A; Table 4). Remarkably, substitutions at most of these positions have been previously shown to affect antigenic and receptor-binding properties of H5 HA (Paulson and de Vries, 2013; Koel et al., 2014; Suttie et al., 2019).

3.2.4 Selection pressure analyses

These analyses were performed to determine, whether site 347 evolved under natural selection pressure and whether pressures varied depending on the virus host species and/or phylogenetic lineage.

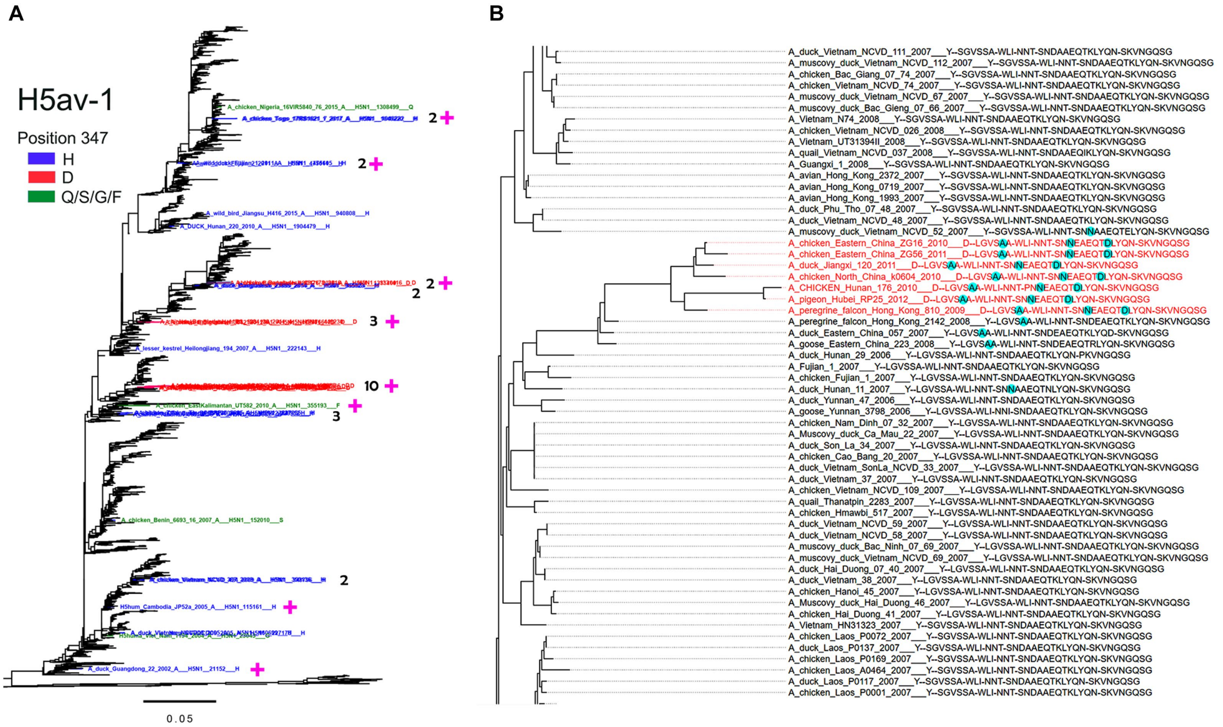


FIGURE 10
Association of substitutions at position 347 in H5av-1 NA with substitutions in HA. **(A)** H5av-1 NA tree rooted to A/chicken/Scotland/1959. The names of NAs with Y347 are not displayed. Taxa with the NA substitution 347 are colored as shown in the legend. For clusters of 2, 3, and 10 mutant sequences, the cluster size is indicated. Magenta plus signs mark sequences with associated substitutions in the HA. **(B)** A section of the H5 HA tree with the cluster of Y347D mutants colored in red. The taxa names, such as A_duck_Jiangxi_120_2011___D--LGVSAA-WLI-NNT-SNNEAEQTDLYQN-SKVNGQSG, include the virus name followed by NA residue 347 and groups of HA residues 133–138, 153–155, 158–160, 185–197, 221–228 separated by minus sign. Cyan highlights amino acid residues that distinguish corresponding HAs from others in the tree section. The full tree in nexus and svg formats can be found in the Supplementary material.

TABLE 3 Prevalence of amino acids at position 347 of N1 NA.

Virus group	SeqN	N	D	Y	F	H	K	Q	S	T	G	%
Avian HxN1 (x = 1–4, 7–12)	918			918								0.00
Avian H6N1	464			410	54							**12.6**
H5av-1	2,627		15	2,592	1	17		1	1			**1.33**
H5av-2	2,315			2,312		3						0.13
H5hum	334			332		1					1	0.60
H5mam	186			185		1						0.54
ALS-1	1,588	1,572	2	14								**1.07**
ALS-2	1,269	1,262	1			1	5					0.55
Classical swine	3,174	3,167	7									0.22
Seasonal	2,551	1,728	823									**32.3**
H1N1pdm, all	32,752	32,723	12	1		2	11		2	1		0.089
H1N1pdm, representative	2,568	2,566	1				1					0.078

SeqN, number of sequences analyzed; %, percentage of amino acids other than Y in avian IVs and other than N in human and swine IVs (percentages greater than 1 are shown in bold).

Using the combined alignment of all sequences ($N = 17{,}856$) we found that site 347 is evolving under overall (pervasive) purifying selection (FEL $dN/dS = 0.20$, value of p for non-neutral evolution $<10^{-20}$). However, a more sensitive method (MEME) identified that this site also experienced episodic diversifying selection (MEME $dN/dS = 8.7$, proportion assigned to $dN/dS > 1 = 10\%$, value of p for episodic positive selection = 0.025). Next, using the combined tree, we estimated dN/dS values for subsets of branches partitioned by

TABLE 4 Substitutions in H5 HA associated with NA substitutions at position 347.

Virus	347	*N*	Substitutions
A/duck/Guangdong/22/2002	H	1	S133L
A/Cambodia/JP52a/2005	H	1	T192I
A/peregrine falcon/Hong Kong/810/2009	D	6	S137A, D187N, K193D*
A/chicken/East Kalimantan/UT581/2010	F	2	L154I, S159N
A/wild duck/Fujian/1/2011	H	2	Q196K + R227S
A/chicken/Bangladesh/3012/2011[#]	D	1	K193Q
A/chicken/Bangladesh/11RS-1984-30/2011[#]	D	1	S185P, D187N, K193Q
A/chicken/Bangladesh/08C178/2016	D	2	D158N (G+), D187N, K193T
A/chicken/Togo/17RS1021-7/2017	H	2	S221P

N, number of HA sequences in the cluster; G+, substitution creates glycosylation sequon; *, D193 was not found in any other H5 HA sequence analyzed; [#], NAs belong to the same cluster of 3 NAs in Figure 10.

TABLE 5 Evolutionary analysis of site 347.

Virus group	BN	dN/dS (CI)	*p*	Inferred substitutions*
Avian HxN1 (x = 1–4, 7–12)	1,211	0.00 (0.00–0.29)	4×10^{-6}	Y(5)
Avian H6N1	643	0.14 (0.00–0.89)	0.03	Y(2), Y-*F*(1)
H5av-1	3,685	0.72 (0.36–1.27)	0.66	Y(8), Y-H(10), Y-D(3), Y-F(1), Y-Q(1), Y-S(1)
H5av-2	2,993	0.21 (0.02–0.86)	0.03	Y(1), Y-H(2)
H5hum + H5mam	618	1.25 (0.24–3.62)	0.68	Y-G(1), Y-H(2)
Classical swine	4,866	0.11 (0.03–0.29)	9×10^{-10}	N(8), N-D(5)
ALS-1	2,603	0.02 (0.00–0.10)	4×10^{-6}	N(16), N-D(1)
ALS-2	1,787	0.23 (0.06–0.61)	7×10^{-4}	N (5), N-D(1), N-K(4)
Seasonal	3,383	0.42 (0.14–0.93)	0.03	N(9), D(2), D-N(5), N-D(3)
H1N1pdm, representative	3,295	0.09 (0.01–0.36)	2×10^{-6}	N(5), N-D(1), N-K(1)
Unlabeled	1,570	0.24 (0.05–0.70)	0.035	Y(7), N(1), D-N(1), N-D(1), N-Y(1), Y-N(1)

dN/dS is estimated using FEL, and profile likelihood confidence intervals (CI, 99%) are shown. The value of p for non-neutral evolution is obtained under a likelihood ratio test, where only the dN/dS ratio for this group is constrained to be 1 in the combined tree with all sequences; the Holm-Bonferroni correction for multiple testing is applied. Substitutions are inferred based on the joint maximum likelihood ancestral codon reconstruction. BN, number of branches tested; P, value of p for $dN/dS \neq 1$; *, X(n), n synonymous substitutions for amino acid X; X-Z(n), n non-synonymous substitutions from amino acid X to amino acid Z.

groups as described above (Figure 7; Table 3) (Note, that we placed H5N1 sequences isolated from mammals or humans in one group to improve statistical power, given the relatively small sizes of each group). We found that 8 out of 10 studied groups of branches evolved subject to purifying selection ($p < 0.05$, LRT with Holm-Bonferroni correction for multiple testing) (Table 5). Overall, unlabeled branches were also subject to purifying selection. As only three substitutions have been accumulated on the combined H5hum + H5mam lineages, there was insufficient statistical power to discern selection. H5av-1 branches encompassed considerably more variation, with 8 synonymous and 16 non-synonymous substitutions, consistent with neutrality (Table 5). However, there was a significant difference ($p = 0.05$, LR test) between the evolutionary regimes along internal branches (dN/dS = 1.3), which, by necessity, represent some host-to-host transmissions, and terminal branches (dN/dS = 0.4), which often reflect within-host evolution. Thus, while we cannot reject the hypothesis of neutrality on all H5av-1 branches, an observation that dN/dS is significantly higher along internal H5av-1 branches, enriched for evolution between hosts (and host species), is consistent with a functional role of this site. We next ran MEME, which is capable of detecting episodic diversifying selection (EDS) affecting only a proportion of branches. Although we detected EDS for the whole NA alignment, no individual group provided statistical support for EDS, likely due to loss of power and relatively few substitutions compared to those accumulated along the entire tree.

We also interrogated the entire NA alignment for evidence of sites that may be co-evolving with site 347 using a Bayesian Graphical Model phylogenetic method (Poon et al., 2007). No evidence of interactions among site 347 and other sites in NA was found.

In summary, negative selection within most groups of viruses indicates that the amino acids at position 347 specifically adapted to particular hosts are maintained, whereas episodic positive selection across the entire tree is consistent with the hypothesis that adaptation occurs during or following some host switches.

4 Discussion

The N1 NAs of avian IVs hydrolyze the Neu5Ac2–6Gal linkage much less efficiently than the Neu5Ac2–3Gal linkage. In contrast, N1

NAs of swine and human IVs show less restricted linkage specificity due to increased activity against the 2–6 linkage and diminished activity against the 2–3 linkage (Mochalova et al., 2007; Gerlach et al., 2012). Previous molecular dynamics simulation studies on N1 NAs from phylogenetically distant avian and human IV isolates suggested that amino acid 347 contributed to the specificity of the enzyme for the Neu5Ac-Gal linkage type (Raab and Tvaroska, 2011; Jongkon and Sangma, 2012; Phanich et al., 2019). Here, we explored this hypothesis by modeling the interaction between sialoglycan substrates and two N1 NAs that differed only by amino acid 347.

Comparison of the complexes of avian-type NA/Y with 3S and 6S revealed the following differences. First, 3S bound to NA/Y in one of its solution-populated conformations, while 6S was forced to adopt a novel conformation that was not populated in its unbound state (Figure 2). Second, the Neu5Ac moiety of the bound 3S was located in the center of the NA catalytic pocket, with optimal contact distances of Neu5Ac to the arginine triad and E276 [residues known to make the major contributions to the binding energy (Taylor and von Itzstein, 1994)], and to the key catalytic residues, such as Y406. In contrast, the Neu5Ac moiety of 6S was tilted in the complex, resulting in a network of contacts that differed from those in both 3S-NA/Y and crystal structures of NA complexes with free Neu5Ac (Table 1; Figures 3A,B). Third, the estimated binding free energy was less favorable in the case of 6S (Table 2). Fourth, catalysis-promoting distortion of the Neu5Ac ring from chair to pseudo-boat was observed in the 3S-NA/Y complex, with little, if any, distortion evident in the 6S-NA/Y complex (Figure 5). Finally, the solvation of the glycosidic bond was less efficient in the 6S-NA/Y complex (Figure 6). All of these differences are expected to decrease the binding avidity and hydrolysis efficiency of 6S in comparison to 3S. Remarkably, the Y347N substitution alleviated most of the mentioned adverse effects by allowing the binding of 6S to NA/N in a solution-populated conformation, improving the orientation of the Neu5Ac moiety and the solvation of the glycosidic linkage, and considerably enhancing the free energy of binding. At the same time, replacement of Y with N negatively impacted binding characteristics of 3S. These findings advance and substantiate previous notions about the significance of amino acid 347 for the NA recognition of the Neu5Ac-Gal linkage. However, substrate binding represents only the first step of the NA-mediated catalysis (Taylor and von Itzstein, 1994). Although catalytic activity correlates with binding affinity (Mochalova et al., 2007; Garcia et al., 2013), other steps, such as donation of the proton from the solvent, formation of the endocyclic sialoside cation intermediate, and release of the asialic part of the substrate may also depend on the type of the Neu5Ac-Gal linkage. Therefore, further analyses of the catalytic activity of the 347 mutants are required to refine the predictions made by modeling.

Our MD simulation data suggest the following molecular mechanisms for the effect of residue 347 on substrate specificity of the NA. We found that optimal binding of 3S to NA/Y depends on favorable polar interactions between the 4-OH group of Gal and the OH group of Y347. In this conformation of bound 3S, the 2-hydroxyl of Gal forms an H-bond with the main chain carbonyl group of V149, while GlcNAc makes a van der Waals contact with the side chain of V149 (Figures 1A, 3A; Table 1). In the case of 6S, the Neu5Ac2–6Gal moiety approaches the side chain of Y347 with the apolar side of the Gal residue. To avoid a steric conflict and unfavorable polar-apolar interactions between Gal and the OH group of Y347, the glycan is shifted toward loop 150; this shift negatively affects interactions with key catalytic residues and reduces binding avidity. In contrast to NA/Y, the less bulky side chain of N347 in the NA/N variant allows accommodation of the Neu5Ac2–6Gal moiety in the catalytic pocket in one of its solution-populated conformations. Furthermore, the 6S-NA/N complex is stabilized by polar contacts of the 3-OH and 4-OH groups of Gal with the side chain of N347 and by interactions of GlcNAc and Gal2 with P431 and Q430, respectively (Table 1). The unfavorable effect of the Y347N substitution on NA complex with 3S can be explained by the loss of stabilizing polar interactions between Gal and residue 347. To compensate for this loss, the glycan tilts and shifts toward loop 150, increasing interactions of the Gal and GlcNAc residues with V149 (Table 1).

The initial notion about the association of amino acid 347 with the IV host range (Fanning et al., 2000) was based on the analysis of a small number of N1 NAs and, to the best of our knowledge, has not been followed by systematic analysis. To fill this gap, we assessed the identity of residue 347 in all currently available N1 NA sequences of IVs in different host species. Remarkably, with two exceptions, Y347 was perfectly conserved in avian IVs regardless of their HA subtype, while N347 was highly conserved in classical swine, avian-like swine, and H1N1pdm virus lineages (Table 3). Natural selection analyses based on dN/dS ratios suggested that position 347 was under pervasive purifying selective pressure in these avian and mammalian viruses (Table 5). The NAs of the avian-origin 1918 pandemic virus and its successors isolated in 1933–1947 also contained N347 (Figure 8A). Combined with the modeling data, these results indicate that efficient cleavage of 2-3-linked sialoglycans associated with Y347 is essential for avian IV fitness, and that the Y347N substitution and resulting changes in substrate specificity of NA are essential for avian virus adaptation to humans and pigs. This pattern correlates perfectly with changes in HA receptor-binding specificity during avian-to-human and avian-to-swine transmission (Matrosovich et al., 2006; Shi et al., 2014; Thompson and Paulson, 2020; Liu et al., 2023).

Interestingly, a functional similarity can be observed between Y347 of N1 NA and conserved Q226 of the avian HA. First, both Q226 of HA and Y347 of NA play important roles in protein binding to one of the solution-dominant conformations of Neu5Ac2–3Gal-terminated receptors via interactions with the 4-hydroxyl group of Gal (Supplementary Figure S3). Second, both residues hinder the accommodation of the Gal residue in solution-dominant conformers of Neu5Ac2–6Gal-terminated receptors. Third, the Q226L substitution in avian influenza viruses of diverse HA subtypes occurs during the virus adaptation to humans and swine and alters its preference for the type of Neu5Ac-Gal linkage.

The first viruses of the ALS lineage isolated in 1979 carried substitutions E190D and/or G225E in the HA and bound to both Neu5Ac2-3Gal- and Neu5Ac2-6-Gal-terminated receptors. The binding preference of the virus for 6-linked receptors and transmissibility in pigs under experimental conditions increased over time (Ito et al., 1998; Matrosovich et al., 2000; Su et al., 2021). Because the Y347N substitution appeared and became fixed after 1981 (Figure 8B), we conclude that it was not essential for the initial avian-to-swine transmission and was selected at a later stage, likely, as a compensation for the increased HA avidity for Neu5Ac2–6Gal.

Seasonal human IVs exhibited a distinct pattern of NA evolution with an initial circulation of N347-containing IVs, their replacement by the D347 variants, and the periodic emergence of new N347-containing lineages (Figure 8A). The D347N substitution was previously found to increase NA avidity and catalytic activity with respect to the synthetic substrate MUNANA (Collins et al., 2009; Rameix-Welti et al., 2011; Duan et al., 2014). Collins et al. (2009) observed this effect in IVs from three independent clades of D-to-N mutants. As suggested by the authors, the substitution increased substrate binding by eliminating unfavorable electrostatic interactions between the negatively charged carboxylic groups of Neu5Ac and D347. It has been speculated that the exchange of N and D at position 347 and corresponding changes in NA catalytic activity reflect epistatic interactions of amino acid 347 with other amino acids of NA and/or with the HA (Collins et al., 2009; Duan et al., 2014). For example, either D or N may be selected to compensate alterations of NA catalytic activity and/or HA receptor-binding activity of the antigenic drift variants of these proteins. Although we found no evidence of interactions between residue 347 and other sites in NA, further studies are needed to test this hypothesis and to characterize underlying mechanisms of N/D switching in at the NA position 347 in seasonal IVs.

Most avian IVs bind to Neu5Ac2–3Gal-terminated receptors, however, avian viruses may differ in the ability to recognize the sub-terminal saccharide parts of the sialoglycans. Thus, mallards and other dabbling ducks (*Anatinae*) carry all IV subtypes except H13 and H16 and are thought to play a key role in the maintenance of IVs in nature (Fouchier and Munster, 2009; Verhagen et al., 2021). Duck-adapted IVs exhibit similar receptor-binding traits, in particular, share a preference for binding to Neu5Ac2–3Gal1-3/4GlcNAc-containing receptors, with low tolerance for fucosylation of the GlcNAc moiety [(Gambaryan et al., 2012, 2018) and references therein]. Observed conservation of the NA residue Y347 in avian IVs with HA subtypes H1-H4 and H7-H12 (Tables 3, 5) is consistent with the high conservation of the HA RBS and receptor specificity of IVs in ducks.

Adaptation of aquatic bird viruses to land-based gallinaceous birds, such as chicken and quail, is often accompanied by substitutions in the HA and alteration of the receptor-binding specificity. In contrast to duck IVs, poultry-adapted viruses typically show high-avidity binding to Neu5Acα2–3Galβ1–4GlcNAc-terminated receptors containing fucose and/or sulfate at the GlcNAc residue. Moreover, some of these viruses, for instance, certain H9N2 and H7N9 lineages, acquire substitutions at the conserved positions of the RBS, such as 190, 225, and 226, which facilitate HA binding to Neu5Ac2–6Gal-terminated receptors [for reviews, see Matrosovich et al. (2008), Sriwilaijaroen and Suzuki (2020), Thompson and Paulson (2020), and Zhao and Pu (2022)]. Remarkably, two groups of poultry-adapted IVs studied here differed from the other avian IVs by the presence of substitutions at NA position 347.

IVs with the H6 subtype HA have a distinctive receptor-binding specificity (Gambaryan et al., 2018) and the broadest host range in birds compared to other IV subtypes (Everest et al., 2020). The Y347F substitution emerged on a single occasion in the chicken H6N1 viruses in parallel with the substitutions at HA positions 137, 228 and/or 190 (Figure 9; Table 5). There is no consensus between different research groups about the effect of these substitutions on receptor-binding specificity. Two studies concluded that the substitutions increase HA binding to Neu5Ac2–6Gal-terminated receptors (Ni et al., 2015; Wang et al., 2015), and one study showed that the substitution E190V increased HA binding to sulfated species of Neu5Ac2–3Gal-containing receptors (Kikutani et al., 2020). In any case, the non-conservative substitutions in the canonical conserved positions of the avian RBS should significantly alter receptor specificity of these H6N1 IVs.

Highly pathogenic H5N1 avian IVs with the Gs/Gd-like HA emerged in 1996 in Southern China, spread to other countries and have been causing continuous outbreaks in poultry and wild birds with occasional infections of mammals including humans. These IVs diverged into multiple H5 HA clades with different NA subtypes (Lycett et al., 2019; Demirev et al., 2023; Yang et al., 2023). Due to the wide geographical spread of H5N1 IVs, their broad host range and tissue tropism, as well as immune pressure imposed by vaccination of poultry, a large number of substitutions in the vicinity of the HA RBS were selected that affected both antigenicity and receptor-binding properties of the HA (Paulson and de Vries, 2013; Koel et al., 2014; Suttie et al., 2019). Most of these substitutions likely reflect antigenic drift and/or adaptation of the HA to different species- and tissue-specific Neu5Ac2–3Gal-containing receptors in wild and domestic avian hosts (Kuchipudi et al., 2021; Zhao adn Pu, 2022); some of the substitutions marginally increase HA avidity for human-type receptors. We found that the Gs/Gd-like lineage contained a number of independently emerging NA mutants (typically with H347 and D347), and that the substitutions in the NA were associated with substitutions in or near the HA RBS (Table 4). For example, all three different NA clusters with D347 shown in Figure 10A carried, among other substitutions, the HA substitution at position 193 from K to D, Q or T. A positively charged amino acid (K/R) at this site is responsible for the high-avidity binding of H5, H7, H13 and H16 subtype IVs to sulfated sialoglycans, such as Su-3SLN and Su-SLex (Gambaryan et al., 2012, 2018, and references therein). Therefore, the substitution of K193 with either a negatively charged or uncharged amino acid should have a significant effect on the receptor-binding properties of H5 HA. It is noteworthy that the Y347H NA mutants had different substitutions in the HA compared to the Y347D mutants, indicating the probable functional interplay between the HA and NA substitutions in these H5N1 IVs. We conclude that the substitutions at position 347 in the NA of both H5N1 and H6N1 were selected to alter the NA catalytic activity and restore the HA/NA balance that was disrupted by the substitutions in the HA. However, it cannot be ruled out that the NA substitutions emerged first, followed by selection of corresponding compensatory substitutions in HA.

In summary, our molecular modeling data combined with the analyses of NA and HA sequences strengthen previous observations on the role of amino acid 347 in the N1 NA recognition of the Neu5Ac-Gal linkage, predict the underlying molecular mechanisms, and suggest that substitutions at NA position 347 represent a novel marker of viral host range, interspecies transmission, and adaptive evolution. These findings call for further studies, including in-depth analyses of the catalytic activity and epistatic interactions of position-347 N1 NA mutants as well as identification of analogous host-specific substitutions within the catalytic domain of other NA subtypes.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

MM conceived and planned the structure of the article and supervised the whole project. MM and MG acquired funding. SE, GR, and MG generated and analyzed MD simulation data, MM prepared the sequence datasets and performed phylogenetic analyses, SKP performed selection analyses. SE, SKP and MM analyzed the results, prepared the figures and tables and wrote the manuscript. All authors contributed to the article and approved the final version of the manuscript.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was funded by the Deutsche Forschungsgemeinschaft (German Research Foundation), project number 197785619-SFB 1021 (MM) and by G Ronzoni Foundation (SE and MG). SK was supported in part by a grant from NIH (GM144468). Open Access funding was provided by the Open Access Publishing Fund of Philipps-Universität Marburg.

Acknowledgments

We gratefully acknowledge the Authors from the Originating laboratories responsible for obtaining the specimens and the Submitting laboratories where genetic sequence data were generated and shared via the GISAID Initiative, on which part of this research is based.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1309156/full#supplementary-material

References

Almond, A., Deangelis, P. L., and Blundell, C. D. (2006). Hyaluronan: the local solution conformation determined by NMR and computer modeling is close to a contracted left-handed 4-fold helix. *J. Mol. Biol.* 358, 1256–1269. doi: 10.1016/j.jmb.2006.02.077

Baum, L. G., and Paulson, J. C. (1991). The N2 neuraminidase of human influenza virus has acquired a substrate specificity complementary to the hemagglutinin receptor specificity. *Virology* 180, 10–15. doi: 10.1016/0042-6822(91)90003-t

Case, D., Babin, V., Berryman, J., Betz, R., Cai, Q., Cerutti, D., et al. (2014). *AMBER 14*. University of California: San Francisco.

Chong, A. K., Pegg, M. S., Taylor, N. R., and von Itzstein, M. (1992). Evidence for a sialosyl cation transition-state complex in the reaction of sialidase from influenza virus. *Eur. J. Biochem.* 207, 335–343. doi: 10.1111/j.1432-1033.1992.tb17055.x

Collins, P. J., Haire, L. F., Lin, Y. P., Liu, J., Russell, R. J., Walker, P. A., et al. (2009). Structural basis for oseltamivir resistance of influenza viruses. *Vaccine* 27, 6317–6323. doi: 10.1016/j.vaccine.2009.07.017

Couceiro, J. N., and Baum, L. G. (1994). Characterization of the hemagglutinin receptor specificity and neuraminidase substrate specificity of clinical isolates of human influenza a viruses. *Mem. Inst. Oswaldo Cruz* 89, 587–591. doi: 10.1590/s0074-02761994000400015

Dai, M., Guo, H., Dortmans, J. C., Dekkers, J., Nordholm, J., Daniels, R., et al. (2016). Identification of residues that affect oligomerization and/or enzymatic activity of influenza virus H5N1 neuraminidase proteins. *J. Virol.* 90, 9457–9470. doi: 10.1128/JVI.01346-16

de Vries, E., Du, W., Guo, H., and de Haan, C. A. M. (2020). Influenza a virus hemagglutinin-neuraminidase-receptor balance: preserving virus motility. *Trends Microbiol.* 28, 57–67. doi: 10.1016/j.tim.2019.08.010

Demirev, A. V., Park, H., Lee, K., Park, S., Bae, J.-Y., Park, M.-S., et al. (2023). Phylodynamics and molecular mutations of the hemagglutinin affecting global transmission and host adaptation of H5Nx viruses. *Transbound. Emerg. Dis.* 2023, 1–14. doi: 10.1155/2023/8855164

Duan, S., Govorkova, E. A., Bahl, J., Zaraket, H., Baranovich, T., Seiler, P., et al. (2014). Epistatic interactions between neuraminidase mutations facilitated the emergence of the oseltamivir-resistant H1N1 influenza viruses. *Nat. Commun.* 5:5029. doi: 10.1038/ncomms6029

Elli, S., Gambacorta, N., Rudd, T. R., Matrosovich, M., and Guerrini, M. (2021). MD simulation of the interaction between sialoglycans and the second sialic acid binding site of influenza a virus N1 neuraminidase. *Biochem. J.* 478, 423–441. doi: 10.1042/BCJ20200670

Elli, S., Macchi, E., Rudd, T. R., Raman, R., Sassaki, G., Viswanathan, K., et al (2014). Insights into the human glycan receptor conformation of 1918 pandemic hemagglutinin-glycan complexes derived from nuclear magnetic resonance and molecular dynamics studies. *Biochemistry* 53, 4122–4135. doi: 10.1021/bi500338r

Everest, H., Hill, S. C., Daines, R., Sealy, J. E., James, J., Hansen, R., et al. (2020). The evolution, spread and global threat of H6Nx avian influenza viruses. *Viruses* 12:673. doi: 10.3390/v12060673

Fanning, T. G., Reid, A. H., and Taubenberger, J. K. (2000). Influenza a virus neuraminidase: regions of the protein potentially involved in virus-host interactions. *Virology* 276, 417–423. doi: 10.1006/viro.2000.0578

Fouchier, R. A., and Munster, V. J. (2009). Epidemiology of low pathogenic avian influenza viruses in wild birds. *Rev. Sci. Tech.* 28, 49–58. doi: 10.20506/rst.28.1.1863

Gambaryan, A. S., Matrosovich, T. Y., Boravleva, E. Y., Lomakina, N. F., Yamnikova, S. S., Tuzikov, A. B., et al. (2018). Receptor-binding properties of influenza viruses isolated from gulls. *Virology* 522, 37–45. doi: 10.1016/j.virol.2018.07.004

Gambaryan, A. S., Matrosovich, T. Y., Philipp, J., Munster, V. J., Fouchier, R. A., Cattoli, G., et al. (2012). Receptor-binding profiles of H7 subtype influenza viruses in different host species. *J. Virol.* 86, 4370–4379. doi: 10.1128/JVI.06959-11

Garcia, J. M., Lai, J. C., Haselhorst, T., Choy, K. T., Yen, H. L., Peiris, J. S., et al. (2013). Investigation of the binding and cleavage characteristics of N1 neuraminidases from avian, seasonal, and pandemic influenza viruses using saturation transfer difference nuclear magnetic resonance. *Influenza Other Respir. Viruses* 8, 235–242. doi: 10.1111/irv.12184

Gasteiger, J., and Marsili, M. (1978). A new model for calculating atomic charges in molecules. *Tetrahedron Lett.* 19, 3181–3184. doi: 10.1016/s0040-4039(01)94977-9

Gerlach, T., Kuhling, L., Uhlendorff, J., Laukemper, V., Matrosovich, T., Czudai-Matwich, V., et al. (2012). Characterization of the neuraminidase of the

H1N1/09 pandemic influenza virus. *Vaccine* 30, 7348–7352. doi: 10.1016/j.vaccine.2012.09.078

Hall, T. A. (1999). "BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT" in *Nucleic acids symposium series* (Oxford: Oxford University Press)

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281

Hufnagel, D. E., Young, K. M., Arendsee, Z. W., Gay, L. C., Caceres, C. J., Rajão, D. S., et al. (2023). Characterizing a century of genetic diversity and contemporary antigenic diversity of N1 neuraminidase in influenza a virus from north American swine. *Virus Evol.* 9:15. doi: 10.1093/ve/vead015

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi: 10.1016/0263-7855(96)00018-5

Ito, T., Couceiro, J. N., Kelm, S., Baum, L. G., Krauss, S., Castrucci, M. R., et al. (1998). Molecular basis for the generation in pigs of influenza a viruses with pandemic potential. *J. Virol.* 72, 7367–7373. doi: 10.1128/JVI.72.9.7367-7373.1998

Janakiraman, M. N., White, C. L., Laver, W. G., Air, G. M., and Luo, M. (1994). Structure of influenza virus neuraminidase B/Lee/40 complexed with sialic acid and a dehydro analog at 1.8-a resolution: implications for the catalytic mechanism. *Biochemistry* 33, 8172–8179. doi: 10.1021/bi00193a002

Jongkon, N., and Sangma, C. (2012). Receptor recognition mechanism of human influenza a H1N1 (1918), avian influenza a H5N1 (2004), and pandemic H1N1 (2009) neuraminidase. *J. Mol. Model.* 18, 285–293. doi: 10.1007/s00894-011-1071-y

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi: 10.1063/1.445869

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285

Kandeil, A., Patton, C., Jones, J. C., Jeevan, T., Harrington, W. N., Trifkovic, S., et al. (2023). Rapid evolution of a(H5N1) influenza viruses after intercontinental spread to North America. *Nat. Commun.* 14:3082. doi: 10.1038/s41467-023-38415-7

Karasin, A. I., West, K., Carman, S., and Olsen, C. W. (2004). Characterization of avian H3N3 and H1N1 influenza a viruses isolated from pigs in Canada. *J. Clin. Microbiol.* 42, 4349–4354. doi: 10.1128/JCM.42.9.4349-4354.2004

Kikutani, Y., Okamatsu, M., Nishihara, S., Takase-Yoden, S., Hiono, T., de Vries, R. P., et al. (2020). E190V substitution of H6 hemagglutinin is one of key factors for binding to sulfated sialylated glycan receptor and infection to chickens. *Microbiol. Immunol.* 64, 304–312. doi: 10.1111/1348-0421.12773

Kirschner, K. N., Yongye, A. B., Tschampel, S. M., Gonzalez-Outeirino, J., Daniels, C. R., Foley, B. L., et al. (2008). GLYCAM06: a generalizable biomolecular force field, Carbohydrates. *J. Comput. Chem.* 29, 622–655. doi: 10.1002/jcc.20820

Kobasa, D., Kodihalli, S., Luo, M., Castrucci, M. R., Donatelli, I., Suzuki, Y., et al. (1999). Amino acid residues contributing to the substrate specificity of the influenza a virus neuraminidase. *J. Virol.* 73, 6743–6751. doi: 10.1128/JVI.73.8.6743-6751.1999

Koel, B. F., van der Vliet, S., Burke, D. F., Bestebroer, T. M., Bharoto, E. E., Yasa, I. W., et al. (2014). Antigenic variation of clade 2.1 H5N1 virus is determined by a few amino acid substitutions immediately adjacent to the receptor binding site. *MBio* 5, e01070–e01014. doi: 10.1128/mBio.01070-14

Kosakovsky Pond, S. L., and Frost, S. D. W. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* 22, 1208–1222. doi: 10.1093/molbev/msi105

Krammer, F., Smith, G. J. D., Fouchier, R. A. M., Peiris, M., Kedzierska, K., Doherty, P. C., et al. (2018). Influenza. *Nat. Rev. Dis. Primers.* 4:3. doi: 10.1038/s41572-018-0002-y

Kuchipudi, S. V., Nelli, R. K., Gontu, A., Satyakumar, R., Surendran Nair, M., and Subbiah, M. (2021). Sialic acid receptors: the key to solving the enigma of zoonotic virus spillover. *Viruses* 13:262. doi: 10.3390/v13020262

Levine, B. G., Stone, J. E., and Kohlmeyer, A. (2011). Fast analysis of molecular dynamics trajectories with graphics processing units-radial distribution function Histogramming. *J. Comput. Phys.* 230, 3556–3569. doi: 10.1016/j.jcp.2011.01.048

Liu, M., van Kuppeveld, F. J., de Haan, C. A., and de Vries, E. (2023). Gradual adaptation of animal influenza a viruses to human-type sialic acid receptors. *Curr. Opin. Virol.* 60:101314. doi: 10.1016/j.coviro.2023.101314

Liu, W. J., Wu, Y., Bi, Y., Shi, W., Wang, D., Shi, Y., et al. (2022). Emerging HxNy influenza a viruses. *Cold Spring Harb. Perspect. Med.* 12:8406. doi: 10.1101/cshperspect.a038406

Lycett, S. J., Baillie, G., Coulter, E., Bhatt, S., Kellam, P., McCauley, J. W., et al. (2012). Estimating reassortment rates in co-circulating Eurasian swine influenza viruses. *J. Gen. Virol.* 93, 2326–2336. doi: 10.1099/vir.0.044503-0

Lycett, S. J., Duchatel, F., and Digard, P. (2019). A brief history of bird flu. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 374:20180257. doi: 10.1098/rstb.2018.0257

Macchi, E., Rudd, T. R., Raman, R., Sasisekharan, R., Yates, E. A., Naggi, A., et al (2016). Nuclear magnetic resonance and molecular dynamics simulation of the interaction between recognition protein H7 of the novel influenza virus H7N9 and glycan cell surface receptors. *Biochemistry.* 55, 6605–6616. doi: 10.1021/acs.biochem.6b00693

Matrosovich, M. N., Gambaryan, A. S., and Klenk, H.-D. (2008). "Receptor specificity of influenza viruses and its alteration during interspecies transmission" in *Avian Influenza*. eds. H.-D. Klenk, M. N. Matrosovich and J. Stech (Basel: Karger), 134–155.

Matrosovich, M. N., Klenk, H.-D., and Kawaoka, Y. (2006). "Receptor specificity, host range and pathogenicity of influenza viruses" in *Influenza virology: Current topics*. ed. Y. Kawaoka (Wymondham: Caister Academic Press), 95–137.

Matrosovich, M., Tuzikov, A., Bovin, N., Gambaryan, A., Klimov, A., Castrucci, M. R., et al. (2000). Early alterations of the receptor-binding properties of H1, H2, and H3 avian influenza virus hemagglutinins after their introduction into mammals. *J. Virol.* 74, 8502–8512. doi: 10.1128/Jvi.74.18.8502-8512.2000

McAuley, J. L., Gilbertson, B. P., Trifkovic, S., Brown, L. E., and McKimm-Breschkin, J. L. (2019). Influenza virus neuraminidase structure and functions. *Front. Microbiol.* 10:39. doi: 10.3389/fmicb.2019.00039

Miller, B. R. 3rd, McGee, T. D. Jr., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012). MMPBSA.Py: an efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* 8, 3314–3321. doi: 10.1021/ct300418h

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015

Mochalova, L., Kurova, V., Shtyrya, Y., Korchagina, E., Gambaryan, A., Belyanchikov, I., et al. (2007). Oligosaccharide specificity of influenza H1N1 virus neuraminidases. *Arch. Virol.* 152, 2047–2057. doi: 10.1007/s00705-007-1024-z

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., et al. (2009). AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30, 2785–2791. doi: 10.1002/jcc.21256

Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., and Kosakovsky Pond, S. L. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* 8:e1002764. doi: 10.1371/journal.pgen.1002764

Ni, F., Kondrashkina, E., and Wang, Q. (2015). Structural and functional studies of influenza virus a/H6 hemagglutinin. *PLoS One* 10:e0134576. doi: 10.1371/journal.pone.0134576

Okonechnikov, K., Golosova, O., and Fursov, M. (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28, 1166–1167. doi: 10.1093/bioinformatics/bts091

Paulson, J. C., and de Vries, R. P. (2013). H5N1 receptor specificity as a factor in pandemic risk. *Virus Res.* 178, 99–113. doi: 10.1016/j.virusres.2013.02.015

Phanich, J., Threeracheep, S., Kungwan, N., Rungrotmongkol, T., and Hannongbua, S. (2019). Glycan binding and specificity of viral influenza neuraminidases by classical molecular dynamics and replica exchange molecular dynamics simulations. *J. Biomol. Struct. Dyn.* 37, 3354–3365. doi: 10.1080/07391102.2018.1514326

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. (2005). Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26, 1781–1802. doi: 10.1002/jcc.20289

Poon, A. F. Y., Lewis, F. I., Pond, S. L. K., and Frost, S. D. W. (2007). An evolutionary-network model reveals stratified interactions in the V3 loop of the HIV-1 envelope. *PLoS Comput. Biol.* 3:e231. doi: 10.1371/journal.pcbi.0030231

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490

R Core Team (2013). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available at: http://www.R-project.org/

Raab, M., and Tvaroska, I. (2011). The binding properties of the H5N1 influenza virus neuraminidase as inferred from molecular modeling. *J. Mol. Model.* 17, 1445–1456. doi: 10.1007/s00894-010-0852-z

Raffaini, G., Elli, S., and Ganazzoli, F. (2006). Computer simulation of bulk mechanical properties and surface hydration of biomaterials. *J. Biomed. Mater. Res. A* 77A, 618–626. doi: 10.1002/jbm.a.30670

Raffaini, G., and Ganazzoli, F. (2006). Adsorption of charged albumin subdomains on a graphite surface. *J. Biomed. Mater. Res. A* 76A, 638–645. doi: 10.1002/jbm.a.30546

Rameix-Welti, M. A., Munier, S., Le Gal, S., Cuvelier, F., Agou, F., Enouf, V., et al. (2011). Neuraminidase of 2007-2008 influenza a(H1N1) viruses shows increased affinity for sialic acids due to the D344N substitution. *Antivir. Ther.* 16, 597–603. doi: 10.3851/IMP1804

Reid, A. H., Fanning, T. G., Janczewski, T. A., and Taubenberger, J. K. (2000). Characterization of the 1918 "Spanish" influenza virus neuraminidase gene. *Proc. Natl. Acad. Sci. U. S. A.* 97, 6785–6790. doi: 10.1073/pnas.100140097

Rudiño-Piñera, E., Tunnah, P., and Lukacik, P. (2006). The crystal structure of type A influenza virus neuraminidase subtype N6 reveals the existence of two separate Neu5Ac binding sites. PDB ID: 1W21. Available at: https://www.rcsb.org/structure/1W21 (Accessed October 24, 2022).

Russell, R. J., Haire, L. F., Stevens, D. J., Collins, P. J., Lin, Y. P., Blackburn, G. M., et al. (2006). The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature* 443, 45–49. doi: 10.1038/nature05114

Sassaki, G. L., Elli, S., Rudd, T. R., Macchi, E., Yates, E. A., Naggi, A., et al. (2013). Human (α2→6) and avian (α2→3) Sialylated receptors of influenza a virus show distinct conformations and dynamics in solution. *Biochemistry* 52, 7217–7230. doi: 10.1021/bi400677n

Seeber, M., Cecchini, M., Rao, F., Settanni, G., and Caflisch, A. (2007). Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics* 23, 2625–2627. doi: 10.1093/bioinformatics/btm378

Shi, Y., Wu, Y., Zhang, W., Qi, J., and Gao, G. F. (2014). Enabling the 'host jump': structural determinants of receptor-binding specificity in influenza a viruses. *Nat. Rev. Microbiol.* 12, 822–831. doi: 10.1038/nrmicro3362

Shtyrya, Y. A., Mochalova, L. V., and Bovin, N. V. (2009). Influenza virus neuraminidase: structure and function. *Acta Nat.* 1, 26–32. doi: 10.32607/20758251-2009-1-2-26-32

Shu, Y., and McCauley, J. (2017). GISAID: global initiative on sharing all influenza data – from vision to reality. *Euro Surveill.* 22:494. doi: 10.2807/1560-7917.ES.2017.22.13.30494

Spielman, S. J., Weaver, S., Shank, S. D., Magalis, B. R., Li, M., and Kosakovsky Pond, S. L. (2019). Evolution of viral genomes: interplay between selection, recombination, and other forces. *Methods Mol. Biol.* 1910, 427–468. doi: 10.1007/978-1-4939-9074-0_14

Sriwilaijaroen, N., and Suzuki, Y. (2020). Host receptors of influenza viruses and coronaviruses-molecular mechanisms of recognition. *Vaccines (Basel)* 8:587. doi: 10.3390/vaccines8040587

Su, W., Harfoot, R., Su, Y. C. F., DeBeauchamp, J., Joseph, U., Jayakumar, J., et al. (2021). Ancestral sequence reconstruction pinpoints adaptations that enable avian influenza virus transmission in pigs. *Nat. Microbiol.* 6, 1455–1465. doi: 10.1038/s41564-021-00976-y

Suttie, A., Deng, Y. M., Greenhill, A. R., Dussart, P., Horwood, P. F., and Karlsson, E. A. (2019). Inventory of molecular markers affecting biological characteristics of avian influenza a viruses. *Virus Genes* 55, 739–768. doi: 10.1007/s11262-019-01700-z

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120

Taylor, N. R., and von Itzstein, M. (1994). Molecular modeling studies on ligand binding to sialidase from influenza virus and the mechanism of catalysis. *J. Med. Chem.* 37, 616–624. doi: 10.1021/jm00031a011

Thompson, A. J., and Paulson, J. C. (2020). Adaptation of influenza viruses to human airway receptors. *J. Biol. Chem.* 296:100017. doi: 10.1074/jbc.REV120.013309

Tian, J., Bai, X., Li, M., Zeng, X., Xu, J., Li, P., et al. (2023). Highly pathogenic avian influenza virus (H5N1) clade 2.3.4.4b introduced by wild birds, China, 2021. *Emerg. Infect. Dis.* 29, 1367–1375. doi: 10.3201/eid2907.221149

Varghese, J. N., Colman, P. M., van Donkelaar, A., Blick, T. J., Sahasrabudhe, A., and McKimm-Breschkin, J. L. (1997). Structural evidence for a second sialic acid binding site in avian influenza virus neuraminidases. *Proc. Natl. Acad. Sci. U. S. A.* 94, 11808–11812. doi: 10.1073/pnas.94.22.11808

Varghese, J. N., McKimm-Breschkin, J. L., Caldwell, J. B., Kortt, A. A., and Colman, P. M. (1992). The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor. *Proteins* 14, 327–332. doi: 10.1002/prot.340140302

Vavricka, C. J., Liu, Y., Kiyota, H., Sriwilaijaroen, N., Qi, J., Tanaka, K., et al. (2013). Influenza neuraminidase operates via a nucleophilic mechanism and can be targeted by covalent inhibitors. *Nat. Commun.* 4:1491. doi: 10.1038/ncomms2487

Verhagen, J. H., Eriksson, P., Leijten, L., Blixt, O., Olsen, B., Waldenstrom, J., et al. (2021). Host range of influenza a VIRUS H1 to H16 in eurasian ducks based on tissue and receptor binding studies. *J. Virol.* 95:873. doi: 10.1128/JVI.01873-20

von Itzstein, M. (2007). The war against influenza: discovery and development of sialidase inhibitors. *Nat. Rev. Drug Discov.* 6, 967–974. doi: 10.1038/nrd2400

Wagner, R., Matrosovich, M., and Klenk, H. D. (2002). Functional balance between haemagglutinin and neuraminidase in influenza virus infections. *Rev. Med. Virol.* 12, 159–166. doi: 10.1002/rmv.352

Wang, F., Qi, J., Bi, Y., Zhang, W., Wang, M., Zhang, B., et al. (2015). Adaptation of avian influenza a (H6N1) virus from avian to human receptor-binding preference. *EMBO J.* 34, 1661–1673. doi: 10.15252/embj.201590960

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033

Weis, A., Katebzadeh, K., Soderhjelm, P., Nilsson, I., and Ryde, U. (2006). Ligand affinities predicted with the MM/PBSA method: dependence on the simulation method and the force field. *J. Med. Chem.* 49, 6596–6606. doi: 10.1021/jm0608210

Xu, D., Newhouse, E. I., Amaro, R. E., Pao, H. C., Cheng, L. S., Markwick, P. R., et al (2009). Distinct glycan topology for avian and human sialopentasaccharide receptor analogues upon binding different hemagglutinins: a molecular dynamics perspective. *J Mol Biol.* 387, 465–491. doi: 10.1016/j.jmb.2009.01.040

Xu, J., Davis, C. T., Christman, M. C., Rivailler, P., Zhong, H., Donis, R. O., et al. (2012). Evolutionary history and phylodynamics of influenza a and B neuraminidase (NA) genes inferred from large-scale sequence analyses. *PLoS One* 7:e38665. doi: 10.1371/journal.pone.0038665

Xu, X., Zhu, X., Dwek, R. A., Stevens, J., and Wilson, I. A. (2008). Structural characterization of the 1918 influenza virus H1N1 neuraminidase. *J. Virol.* 82, 10493–10501. doi: 10.1128/JVI.00959-08

Yang, J., Zhang, C., Yuan, Y., Sun, J., Lu, L., Sun, H., et al. (2023). Novel avian influenza virus (H5N1) clade 2.3.4.4b reassortants in migratory birds, China. *Emerg. Infect. Dis.* 29, 1244–1249. doi: 10.3201/eid2906.221723

Yongkiettrakul, S., Nivitchanyong, T., Pannengpetch, S., Wanitchang, A., Jongkaewwattana, A., and Srimanote, P. (2013). Neuraminidase amino acids 149 and 347 determine the infectivity and oseltamivir sensitivity of pandemic influenza a/H1N1 (2009) and avian influenza a/H5N1. *Virus Res.* 175, 128–133. doi: 10.1016/j.virusres.2013.04.011

Zhao, C., and Pu, J. (2022). Influence of host sialic acid receptors structure on the host specificity of influenza viruses. *Viruses* 14:2141. doi: 10.3390/v14102141

frontiers | Frontiers in Microbiology

TYPE Review
PUBLISHED 21 December 2023
DOI 10.3389/fmicb.2023.1272498

OPEN ACCESS

EDITED BY
Shailendra Saxena,
King George's Medical University, India

REVIEWED BY
Ohad Shifman,
Israel Institute for Biological Research, Israel
Milad Zandi,
Tehran University of Medical Sciences, Iran

*CORRESPONDENCE
D. Katterine Bonilla-Aldana
✉ dbonilla@continental.edu.pe

RECEIVED 04 August 2023
ACCEPTED 16 November 2023
PUBLISHED 21 December 2023

CITATION
Taha AM, Katamesh BE, Hassan AR, Abdelwahab OA, Rustagi S, Nguyen D, Silva-Cajaleon K, Rodriguez-Morales AJ, Mohanty A, Bonilla-Aldana DK and Sah R (2023) Environmental detection and spreading of mpox in healthcare settings: a narrative review. *Front. Microbiol.* 14:1272498.
doi: 10.3389/fmicb.2023.1272498

Environmental detection and spreading of mpox in healthcare settings: a narrative review

Amira Mohamed Taha[1,2], Basant E. Katamesh[3,4], Abdul Rhman Hassan[3], Omar Ahmed Abdelwahab[2,5], Sarvesh Rustagi[6], Dang Nguyen[7], Kenneth Silva-Cajaleon[8], Alfonso J. Rodriguez-Morales[8,9,10], Aroop Mohanty[11], D. Katterine Bonilla-Aldana[12]* and Ranjit Sah[13,14]

[1]Faculty of Medicine, Fayoum University, Fayoum, Egypt, [2]Medical Research Group of Egypt (MRGE), Negida Academy, Arlington, MA, United States, [3]Faculty of Medicine, Tanta University, Tanta, Egypt, [4]Mayo Clinic, Rochester, MN, United States, [5]Faculty of Medicine, Al-Azhar University, Cairo, Egypt, [6]School of Applied and Life Sciences, Uttaranchal University, Dehradun, India, [7]Massachusetts General Hospital, Corrigan Minehan Heart Center and Harvard Medical School, Boston, MA, United States, [8]Faculty of Environmental Sciences, Universidad Científica del Sur, Lima, Peru, [9]Grupo de Investigación Biomedicina, Faculty of Medicine, Fundación Universitaria Autónoma de lasAméricas-Institución Universitaria Visión de las Américas, Pereira, Colombia, [10]Gilbert and Rose-Marie Chagoury School of Medicine, Lebanese American University, Beirut, Lebanon, [11]Department of Clinical Microbiology, All India Institute of Medical Sciences, Gorakhpur, India, [12]Research Unit, Universidad Continental, Huancayo, Peru, [13]Tribhuvan University Teaching Hospital, Kathmandu, Nepal, [14]Department of Clinical Microbiology, DY Patil Medical College, Hospital and Research Centre, DY Patil Vidyapeeth, Pune, India

Monkeypox virus (MPXV), which causes Monkeypox (Mpox), has recently been found outside its usual geographic distribution and has spread to 117 different nations. The World Health Organization (WHO) designated the epidemic a Public Health Emergency of International Concern (PHEIC). Humans are at risk from MPXV's spread, which has raised concerns, particularly in the wake of the SARS-CoV-2 epidemic. The risk of virus transmission may rise due to the persistence of MPXV on surfaces or in wastewater. The risk of infection may also increase due to insufficient wastewater treatment allowing the virus to survive in the environment. To manage the infection cycle, it is essential to investigate the viral shedding from various lesions, the persistence of MPXV on multiple surfaces, and the length of surface contamination. Environmental contamination may contribute to virus persistence and future infection transmission. The best possible infection control and disinfection techniques depend on this knowledge. It is thought to spread mainly through intimate contact. However, the idea of virus transmission by environmental contamination creates great concern and discussion. There are more cases of environmental surfaces and wastewater contamination. We will talk about wastewater contamination, methods of disinfection, and the present wastewater treatment in this review as well as the persistence of MPXV on various environmental surfaces.

KEYWORDS

Mpox, contamination, surface, hospital, epidemiology

1 Introduction

1.1 Aetiology

The zoonotic viral disease monkeypox (Mpox), endemic in central and western Africa, has recently spread to numerous nations in both endemic and nonendemic regions. Over 92,048 cases had been reported up to December 2023 across 117 countries, 110 (~94%) of which have not historically reported mpox. Its incubation period varies from a few days to 3 weeks. Unlike the smallpox virus, the monkeypox virus (MPXV) has a wide range of animal reservoirs that spread it to humans sporadically (Ježek et al., 1987; Di Giulio and Eckburg, 2004; CDC, 2022; Kaler et al., 2022; Moore and Zahra, 2022). Its prodromal symptoms, which initially resemble the flu, are followed by lymphadenopathy and a rash on the face before spreading throughout the body.

However, the clinical picture in the current outbreak is abnormal; according to research done on 23 patients, 95% had a rash, almost two-thirds had 10 or more lesions, 73% had anogenital lesions, and 41% had mucosal lesions (Ladnyj et al., 1972; Learned et al., 2005; Beer and Rao, 2019; Thornhill et al., 2022a).

Additionally, roughly 10% of patients had a single genital ulcer, and others had many lesions at once (Thornhill et al., 2022a). The majority of the time, it is a self-limiting disease that goes away 3–4 weeks after the first signs and symptoms arise, with crust formation and desquamation occurring over the next 7–14 days. After the lesion has healed, there may be light scarring that later darkens.

1.2 Epidemiology

Although Mpox outbreaks have grown since their initial detection, they were primarily confined to Africa. In 2003, the United States of America (USA) reported the first cases outside of Africa, with 53 with mpox confirmed infections. United Kingdom (UK) health officials confirmed the Mpox case of a tourist who returned from Nigeria in May 2022. Since then, recorded cases have grown in other nations, including those without a history of travel to endemic regions, but through community transmission, with only a few cases involving women and children. In contrast, most cases involved having sex with men (MSM) (CDC, 2003a; Isidro et al., 2022; Kipkorir et al., 2022; Kozlov, 2022; Mahase, 2022; Americo et al., 2023).

1.3 MPXV clades

Central African and western African are the two primary clades of MPXV, and they are now referred to as clades I and II, respectively. With a case fatality rate of up to 10%, clade I is more severe than clade II, which is less severe with a case fatality rate of less than 1%. Clade II would be divided further into clades IIa and IIb. The MPXV responsible for the current outbreak is related to clade IIb, which is related to clade IIa, responsible for the outbreak in Nigeria in 2017.

That is determined through phylogenetic analysis. It is uncertain what genetic variations are responsible for the greater incidence of clade IIb transmission in humans (Gammon et al., 2010; Isidro et al., 2022; Kipkorir et al., 2022; Kozlov, 2022; Mahase, 2022; Americo et al., 2023). DNA viruses like MPXV typically do not show many mutations. However, in the current outbreak, isolates from 2022 were discovered to have 10 alterations in the MPXV replication complex (RC) and other viral proteins. The potential effects of each mutation are not now fully understood. More study is required to understand further how these recently generated mutations function (Arita et al., 1985; Ježek et al., 1988a; CDC, 2003a; Hutson et al., 2009; Gammon et al., 2010; Isidro et al., 2022; Kannan et al., 2022; Kipkorir et al., 2022; Kozlov, 2022; Mahase, 2022; Americo et al., 2023).

1.4 MPXV transmission

Animal interaction was the leading cause of human infection. However, it could be challenging to pinpoint the precise animal interaction that caused a case in areas where different species are encountered. Rarely in secondary cases, but occasionally in primary human patients, does Human-to-human transfer happen. The evidence suggests that family members or those who care for a Mpox patients are at a higher risk for contracting an infection, even though the efficiency of Mpox human-to-human transmission appears to be less than that of smallpox. It did occur up to 11.7% of household contacts of patients who were not protected against smallpox (Arita et al., 1985; Ježek et al., 1988a; Goldmann, 2000; Hutson et al., 2009; Gammon et al., 2010; Spicknall et al., 2010; Kannan et al., 2022; León-Figueroa et al., 2022; Salvato et al., 2022; Hernaez et al., 2023; Karagoz et al., 2023; CDC, 2023a).

The environment, the virus's ability to infect the host, viral shedding, the host's susceptibility, and behavior all impact the virus's spread. Additional essential environmental elements that affect transmission include host density and mobility patterns. Lesion exudate, crust material, respiratory droplets, and viral shedding in faeces are all possible routes for human transmission (Goldmann, 2000; León-Figueroa et al., 2022; Salvato et al., 2022; Atkinson et al., 2022a; Hernaez et al., 2023; Karagoz et al., 2023; Mellon et al., 2023; Sharkey et al., 2023; CDC, 2023a).

It is most likely spread through intimate contact with an infected person and possibly transmitted by big droplets that settle after about one meter; however, the evidence of respiratory transmission is still questionable (Goldmann, 2000; Salvato et al., 2022; Atkinson et al., 2022a; Hernaez et al., 2023). Mpox is spread by close, personal contact, mainly through sexual contact in the current outbreak. Indirect contact with contaminated surfaces and objects is another route that could contribute to viral transmission. A recent investigation in the UK revealed the presence of MPXV DNA on several surfaces from locations visited by a tourist returning from Nigeria with a confirmed case of Mpox (Atkinson et al., 2022a). According to a different investigation, fomite transmission was regarded as a plausible primary infection source in two cases. These studies support environmental assessments since polluted fomite can spread infectious diseases. Fomites might contract the virus by coming into intimate contact with human fluids or secretions, touching contaminated hands, or contacting respiratory droplets that land on various surfaces (Lu et al., 2022; Palich et al., 2022; Atkinson et al., 2022a; Girón-Guzmán et al., 2023; Hernaez et al., 2023; La Rosa et al., 2023; Mellon et al., 2023; Sharkey et al., 2023; CDC, 2023b). Despite patients wearing surgical masks, a different investigation found viral DNA in the air in the outpatient consultation room during visits from Mpox patients. Therefore, the viral particles found may have been transmitted through the air from the skin, sores on the genitalia or oropharynx, or respiratory secretions. Despite this, neither healthcare professionals nor consulting Mpox-negative patients reported any MPXV symptoms during the trial period or in the 21 days that followed (Mellon et al., 2023). Infections of the Mpox virus are transmitted and closely watched due to wastewater (Girón-Guzmán et al., 2023; La Rosa et al., 2023; Sharkey et al., 2023). Since human fluids that are contaminated with the virus may be present in wastewater, it is possible that MPXV can spread there. Therefore, public health officials' wastewater analysis can provide a clear picture of the scope of the Mpox outbreak and potentially aid in its containment (Ježek et al., 1988b; Nakoune et al., 2017; Adler et al., 2022; Ferré et al., 2022; Lu et al., 2022; Ogoina and Yinka-Ogunleye, 2022; Palich et al., 2022; Riopelle et al., 2022; Thornhill et al., 2022b; Owhonda et al., 2023; Wieder-Feinsod et al., 2023; CDC, 2023b,c).

The possibility of cross-border diffusion and subsequent transmission of Mpox has increased due to the ambiguity surrounding the outbreak's containment and the risk of social transmission.

Unlike earlier cases, most current cases do not have confirmed travel connections to endemic regions, indicating a crucial role in community and environmental transmission. That review covers the most recent research on the role of the environment in the transmission of Mpox, the importance of wastewater surveillance and monitoring, and the potential transmission of illnesses pathways (Ježek et al., 1988b; Adler et al., 2022; Lu et al., 2022; Palich et al., 2022; Thornhill et al., 2022b; Girón-Guzmán et al., 2023; La Rosa et al., 2023; Sharkey et al., 2023; Wieder-Feinsod et al., 2023; CDC, 2023b,c).

2 Viral shedding

Identifying viral shedding to improve treatment and stop further transmission is essential. Because infected skin lesions have the largest viral loads, the highest favorable rates, and the least invasive, quantitative real-time polymerase chain reaction (qPCR) techniques, have been suggested for detecting MPXV in these infected skin lesion swabs. These assessments are most effective when combined with clinical and epidemiological data, such as vaccination history (Ježek et al., 1988b; Adler et al., 2022; Lu et al., 2022; Ogoina and Yinka-Ogunleye, 2022; Palich et al., 2022; Riopelle et al., 2022; Thornhill et al., 2022b; Girón-Guzmán et al., 2023; La Rosa et al., 2023; Owhonda et al., 2023; Wieder-Feinsod et al., 2023; CDC, 2023b,c).

MPXV was found in various swabs, including anal, contaminated objects, skin, saliva, oropharynx, conjunctiva, vaginal fluids, breast milk, semen, blood, urine, and faeces. Given that the viral load in diverse clinical samples coincided with the quantity of replication-competent MPXV found, a higher viral concentration, as assessed by qPCR, can forecast a larger potential for infectivity. After the sickness first appears, the virus might shed for up to 3 weeks. However, it persisted longer in samples with cycle threshold (Ct) values higher than 35, lasting up to 8 weeks for semen and up to 10 weeks for saliva samples, and 6 to 10 weeks for respiratory samples (Ježek et al., 1986, 1988b; Nakoune et al., 2017; Ogoina et al., 2019; Vaughan et al., 2020; Adler et al., 2022; Ferré et al., 2022; Ogoina and Yinka-Ogunleye, 2022; Riopelle et al., 2022; Zachary and Shenoy, 2022; Thornhill et al., 2022b; Owhonda et al., 2023; Wieder-Feinsod et al., 2023).

Numerous diagnostic procedures, such as qPCR, immunohistochemistry, serology, and electron microscopy, could be used to determine the presence of Mpox. The best method for experimental diagnosis is known to be qPCR. However, in an outbreak like this, it may not always be practical because every diagnostic process has to adhere to strict diagnostic and biosafety performance standards regarding sampling technique, storage, and other biosafety requirements, which delays the diagnostic process and causes additional costs. Furthermore, it only has a limited impact in rural places with few resources. Given the rising prevalence of Mpox, quick screening techniques, like those found in SARS-CoV-2 kits, should be created to be employed as a large-scale diagnostic approach for screening and detecting vulnerable people (Ježek et al., 1986; CDC, 2003b; Nakoune et al., 2017; Ogoina et al., 2019; Vaughan et al., 2020; Gould et al., 2022; Morgan et al., 2022; Nörz et al., 2022; Ogoina and Ogunsola, 2022; Pfeiffer et al., 2022; Zachary and Shenoy, 2022).

CDC stated that a patient is no longer contagious when all scabs have gone off, contradicting the traditional view that Mpox patients are infectious until all lesions have crusted. Recent research has shown that MPXV DNA can still be found in the upper respiratory system days after lesions have healed. It is uncertain whether patients with crusted skin lesions and positive upper respiratory tract swabs are contagious. Further study is needed in this area because the healing of skin lesions served as the leading indicator of the Mpox patients' infectiousness. These findings might prompt additional adjustments to the rules for discharge and quarantine days, with immediate implications for using healthcare resources (CDC, 2003b; Gould et al., 2022; Morgan et al., 2022; Nörz et al., 2022; Ogoina and Ogunsola, 2022; Pfeiffer et al., 2022; Atkinson et al., 2022a,b).

Misdiagnosis of Mpox as another sexually transmitted disease (STD) can delay isolation and result in inadequate treatment, prolonging and worsening the current epidemic. According to several studies, misdiagnosis occurred due to the clinical presentation, which was unusual compared to African outbreaks. A case report showed an atypical distribution of rash that had not happened on the face, in contrast to the recognized national case definition. Only 11% of patients in the UK experienced the rash, the primary diagnostic sign of Mpox, and 20% did not share a prodrome before the rash. Clinical symptoms that are distinct from those that have previously been documented may be the result of an atypical method of transmission. A recent study found that during the 2017–2018 Nigerian outbreak, risky sexual conduct was highly prevalent and that 81.2% of Mpox-infected patients had genital ulcers. Due to this abnormal presentation of Mpox, healthcare professionals may have incorrectly diagnosed Mpox patients with ordinary STDs without thoroughly considering the range of other possible differential diagnoses. Some of the outbreak's less frequently reported cases may also be due to the less severe clinical symptoms associated with clade 2 MPXV. Patients who feel good might also not seek medical attention (Patrono et al., 2020; Jonge et al., 2022; Nörz et al., 2022; Peiró-Mestres et al., 2022; Pfeiffer et al., 2022; Sharkey et al., 2022; Atkinson et al., 2022a,b; CDC, 2023d).

Semen samples from asymptomatic people in this outbreak contained MPXV DNA, suggesting hat asymptomatic transmission may have contributed to the virus's spread. In a recent investigation, rectal swabs from 200 asymptomatic MSM cases revealed positive Mpox qPCR results. To establish the presence of a contagious virus, isolating it from tested samples is necessary because qPCR results cannot precisely predict the chance of infection. The existence of asymptomatic Mpox cases raises concerns that the true number of patients may be underestimated and that the virus may be spread to close contacts even when there are no symptoms. Collecting more accurate epidemiologic data on the prevalence, clinical signs, transmission mechanism, and asymptomatic cases in the general population and individuals with other STDs is necessary (Patrono et al., 2020; Jonge et al., 2022; Peiró-Mestres et al., 2022; Sharkey et al., 2022; Wurtzer et al., 2022; Gazecka et al., 2023; Tiwari et al., 2023; Wannigama et al., 2023; CDC, 2023d).

3 Environmental contamination of MPXV

3.1 MPXV persistence on surfaces in healthcare settings

The contamination of healthcare settings by MPXV and its ability to cause infection is also vital when dealing with the recent outbreak, significantly helping protect healthcare providers from getting infected (Nakoune et al., 2017). Healthcare professionals are a high-risk group for getting infected and infecting others as they are exposed to many patients daily.

Few cases of healthcare-associated human Mpox have been recorded in endemic African nations throughout history. These cases were documented in the Democratic Republic of the Congo in 1983 (Ježek et al., 1986), the Republic of the Congo in 2003 (Learned et al., 2005), the Central African Republic in 2015 and 2016 (Nakoune et al., 2017), and Nigeria in 2017 and 2018 (Ogoina et al., 2019). Although all these exposures took place in hospital Settings, the precise routes of spread could not be determined. Between 2003 and 2021, at least 250 healthcare professionals had varying levels of unprotected contact with MPXV while working in a hospital environment. Yet, only one incidence of nosocomial transmission is documented in the literature (Vaughan et al., 2020; Zachary and Shenoy, 2022). That case report represented a healthcare worker identified as having Mpox in the UK in 2018 and whose only known exposure risk was replacing a patient's possibly contaminated bedding (Vaughan et al., 2020).

Further, fomite transmission has been reported as a single route of transmission of Mpox in two cases out of 3,924 cases (Salvato et al., 2022). In this study, two cases of fomite transmission involved healthcare workers who spent an hour at the patient's home, wore PPE (N95 masks, eye protection, gowns), only used gloves when taking clinical samples from patients, and avoided direct physical contact with the patient. Authors reported that Mpox could have been transmitted exclusively through fomites in those two cases in contrast to 2,420 out of 3,924 cases with direct physical or sexual contact as the transmission mode (Salvato et al., 2022). Although indirect transmission through contaminated objects or surfaces has been reported, the evidence of fomite transmission for Mpox is currently limited (Vaughan et al., 2020; Kozlov, 2022). However, it is still a rare method of transmission. Other transmission methods are already considered, and variable preventive measures are applied for their prevention.

After demonstrating the evidence of the possibility of nosocomial transmission and evidence of surface contamination at the homes of the infected persons, the question of whether surface contamination in hospital settings and whether contaminated objects are a cause of Mpox dissemination is one the scientific community is preoccupied with (Kozlov, 2022; Ogoina and Ogunsola, 2022).

A study conducted in Germany evaluated the contamination of surfaces in hospital rooms and found that the most significant virus loads were found in toilets. However, contamination was present on all surfaces contacted directly by the patients. It was also detected on the patients' mobile phones, toilet seats, and chair seat surfaces (Nörz et al., 2022). Fabrics used frequently by the patients were also contaminated with viruses. After touching these fabrics by investigators, the investigator's gloves were swabbed on the palmar side, and the contamination on that side was determined. Interestingly, these samples showed positive culture, which supports the evidence of infection from touching and objects used or touched by any infected person. In addition, all hand-contact sites, such as the door handle, in the anteroom that were investigated showed positive qPCR findings (CDC, 2003b; Ogoina et al., 2019; Vaughan et al., 2020; Gould et al., 2022; Morgan et al., 2022; Nörz et al., 2022; Ogoina and Ogunsola, 2022; Pfeiffer et al., 2022; Zachary and Shenoy, 2022).

A more extensive study in a UK hospital (Gould et al., 2022) found MPXV DNA in 56 (93%) of 60 surface swab samples taken from patient bedrooms and toilets. Some of the samples that tested positive were from places patients were not likely to have touched directly, such as the air vent over the bathroom door and the air vent above the bedroom door, indicating non-contact contamination, presumably through respiratory droplets or particles suspended in the air after changing bedding.

The DNA of the virus was also detected from two samples swapped from the gloves. The discovery of the contagious MPXV in air samples taken during a bedding change emphasizes the necessity of respiratory protection for healthcare professionals while conducting activities that could contain infectious material in contaminated settings (Ježek et al., 1986; CDC, 2003b; Nakoune et al., 2017; Ogoina et al., 2019; Vaughan et al., 2020; Ferré et al., 2022; Gould et al., 2022; Morgan et al., 2022; Nörz et al., 2022; Ogoina and Ogunsola, 2022; Ogoina and Yinka-Ogunleye, 2022; Pfeiffer et al., 2022; Zachary and Shenoy, 2022; Atkinson et al., 2022b).

There was variability in the frequency of virus detection from one patient's room to another. That can be attributed to disease severity as the extent of contamination varies depending on viral load. Patients suffering from severe illness are more likely to have a higher viral load, leading to increased shedding and potentially higher environmental contamination. Also, the period during the patient's illness when environmental sampling was done, staff or patient behavior, and Infection control measures. However, no variation in cleaning procedures might account for the variations in environmental sample data, and none of the patients' clinical features can account for the variations in the findings of the air sampling for the various isolation rooms examined.

Other factors might have accounted for this variability, necessitating further investigation (Ježek et al., 1986; CDC, 2003b; Nakoune et al., 2017; Ogoina et al., 2019; Vaughan et al., 2020; Ferré et al., 2022; Gould et al., 2022; Morgan et al., 2022; Nörz et al., 2022; Ogoina and Ogunsola, 2022; Pfeiffer et al., 2022; Zachary and Shenoy, 2022; Atkinson et al., 2022b).

In a cross-sectional study, Hernaez et al. (2023) assessed the presence of MPXV DNA in saliva, exhaled droplets from a mask, and aerosols from patients with qPCR-confirmed Mpox infection attending two healthcare centers in Spain. They reported high viral load detected by qPCR in saliva samples and viable virus in 66% of qPCR-positive saliva samples. These findings signify that saliva might assist in contaminating surfaces with the infectious virus, respiratory droplets, aerosols. According to studies above (Gould et al., 2022; Nörz et al., 2022), MPXV DNA has been found by qPCR on various surfaces in hospital rooms occupied by Mpox patients. Some contained the contagious virus, which the previous findings could explain. The exact persistence of MPXV in healthcare settings and its availability to cause infection is not identified definitively. However, the evidence of its persistence for 15 days in a home setting is alarming and should increase our attention (Morgan et al., 2022). Thus, measuring its infectivity power and its existence raises a major concern. Numerous variables might affect a human being successfully being infected with a virus; therefore, just because a virus is detected in environmental samples does not always indicate that transmission resulting in infection would occur if that virus were exposed to a person (Gould et al., 2022). These variables include modes of transmission, host susceptibility, environmental elements that may impair the virus capacity to replicate and infect cells, and the quantity of virus to which a person is exposed (Gould et al., 2022). The infectious dosage ofMPXV in people is unknown and may vary depending on the body part exposed to the virus.

Various lessons should be considered during this outbreak to protect healthcare providers and prevent the spread of the virus in hospital settings. Firstly, the pervasive surface contamination of the patient care environment necessitates a methodical, standard approach to surface disinfection of hospital settings and homes of patients with Mpox. Secondly, the discovery of MPXV DNA on personal protective equipment (PPE), in doffing areas, and air samples were taken at different distances from the patient bed and during bed sheet changes emphasizes the significance of using and removing PPE properly by healthcare professionals to prevent exposure to Mpox while providing patient care (Gould et al., 2022).

Third, since changing bedding may spread MPXV particles, staff members should wear surgical masks (with or without a face shield) at the very least to protect their mucous membranes (CDC, 2003b; Gould et al., 2022; Ogoina and Ogunsola, 2022). Therefore, disinfectants must be used for sites of isolation of Mpox patients.

3.2 MPXV persistence on surfaces in households

Surface contamination may contribute to the spread of infection and is therefore essential to study, especially in the presence of asymptomatic cases that may contribute to virus transmission. Several studies have been conducted to measure MPXV contamination on different surfaces (Nörz et al., 2022; Pfeiffer et al., 2022; Atkinson et al., 2022b). Despite this, little is known about the extent of surface contamination by MPXV. There are several gaps in knowledge, such as the duration of contamination of different surfaces, and further studies are necessary to provide more detailed and accurate information.

CDC and the Utah Department of Health and Human Services (UDHHS) investigated MPXV surface contamination on frequently used household surfaces in May 2022 (Pfeiffer et al., 2022). Patients were still symptomatic at the time of sample collection, obtained from 9 different locations throughout the house and examined using qPCR. Of them, 21 of the 30 samples from porous objects tested positive for Mpox, and 17 of 25 non-porous items developed positive qPCR results. Fabric surfaces such as blankets and chaise lounge tested, and hard surfaces such as light handles and a keyboard tested positive. However, none of the surfaces that tested positive by qPCR yielded a positive viral culture test. It is suggested that cleaning or passing of time may have affected the viral load and inactivated the virus (Ježek et al., 1986; CDC, 2003b; Ogoina et al., 2019; Vaughan et al., 2020; Gould et al., 2022; Morgan et al., 2022; Nörz et al., 2022; Ogoina and Ogunsola, 2022; Zachary and Shenoy, 2022).

Additionally, environmental sampling was carried out in the house of a resident in the USA who had confirmed infection with the west African clade after travelling to Nigeria 15 days after the patient left the residence. Results suggested substantial MPXV DNA contamination as seven samples yielded viable viruses isolated in cell culture. That study demonstrated no difference in qPCR positivity between porous and non-porous surfaces but noted a significant difference in the virus detected in cultures. That suggests that porous surfaces may pose a higher risk of Mpox infection (Morgan et al., 2022).

In June 2022, another study investigated the presence and extent of surface contamination in the hospital rooms of two patients infected with MPXV on the fourth hospitalization day. All surfaces touched by both patients showed viral contamination, with the highest detected on bathroom surfaces. According to the number of viral copies per cm^2 in the first patient, the highest viral load was found on the tap control lever, the seating surface toilet seat, and the mattress cover. In the second, the highest viral load was found on the towel in bed, soap dispenser lever, and pillowcase used to cover cooling packs, followed by the glove used by the examiner. After handling fabrics, the examiners' gloves were immediately swabbed from the palmer side and found to have contamination levels in both patients (Nörz et al., 2022). The investigators had successfully isolated MPXV using three different samples, each containing a minimum of 10^6 virus copies. As a result, contaminated surfaces that carry such viral loads or higher may be contagious (Pfeiffer et al., 2022; Atkinson et al., 2022a; CDC, 2023d).

The CDC in the United States of America has advised that close contacts of Mpox patients should watch for any MPXV symptoms for 3 weeks following the last exposure. In addition, the CDC recommended getting vaccinated and seeking medical attention if an unexplained rash emerged after touching an infected person. Sharing dishes, towels, bed sheets, clothes, drinking glasses, or other private belongings is also not advised (CDC, 2003b; Brown and Leggat, 2016; Patrono et al., 2020; Dye and Kraemer, 2022; Gould et al., 2022; Jonge et al., 2022; Morgan et al., 2022; Nörz et al., 2022; Ogoina and Ogunsola, 2022; Peiró-Mestres et al., 2022; Pfeiffer et al., 2022; Sharkey et al., 2022; Wurtzer et al., 2022; Atkinson et al., 2022a,b; Gazecka et al., 2023; Tiwari et al., 2023; Wannigama et al., 2023; CDC, 2023d). It is strongly recommended to disinfect surfaces, and this should be done with cleaning supplies from the CDC and Environmental Protection Agency lists to prevent cross-contamination. Disposable gloves, a face mask, and clothing covering their arms and legs are all recommended when disinfecting. These suggestions are examples of ways to stop the transmission of Mpox; however, other nations may have different laws; therefore, they may not apply there (Klein, 1963; Kampf et al., 2002; Wutzler and Sauerbrei, 2004; Kampf et al., 2007; Rabenau et al., 2010; Steinmann et al., 2012; Brown and Leggat, 2016; Petersen et al., 2016; Becker et al., 2017; Siddharta et al., 2017; Dye and Kraemer, 2022; Ogoina and Ogunsola, 2022; Upadhayay et al., 2022).

3.3 MPXV contaminating wastewater

Another concern was the contamination of wastewater. It is not fully known how MPXV ends up in sewage. Excretion and secretion of MPXV through faeces were observed in infected animals and humans (Patrono et al., 2020; Peiró-Mestres et al., 2022). Another plausible transmission mode is by skin flakes from affected body areas that wash into wastewater; this is supported by the high viral concentration found in blisters and scabs (Jonge et al., 2022).

According to data from various body fluids, viral shedding can occur from several body sites, including saliva, semen, urine, and faeces. That suggests that bodily fluids may play a part in the spread of disease (Peiró-Mestres et al., 2022).

To check for the existence of MPXV nucleic acids, a team of researchers examined the presence of DNA and RNA extracts from wastewater. Viral particles were investigated by examining the presence of packaged DNA in the viral capsid. Wastewater included human beta-2-microglobulin (B2M) RNA, indicating that excretion and washing can cause the nucleic acids found in human cells to be discovered in sewage (Jonge et al., 2022; Ogoina and Ogunsola, 2022; Peiró-Mestres et al., 2022; Sharkey et al., 2022; Wurtzer et al., 2022; Gazecka et al., 2023; Tiwari et al., 2023). Midway through July, MPXV was found in the wastewater treatment facility for the first time. Over time, higher quantities were observed. Data also revealed a correlation between wastewater levels and the amount of reported clinical cases (Brown and Leggat, 2016; Petersen et al., 2016; Dye and Kraemer, 2022; Ogoina and Ogunsola, 2022; Tiwari et al., 2023).

In addition, a recent study in Amsterdam detected MPXV DNA in wastewater samples from two wastewater treatment facilities, Schiphol Airport, and five different city districts. The sample was deemed positive if both the generic and west-African qPCR were present. Out of 108 wastewater samples, MPXV DNA was discovered in 45 of them. Because infected people may have passed through, samples from the airport occasionally tested positive for MPXV DNA. A sporadic rise in the number of positive samples taken in Schiphol was associated with an increase in confirmed cases. That is consistent with the study's findings described earlier. To verify the DNA specificity, qPCR and an extra traditional qPCR were performed on a predetermined subset of the obtained samples (Klein, 1963; Kampf et al., 2002; Wutzler and Sauerbrei, 2004; Kampf et al., 2007; McDevitt et al., 2007; CDC, 2008; McDevitt et al., 2008; Russell, 2008; Eterpi et al., 2009; Rabenau et al., 2010; de Oliveira et al., 2011; Steinmann et al., 2012; Bleichert et al., 2014; Eggers et al., 2015; Brown and Leggat, 2016; Campagna et al., 2016; Petersen et al., 2016; Becker et al., 2017; Siddharta et al., 2017; Dye and Kraemer, 2022; Ogoina and Ogunsola, 2022; Upadhayay et al., 2022; Meister et al., 2023).

Researchers collected wastewater samples between May and August 2022 from 63 sewered and non-sewered places in Bangkok's city centre. qPCR was used to measure the MPXV DNA copy counts, and Sanger sequencing was used to validate the results as positive. Beginning in the second week of June 2022, wastewater samples containing MPXV DNA had a mean copy number of 16.4 copies/mL. Sanger sequencing of positive samples confirmed the MPXV's existence. According to preliminary analyses, the MPXV DNA from wastewater samples also belonged to the West African lineage. All places where MPXV DNA was detected had closed, un-severed sewage systems and were either public or commercial (Wannigama et al., 2023).

Investigators in Spain found MPXV DNA in 56 of 312 wastewater samples they took from various country locations. Their CT scores varied from 39.98 to 34.5. Interestingly, samples taken from Madrid WWTPs during week 23 of 2022, when just 275 cumulative cases had been reported across the entire region, consistently contained MPXV DNA. Additionally, some WWTPs with limited confirmed clinical patients reported intermittent detection (Girón-Guzmán et al., 2023). That could mean that silent infections are occurring more frequently than anticipated and that the reported clinical cases appear to be underestimated. To estimate the actual number of Mpox cases, wastewater monitoring is crucial during an outbreak. Another Polish study that revealed no relationship between the quantity of hospitalized patients in a given area and MPXV detection in WWTPs supports this result (Gazecka et al., 2023). That is in contrast to a study that found that the discovery and quantification of the MPXV genome in sewersheds in Paris coincided temporally with the discovery of the first case of infection and the spread of the disease throughout the population connected to the sewage system (Wurtzer et al., 2022).

Wastewater surveillance is a well-established supplemental epidemiologic tool that has been effectively employed for viral infectious illnesses, including SARS-CoV-2 and polio; therefore, investigating potential strategies for monitoring MPXV through these systems is crucial. Three distinct PCR tests that were previously developed for clinical samples were put to the test by a research team. To lessen the impact of nucleotide mismatches, they altered the tests by making alterations to the primer and probe sequences. Using real-time or nested PCR and sequencing, three samples out of 20 tested positive for MPXV, demonstrating that these techniques can be used for wastewater-based epidemiology for Mpox outbreaks and offering fundamental resources, such as analytical techniques (La Rosa et al., 2023; Tiwari et al., 2023).

Untreated wastewater can track the movement and dispersal of many diseases. As it contains a variety of biological materials, such as skin, vesicular fluid, saliva, semen, faeces, and respiratory and nasal secretions, it is perfect for monitoring. A customizable platform called wastewater-based surveillance (WBS) can show in real-time when infectious pathogens are being shed; genetic material may be discovered days before symptoms or a healthcare facility's confirmation of infection. WBS has advantages and disadvantages; benefits include cost-effectiveness, independence from testing capacity, patient permission to test, and data utilization by third parties. Disadvantages include the absence of standardization and interference from different substances; since WBE is not always sensitive enough to pick up minute amounts of pathogens, it might be challenging to obtain reliable results from wastewater samples (Gerba et al., 1980; Farahbakhsh and Smith, 2004; John and Rose, 2005; McDevitt et al., 2007; CDC, 2008; McDevitt et al., 2008; Russell, 2008; Gundy et al., 2009; de Oliveira et al., 2011; Bleichert et al., 2014; Campagna et al., 2016; Sassi et al., 2018; Meister et al., 2023).

Detection of MPXVs in wastewater depends on many variables, including infection rate, water flow, viral shedding, and the process used for detection and analysis. Cross-reaction between the assay and non-targeted pathogens is a challenging possibility as wastewater contains many different microbes from multiple sources. The most concerning challenge with WBS is the lack of standard procedures and methodology used to collect samples, measure viral concentrations, extract DNA, and interpret data. Due to the scarcity of essential data on Mpox infection, it's challenging to create a prediction model or to establish a correlation between reported clinical cases and available WBS data. Further research is necessary to fill in critical knowledge gaps, such as the persistence of MPXV in different environments, MPXV DNA shedding in other body fluids, and the extent of MPXV infectivity in wastewater. Filling in these gaps is necessary to quickly develop a robust global WBS network (Klein, 1963; Wutzler and Sauerbrei, 2004; McDevitt et al., 2007, 2008; Eterpi et al., 2009; Rabenau et al., 2010; de Oliveira et al., 2011; Bleichert et al., 2014; Eggers et al., 2015; Campagna et al., 2016; Meister et al., 2023).

4 Controlling Mpox transmission

4.1 Disinfection of surfaces

The use of disinfectants to eliminate MPXV from the surfaces and environment is now considered a vital issue after proving its presence on the surfaces and in the environment and its possibility to cause infection (Gould et al., 2022; Nörz et al., 2022; Ogoina and Ogunsola, 2022; Atkinson et al., 2022a). Therefore, disinfection around the confirmed cases may be necessary to lessen the risk of viral transmission via contaminated surfaces. Consequently, it is crucial to understand which disinfectants and biocidal chemicals are efficient against MPXV and other orthopoxviruses, even though it is doubtful that mpox would cause a worldwide health disaster (Dye and Kraemer, 2022).

Vaccinia virus has been extensively studied and used as a surrogate for other orthopoxviruses, including MPXV, due to its availability, ease of handling, and well-established protocols for inactivation an disinfection (Brown and Leggat, 2016). Many disinfection protocols and guidelines for orthopoxviruses, including those recommended by the World Health Organization, are based on studies using vaccinia virus as a surrogate (Petersen et al., 2016). While there may be some differences between the vaccinia virus and MPXV in their behavior in certain disinfection conditions (Upadhayay et al., 2022), vaccinia virus as a surrogate is generally considered appropriate, especially when primary data for MPXV is unavailable.

Ethanol (ranging 50%–90%) proved effective in suspension experiments against the vaccinia virus strain Elstree and the modified vaccinia virus Ankara (MVA) at concentrations ranging from 50% to 95% within 1 min, even with varied organic loads (Kampf et al., 2002, 2007; Steinmann et al., 2012; Siddharta et al., 2017). Phosphoric acid added to ethanol at 45% also worked for 30 s. However, ethanol only showed marginal effectiveness in doses of 40% or less (Steinmann et al., 2012). Formulas containing two kinds of alcohol and a combined alcohol content between 75% and 77.8% were likewise quite effective in 15 s (Kampf et al., 2002, 2007).

Peroxides also showed good effectiveness against the vaccinia virus (Kampf et al., 2002). In suspension experiments, hydrogen peroxide proved efficient against the vaccinia virus at 14.4% concentration in 30 s (Wutzler and Sauerbrei, 2004; Rabenau et al., 2010; Becker et al., 2017). Additionally, it was shown that peracetic acid rapidly rendered vaccinia viruses inactive at concentrations between 0.005 and 0.2% within 1 min and with 10% fetal calf serum (FCS) as an organic load (Rabenau et al., 2010; Becker et al., 2017). Glutaraldehyde was successful in suspension tests against the vaccinia virus strains Elstree and MVA at concentrations between 0.05 and 0.5% within 5 min, often in the presence of 10% FCS (Klein, 1963; Rabenau et al., 2010). However, significant activity against vaccinia viruses was not documented at shorter contact periods of 30 s or 2 min (Klein, 1963; Rabenau et al., 2010).

With a low organic load, chlorine proved effective against vaccinia viruses in suspension testing at 0.64% in 1 min and 0.525% in 3 min (Eterpi et al., 2009; Eggers et al., 2015). Lower concentrations either needed longer exposure durations or were not powerful enough. However, increased albumin content as an organic load decreased the virucidal effectiveness (Eterpi et al., 2009). Under clean and unclean test circumstances, iodine was also potent against vaccinia viruses at concentrations between 0.045 and 1% within 1 min (de Oliveira et al., 2011; Eggers et al., 2015). However, under dirty test circumstances on artificially contaminated stainless-steel carriers, sodium hypochlorite (0.25% and 2.5%) proved also efficient against the vaccinia virus in one minute (Eterpi et al., 2009; de Oliveira et al., 2011; Eggers et al., 2015).

An alkaline cleanser at 0.9% was shown to inactivate the vaccinia virus in 10 min under unclean conditions on artificially polluted stainless-steel carriers (Eterpi et al., 2009). Ultraviolet light (254 nm) has been shown to inactivate aerosolized vaccinia virus strain WR in a tabletop one-pass aerosol chamber in 7.6 s (McDevitt et al., 2007, 2008).

The vaccinia virus was more vulnerable to UVC with decreased relative air humidity (McDevitt et al., 2007). Similar outcomes were seen when the Western Reserve strain of the vaccinia virus was exposed to UVC light for 10 min through aerosol (254 nm) (McDevitt et al., 2007, 2008).

In summary, vaccinia viruses could be rendered inactive by a 1 min application of 70% ethanol, 0.25% and 2.5% sodium hypochlorite, 14.4% hydrogen peroxide, 0.64% chlorine, and 0.045 and 1% iodine. It also can be killed by a 3 min application of 99.9 copper or 0.525% chlorine, 5 min of 0.55% ortho-phthalaldehyde, 10 min of 2% glutaraldehyde, 0.2% peracetic acid, 0.9% alkaline cleanser, or 254 nm ultraviolet light. Few studies investigated the efficacy of some disinfectants on Mpox. For example, one study found that in 3 min, copper with a purity of 99.9% was as effective against the vaccinia virus as it was against MPXV (Bleichert et al., 2014). Another study on the hand rub formulations and alcohols recommended by the WHO found that, despite these disinfectants being effective, MPXV showed the greatest stability compared with other enveloped viruses (Meister et al., 2023). It is crucial to note that the efficacy of a disinfectant against MPXV can be influenced by parameters such as the kind and amount of organic material present, disinfectant concentration and contact time, and environmental temperature and humidity (CDC, 2008; Russell, 2008; Campagna et al., 2016). However, we recommend validating disinfection protocols using MPXV to ensure their effectiveness in real-world settings.

4.2 Treatment of wastewater

Treatment of wastewater is a crucial step to avoid the further spread of diseases. Conventional treatment plants consist of a primary sedimentation step, secondary aerobic treatment and chlorination before reuse and discharge. The survival of viruses is mainly dependent on temperature; high temperatures denature proteins and increase enzymatic activity, decreasing viral survival rates (Gerba et al., 1980; John and Rose, 2005). Solvents and detergents also play a role as they remove viral envelopes. Anaerobic treatment denatures proteins and nucleic acids (Gundy et al., 2009; Sassi et al., 2018).

Removal of viruses is usually greater in the secondary treatment plant; this is attributed to sludge and suspended solids in the secondary effluent to which viruses attach. Viruses are mainly adsorbed in the second treatment (John and Rose, 2005). Viral inactivation especially takes place in tertiary treatment, which is the most crucial step to stop the transmission of infection. Advanced treatment is membrane separation, used in water reuse. It also plays a role in removing viruses from wastewater. A study revealed that using foul membranes for microfiltration removed more viruses than clean membranes (Farahbakhsh and Smith, 2004).

The currently implemented wastewater treatment systems are adequate at removing MPXV from wastewater and helping prevent the spread of infection. However, further studies may be needed to provide more information regarding the optimal temperature and best-suited disinfectant to rid wastewater from MPXV completely.

5 Conclusion

Current evidence demonstrates the contamination of surfaces by infected persons with mpox that has been proven to be stable and viable in the environment, surviving for a varying time depending on factors that impact virus survival, such as the contaminated surface, humidity, and temperature. This is to be emphasized that positive qPCR results do not necessarily means viral infectivity. It is essential to keep in mind that indirect transmission of mpox through contaminated surfaces may present a significant issue for international health organizations by highlighting the risks of long-distance transmission, bringing back the virus into an area that has attained regional elimination with extended outbreak duration. Due to asymptomatic cases, it is currently more challenging to identify the infection's origin and spread. Thus, personal protective measures, such as hand washing and regular disinfection, should reduce environmental contamination and the possibility of virus transmission.

Wastewater surveillance is imperative in developing systems for efficient, timely control of diseases. However, poor urban planning structures, growing populations, and transportation challenges worsen the wastewater treatment problem in these countries. As a result, appropriate wastewater treatment services must be established.

In summary, focusing on interventions related to the highest-risk-associated settings for mpox infection spread and transmission is an effective strategy for preventing and reducing the risk of mpox spread. Furthermore, our findings could form the base for future research into mpox transmission and persistence in the environment. It assists in preventing and reducing the negative consequences of this disease. Further research is needed to fill in knowledge gaps, to help understand the persistence of the virus on surfaces in healthcare settings and the environment, viral shedding and DNA levels in different body fluids and skin lesions, and the persistence of MPXV DNA and its infectivity in wastewater. Improvements in the analytical methods used to detect MPXV DNA in sewage are also needed.

Author contributions

AT: Conceptualization, Writing – original draft, Writing – review & editing. BK: Conceptualization, Writing – original draft, Writing – review & editing. AH: Conceptualization, Writing – original draft, Writing – review & editing. OA: Conceptualization, Writing – original draft, Writing – review & editing. SR: Conceptualization, Writing – original draft, Writing – review & editing. DN: Conceptualization, Writing – original draft, Writing – review & editing. KS-C: Writing – review & editing. AR-M: Writing – review & editing. AM: Writing – review & editing. DB-A: Writing – review & editing. RS: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. APCs for this article were covered by Universidad Continental, Huancayo, Peru.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Adler, H., Gould, S., Hine, P., Snell, L. B., Wong, W., Houlihan, C. F., et al. (2022). Clinical features and management of human monkeypox: a retrospective observational study in the UK. *Lancet Infect. Dis.* 22, 1153–1162. doi: 10.1016/S1473-3099(22)00228-6

Americo, J. L., Earl, P. L., and Moss, B. (2023). Virulence differences of mpox (monkeypox) virus clades I, IIa, and IIb. 1 in a small animal model. *Proc. Natl. Acad. Sci. U.S.A.* 120:e2220415120. doi: 10.1073/pnas.2220415120

Arita, I., Ježek, Z., Khodakevich, L., and Ruti, K. (1985). Human monkeypox: a newly emerged orthopoxvirus zoonosis in the tropical rain forests of Africa. *Am. J. Trop. Med. Hyg.* 34, 781–789. doi: 10.4269/ajtmh.1985.34.781

Atkinson, B., Burton, C., Pottage, T., Thompson, K. A., Ngabo, D., Crook, A., et al. (2022a). Infection-competent monkeypox virus contamination identified in domestic settings following an imported case of monkeypox into the UK. *Environ. Microbiol.* 24, 4561–4569. doi: 10.1111/1462-2920.16129

Atkinson, B., Gould, S., Spencer, A., Onianwa, O., Furneaux, J., Grieves, J., et al. (2022b). Monkeypox virus contamination in an office-based workplace environment. *J. Hosp. Infect.* 130, 141–143. doi: 10.1016/j.jhin.2022.08.009

Becker, B., Bischoff, B., Brill, F. H. H., Steinmann, E., and Steinmann, J. (2017). Virucidal efficacy of a sonicated hydrogen peroxide system (Trophon® EPR) following

European and German test methods. *GMS Hyg. Infect. Control* 12:Doc02. doi: 10.3205/dgkh000287

Beer, E. M., and Rao, V. B. (2019). A systematic review of the epidemiology of human monkeypox outbreaks and implications for outbreak strategy. *PLoS Negl. Trop. Dis.* 13:e0007791. doi: 10.1371/journal.pntd.0007791

Bleichert, P., Espírito Santo, C., Hanczaruk, M., Meyer, H., and Grass, G. (2014). Inactivation of bacterial and viral biothreat agents on metallic copper surfaces. *Biometals* 27, 1179–1189. doi: 10.1007/s10534-014-9781-0

Brown, K., and Leggat, P. A. (2016). Human monkeypox: current state of knowledge and implications for the future. *Trop. Med. Infect. Dis.* 1:8. doi: 10.3390/tropicalmed1010008

Campagna, M. V., Faure-Kumar, E., Treger, J. A., Cushman, J. D., Grogan, T. R., Kasahara, N., et al. (2016). Factors in the selection of surface disinfectants for use in a laboratory animal setting. *J. Am. Assoc. Lab. Anim. Sci.* 55, 175–188. Available at: https://pubmed.ncbi.nlm.nih.gov/27025810/

CDC (2003a). Update: multistate outbreak of monkeypox—Illinois, Indiana, Kansas, Missouri, Ohio, and Wisconsin, 2003. *MMWR Morb. Mortal. Wkly Rep.* 52, 561–564. Available at: www.cdc.gov

CDC. *Environmental infection control guidelines CDC.* (2003b). Available at: www.cdc.gov

CDC. *Efficacy disinfection & sterilization guidelines guidelines library infection control CDC.* (2008). Available at: www.cdc.gov

CDC. (2022) *2022 Mpox outbreak global map*. Available at: www.cdc.gov

CDC. *How it spreads Mpox poxvirus CDC.* (2023a). Available at: www.cdc.gov

CDC. *Science brief: detection and transmission of Mpox (formerly monkeypox) virus during the 2022 clade IIb outbreak.* (2023b). Available at: www.cdc.gov

CDC. *Clinical recognition Mpox poxvirus CDC.* (2023c). Available at: www.cdc.gov

CDC. *If you're a close contact Mpox poxvirus CDC.* (2023d). Available at: www.cdc.gov

de Oliveira, T. M. L., Rehfeld, I. S., Guedes, M. I. M. C., Ferreira, J. M. S., and Kroon, E. G. (2011). Lobato ZIP susceptibility of vaccinia virus to chemical disinfectants. *Am. J. Trop. Med. Hyg.* 85, 152–157. doi: 10.4269/ajtmh.2011.11-0144

Di Giulio, D. B., and Eckburg, P. B. (2004). Human monkeypox: an emerging zoonosis. *Lancet Infect. Dis.* 4, 15–25. doi: 10.1016/S1473-3099(03)00856-9

Dye, C., and Kraemer, M. U. G. (2022). Investigating the monkeypox outbreak: Here's what we know, and what we need to know. *BMJ*, 19–20. doi: 10.1136/bmj.o1314

Eggers, M., Eickmann, M., and Zorn, J. (2015). Rapid and effective virucidal activity of povidone-iodine products against Middle East respiratory syndrome coronavirus (MERS-CoV) and modified vaccinia virus Ankara (MVA). *Infect. Dis. Ther.* 4, 491–501. doi: 10.1007/s40121-015-0091-9

Eterpi, M., McDonnell, G., and Thomas, V. (2009). Disinfection efficacy against parvoviruses compared with reference viruses. *J. Hosp. Infect.* 73, 64–70. doi: 10.1016/j.jhin.2009.05.016

Farahbakhsh, K., and Smith, D. W. (2004). Removal of coliphages in secondary effluent by microfiltration—mechanisms of removal and impact of operating parameters. *Water Res.* 38, 585–592. doi: 10.1016/j.watres.2003.10.018

Ferré, V. M., Bachelard, A., Zaidi, M., Armand-Lefevre, L., Descamps, D., Charpentier, C., et al. (2022). Detection of monkeypox virus in anorectal swabs from asymptomatic men who have sex with men in a sexually transmitted infection screening program in Paris, France. *Ann. Intern. Med.* 175, 1491–1492. doi: 10.7326/M22-2183

Gammon, D. B., Gowrishankar, B., Duraffour, S., Andrei, G., Upton, C., and Evans, D. H. (2010). Vaccinia virus-encoded ribonucleotide reductase subunits are differentially required for replication and pathogenesis. *PLoS Pathog.* 6:e1000984. doi: 10.1371/journal.ppat.1000984

Gazecka, M., Sniezek, J., Maciolek, K., Kowala-Piaskowska, A., and Zmora, P. (2023). Mpox virus detection in the wastewater and the number of hospitalized patients in the Poznan metropolitan area, Poland. *Int. J. Infect. Dis.* 133, 75–77. doi: 10.1016/j.ijid.2023.05.014

Gerba, C. P., Goyal, S. M., Hurst, C. J., and Labelle, R. L. (1980). Type and strain dependence of enterovirus adsorption to activated sludge, soils and estuarine sediments. *Water Res.* 14, 1197–1198. doi: 10.1016/0043-1354(80)90176-1

Girón-Guzmán, I., Díaz-Reolid, A., Truchado, P., Carcereny, A., García-Pedemonte, D., Hernáez, B., et al. (2023). Spanish wastewater reveals the current spread of monkeypox virus. *Water Res.* 231:119621. doi: 10.1016/j.watres.2023.119621

Goldmann, D. A. (2000). Transmission of viral respiratory infections in the home. *Pediatr. Infect. Dis. J.* 19, S97–S102. doi: 10.1097/00006454-200010001-00002

Gould, S., Atkinson, B., Onianwa, O., Spencer, A., Furneaux, J., Grieves, J., et al. Air and surface sampling for monkeypox virus in UK hospitals. *medRxiv*. (2022): Available at: https://doi.org/10.1101/2022.07.21.22277864. [Epub ahead of preprint]

Gundy, P. M., Gerba, C. P., and Pepper, I. L. (2009). Survival of Coronaviruses in Water and Wastewater. *Food Environ. Virol.* 1, 10–14. doi: 10.1007/s12560-008-9001-6

Hernaez, B., Muñoz-Gómez, A., Sanchiz, A., Orviz, E., Valls-Carbo, A., Sagastagoitia, I., et al. (2023). Monitoring monkeypox virus in saliva and air samples in Spain: a cross-sectional study. *Lancet Microbe* 4, e21–e28. doi: 10.1016/S2666-5247(22)00291-9

Hutson, C. L., Olson, V. A., Carroll, D. S., Abel, J. A., Hughes, C. M., Braden, Z. H., et al. (2009). A prairie dog animal model of systemic orthopoxvirus disease using west African and Congo basin strains of monkeypox virus. *J. Gen. Virol.* 90, 323–333. doi: 10.1099/vir.0.005108-0

Isidro, J., Borges, V., Pinto, M., Sobral, D., Santos, J. D., Nunes, A., et al. (2022). Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus. *Nat. Med.* 28, 1569–1572. doi: 10.1038/s41591-022-01907-y

Ježek, Z., Arita, I., Mutombo, M., Dunn, C., Nakano, J. H., and Szczeniowski, M. (1986). Four generations of probable person-to-person transmission of human monkeypox. *Am. J. Epidemiol.* 123, 1004–1012. doi: 10.1093/oxfordjournals.aje.a114328

Ježek, Z., Grab, B. S., Szczeniowski, M. V., Paluku, K. M., and Mutombo, M. (1988a). Human monkeypox: secondary attack rates. *Bull. World Health Organ.* 66, 465–470.

Ježek, Z., Szczeniowski, M., Paluku, K., and Mutombo, M. (1987). Human monkeypox: clinical features of 282 patients. *J. Infect. Dis.* 156, 293–298. doi: 10.1093/infdis/156.2.293

Ježek, Z., Szczeniowski, M. P., Paluku, K. M., Mutombo, M., and Grab, B. (1988b). Human monkeypox: confusion with chickenpox. *Acta Trop.* 45, 297–307.

John, D. E., and Rose, J. B. (2005). Review of factors affecting microbial survival in groundwater. *Environ. Sci. Technol.* 39, 7345–7356. doi: 10.1021/es047995w

Jonge, E. F. D., Peterse, C. M., Koelewijn, J. M., van der Drift, A. R., van der Beek, R. F. H. J., Nagelkerke, E., et al. (2022). Science of the Total Environment The detection of monkeypox virus DNA in wastewater samples in the Netherlands. *Sci. Total Environ.* 852:158265. doi: 10.1016/j.scitotenv.2022.158265

Kaler, J., Hussain, A., Flores, G., Kheiri, S., and Desrosiers, D. (2022). Monkeypox: a comprehensive review of transmission, pathogenesis, and manifestation. *Cureus* 14, e26531–e26511. doi: 10.7759/cureus.26531

Kampf, G., Rudolf, M., Labadie, J. C., and Barrett, S. P. (2002). Spectrum of antimicrobial activity and user acceptability of the hand disinfectant agent Sterillium® Gel. *J. Hosp. Infect.* 52, 141–147. doi: 10.1053/jhin.2002.1281

Kampf, G., Steinmann, J., and Rabenau, H. (2007). Suitability of vaccinia virus and bovine viral diarrhea virus (BVDV) for determining activities of three commonly-used alcohol-based hand rubs against enveloped viruses. *BMC Infect. Dis.* 7, 1–6. doi: 10.1186/1471-2334-7-5

Kannan, S. R., Sachdev, S., Reddy, A. S., Kandasamy, S. L., Byrareddy, S. N., Lorson, C. L., et al. (2022). Mutations in the monkeypox virus replication complex: potential contributing factors to the 2022 outbreak. *J. Autoimmun.* 133:102928. doi: 10.1016/j.jaut.2022.102928

Karagoz, A., Tombuloglu, H., Alsaeed, M., Tombuloglu, G., AlRubaish, A. A., Mahmoud, A., et al. (2023). Monkeypox (mpox) virus: classification, origin, transmission, genome organization, antiviral drugs, and molecular diagnosis. *J. Infect. Public Health* 16, 531–541. doi: 10.1016/j.jiph.2023.02.003

Kipkorir, V., Dhali, A., Srichawla, B., Kutikuppala, S., Cox, M., Ochieng, D., et al. (2022). The re-emerging monkeypox disease. *Trop. Med. Int. Health* 27, 961–969. doi: 10.1111/tmi.13821

Klein, M. (1963). Antiviral action of germicides. *Soap Chem. Spec.* 39, 70–72.

Kozlov, M. (2022). How does monkeypox spread? What scientists know. *Nature* 608, 655–656. doi: 10.1038/d41586-022-02178-w

La Rosa, G., Mancini, P., Veneri, C., Ferraro, G. B., Lucentini, L., Iaconelli, M., et al. (2023). Detection of monkeypox virus DNA in airport Wastewater, Rome. *Emerg. Infect. Dis.* 29, 193–196. doi: 10.3201/eid2901.221311

Ladnyj, I. D., Ziegler, P., and Kima, E. (1972). A human infection caused by monkeypox virus in Basankusu Territory, Democratic Republic of the Congo. *Bull. World Health Organ.* 46, 593–597.

Learned, L. A., Wassa, M. G., Li, D. W., Olson, Y., Karem, V. A., Stempora, K., et al. (2005). Extended interhuman transmission of monkeypox in a hospital community in the Republic of the Congo, 2003. *Am. J. Trop. Med. Hyg.* 73, 428–434. doi: 10.4269/ajtmh.2005.73.428

León-Figueroa, D. A., Barboza, J. J., Garcia-Vasquez, E. A., Bonilla-Aldana, D. K., Diaz-Torres, M., Saldaña-Cumpa, H. M., et al. (2022). Epidemiological situation of monkeypox transmission by possible sexual contact: a systematic review. *Trop. Med. Infect. Dis.* 7. doi: 10.3390/tropicalmed7100267

Lu, T., Wu, Z., Jiang, S., Lu, L., and Liu, H. (2022). The current emergence of monkeypox: the recurrence of another smallpox? *Biosaf. Health* 4, 369–375. doi: 10.1016/j.bsheal.2022.09.004

Mahase, E. (2022). Seven monkeypox cases are confirmed in England. *BMJ* 377:o1239. doi: 10.1136/bmj.o1239

McDevitt, J. J., Lai, K. M., Rudnick, S. N., Houseman, E. A., First, M. W., and Milton, D. K. (2007). Characterization of UVC light sensitivity of vaccinia virus. *Appl. Environ. Microbiol.* 73, 5760–5766. doi: 10.1128/AEM.00110-07

McDevitt, J. J., Milton, D. K., Rudnick, S. N., and First, M. W. (2008). Inactivation of poxviruses by upper-room UVC light in a simulated hospital room environment. *PLoS One* 3:e3186. doi: 10.1371/journal.pone.0003186

Meister, T. L., Tao, R., Brüggemann, Y., Todt, D., Steinmann, J., Timm, J., et al. (2023). Efficient inactivation of monkeypox virus by World Health Organization–recommended hand rub formulations and alcohols. *Emerg. Infect. Dis.* 29, 189–192. doi: 10.3201/eid2901.221429

Mellon, G., Rubenstein, E., Antoine, M., Ferré, V. M., Gabassi, A., Molina, J. M., et al. (2023). Air detection of monkeypox virus in a dedicated outpatient clinic room for monkeypox infection diagnosis. *J. Infect.* 86, 256–308. doi: 10.1016/j.jinf.2022.12.025

Moore, M., and Zahra, F. (2022). "Monkeypox" in *Stat Pearls* (Treasur Island (FL): Stat Pearls Publishing)

Morgan, C. N., Whitehill, F., Doty, J. B., Schulte, J., Matheny, A., Stringer, J., et al. (2022). Environmental persistence of monkeypox virus on surfaces in household of person with travel-associated infection, Dallas, Texas, USA, 2021. *Emerg. Infect. Dis.* 28, 1982–1989. doi: 10.3201/eid2810.221047

Nakoune, E., Lampaert, E., Ndjapou, S. G., Janssens, C., Zuniga, I., Van Herp, M., et al. (2017). A nosocomial outbreak of human monkeypox in the central African Republic. *Open Forum Infect. Dis.* 4, 1–4. doi: 10.1093/ofid/ofx168

Nörz, D., Pfefferle, S., Brehm, T. T., Franke, G., Grewe, I., Knobling, B., et al. (2022). Evidence of surface contamination in hospital rooms occupied by patients infected with monkeypox. *Euro Surveill.* 27, 1–6. doi: 10.2807/1560-7917.ES.2022.27.26.2200477

Ogoina, D., Izibewule, J. H., Ogunleye, A., Ederiane, E., Anebonam, U., Neni, A., et al. (2019). The 2017 human monkeypox outbreak in Nigeria-report of outbreak experience and response in the Niger Delta University Teaching Hospital, Bayelsa State, Nigeria. *PLoS One* 14:e0214229. doi: 10.1371/journal.pone.0214229

Ogoina, D., and Ogunsola, F. T. (2022). Monkeypox and the healthcare environment. *Lancet Microbe* 3, e889–e890. doi: 10.1016/S2666-5247(22)00286-5

Ogoina, D., and Yinka-Ogunleye, A. (2022). Sexual history of human monkeypox patients seen at a tertiary hospital in Bayelsa, Nigeria. *Int. J. STD AIDS* 33, 928–932. doi: 10.1177/09564624221119335

Owhonda, G., Wali, I. A., Nwadiuto, I., Abikor, V., Solomon, L., Luke, A., et al. (2023). Monkeypox disease with predominant genital symptoms: A Nigerian case report. *Arch. Curr. Res. Int.* 23, 36–42. doi: 10.9734/acri/2023/v23i2557

Palich, R., Burrel, S., Monsel, G., Nouchi, A., Bleibtreu, A., Seang, S., et al. (2022). Viral loads in clinical samples of men with monkeypox virus infection: a French case series. *Lancet Infect. Dis.* 23, 74–80. doi: 10.1016/S1473-3099(22)00586-2

Patrono, L. V., Pléh, K., Samuni, L., Ulrich, M., Röthemeier, C., Sachse, A., et al. (2020). Monkeypox virus emergence in wild chimpanzees reveals distinct clinical outcomes and viral diversity. *Nat. Microbiol.* 5, 955–965. doi: 10.1038/s41564-020-0706-0

Peiró-Mestres, A., Fuertes, I., Camprubí-Ferrer, D., Marcos, M., Vilella, A., Navarro, M., et al. (2022). Frequent detection of monkeypox virus DNA in saliva, semen, and other clinical samples from 12 patients, Barcelona, Spain, May to June 2022. *Euro Surveill.* 27. doi: 10.2807/1560-7917.ES.2022.27.28.2200503

Petersen, B. W., Harms, T. J., Reynolds, M. G., and Harrison, L. H. (2016). Use of vaccinia virus smallpox vaccine in laboratory and health care personnel at risk for occupational exposure to orthopoxviruses—ecommendations of the Advisory Committee on Immunization Practices (ACIP), 2015. *MMWR Morb. Mortal. Wkly Rep.* 65, 257–262. doi: 10.15585/mmwr.mm6510a2

Pfeiffer, J. A., Collingwood, A., Rider, L. E., Minhaj, F. S., Matheny, A. M., Kling, C., et al. (2022). High-contact object and surface contamination in a household of persons with monkeypox virus. Infection-Utah, June 2022. *MMWR Morb. Mortal. Wkly Rep.* 71, 1092–1094. doi: 10.15585/mmwr.mm7134e1

Rabenau, H. F., Rapp, I., and Steinmann, J. (2010). Can vaccinia virus be replaced by MVA virus for testing virucidal activity of chemical disinfectants? *BMC Infect. Dis.* 10, 1–8. doi: 10.1186/1471-2334-10-185

Riopelle, J. C., Munster, V. J., and Port, J. R.. *Atypical and unique transmission of monkeypox virus during the 2022 outbreak: an overview of the current state of knowledge.* (2022).

Russell, A. D. (2008). "Factors influencing the efficacy of antimicrobial agents" in *Principles and practice of disinfection, preservation and sterilization* (London: Blackwell), 98–127.

Salvato, R. S., Rodrigues Ikeda, M. L., Barcellos, R. B., Godinho, F. M., Sesterheim, P., Bitencourt, L. C. B., et al. (2022). Possible occupational infection of healthcare workers with monkeypox virus, Brazil. *Emerg. Infect. Dis.* 28, 2520–2523. doi: 10.3201/eid2812.221343

Sassi, H. P., Ikner, L. A., Abd-Elmaksoud, S., Gerba, C. P., and Pepper, I. L. (2018). Comparative survival of viruses during thermophilic and mesophilic anaerobic digestion. *Sci. Total Environ.* 615, 15–19. doi: 10.1016/j.scitotenv.2017.09.205

Sharkey, M. E., Babler, K. M., Abelson, S. M., Alsuliman, B., Amirali, A., Comerford, S., et al. First detection of the monkeypox virus using wastewater-based surveillance in Miami-Dade County. (2022), Available at: http://dx.doi.org/10.2139/ssrn.4236277

Sharkey, M. E., Babler, K. M., Shukla, B. S., Abelson, S. M., Alsuliman, B., Amirali, A., et al. (2023). Monkeypox viral nucleic acids detected using both DNA and RNA extraction workflows. *Sci. Total Environ.* 890:164289. doi: 10.1016/j.scitotenv.2023.164289

Siddharta, A., Pfaender, S., Vielle, N. J., Dijkman, R., Friesland, M., Becker, B., et al. (2017). Virucidal activity of World Health Organization–recommended formulations against enveloped viruses, including Zika, Ebola, and emerging coronaviruses. *J. Infect. Dis.* 215, 902–906. doi: 10.1093/infdis/jix046

Spicknall, I. H., Koopman, J. S., Nicas, M., Pujol, J. M., Li, S., and Eisenberg, J. N. (2010). Informing optimal environmental influenza interventions: how the host, agent, and environment alter dominant routes of transmission. *PLoS Comput. Biol.* 6:e1000969-e. doi: 10.1371/journal.pcbi.1000969

Steinmann, J., Paulmann, D., Becker, B., Bischoff, B., and Steinmann, E. (2012). Comparison of virucidal activity of alcohol-based hand sanitizers versus antimicrobial hand soaps *in vitro* and *in vivo*. *J. Hosp. Infect.* 82, 277–280. doi: 10.1016/j.jhin.2012.08.005

Thornhill, J. P., Barkati, S., Walmsley, S., Rockstroh, J., Antinori, A., Harrison, L. B., et al. (2022a). Monkeypox virus infection in humans across 16 countries—April–June 2022. *N. Engl. J. Med.* 387, 679–691. doi: 10.1056/NEJMoa2207323

Thornhill, J. P., Palich, R., Ghosn, J., Walmsley, S., Moschese, D., Cortes, C. P., et al. (2022b). Human monkeypox virus infection in women and non-binary individuals during the 2022 outbreaks: a global case series. *Lancet* 400, 1953–1965. doi: 10.1016/S0140-6736(22)02187-0

Tiwari, A., Adhikari, S., Kaya, D., Islam, A., Malla, B., Sherchan, S. P., et al. (2023). Science of the total environment monkeypox outbreak: wastewater and environmental surveillance perspective. *Sci. Total Environ.* 856:159166. doi: 10.1016/j.scitotenv.2022.159166

Upadhayay, S., Arthur, R., Soni, D., Yadav, P., Navik, U., Singh, R., et al. (2022). Monkeypox infection: the past, present, and future. *Int. Immunopharmacol.* 113:109382. doi: 10.1016/j.intimp.2022.109382

Vaughan, A., Aarons, E., Astbury, J., Brooks, T., Chand, M., Flegg, P., et al. (2020). Human-to-human transmission of monkeypox virus, United Kingdom, October 2018. *Emerg. Infect. Dis.* 26, 782–785. doi: 10.3201/eid2604.191164

Wannigama, D. L., Amarasiri, M., Hongsing, P., Hurst, C., Modchang, C., Chadsuthi, S., et al. (2023). Multiple traces of monkeypox detected in non-sewered wastewater with sparse sampling from a densely populated metropolitan area in Asia. *Sci. Total Environ.* 858:159816. doi: 10.1016/j.scitotenv.2022.159816

Wieder-Feinsod, A., Zilberman, T., Erster, O., Kolasko, G. W., Biber, A., Gophen, R., et al. (2023). Overlooked monkeypox cases among men having sex with men during the 2022 outbreak—a retrospective study. *Int. J. Infect. Dis.* 128, 58–60. doi: 10.1016/j.ijid.2022.12.014

Wurtzer, S., Levert, M., Dhenain, E., Boni, M., Tournier, J. N., Londinsky, N., et al. (2022). First detection of monkeypox virus genome in sewersheds in France: the potential of wastewater-based epidemiology for monitoring emerging disease. *Environ. Sci. Technol. Lett.* 9, 991–996. doi: 10.1021/acs.estlett.2c00693

Wutzler, P., and Sauerbrei, A. (2004). Virucidal activity of the new disinfectant monopercitric acid. *Lett. Appl. Microbiol.* 39, 194–198. doi: 10.1111/j.1472-765X.2004.01561.x

Zachary, K. C., and Shenoy, E. S. (2022). Monkeypox transmission following exposure in healthcare facilities in nonendemic settings: low risk but limited literature. *Infect. Control Hosp. Epidemiol.* 43, 920–924. doi: 10.1017/ice.2022.152

Glossary

qPCR	Quantitative polymerase chain reaction
WBS	Wastewater-based surveillance
DNA	Double-stranded nucleic acid
RNA	Ribonucleic acid
MPXV	Monkeypox virus
RT-PCR	Reverse transcription polymerase chain reaction
UV	Ultraviolet
CDC	Centers for Disease Control and Prevention
UK	United Kingdom
STD	Sexually transmitted diseases
MVA	Modified vaccinia virus Ankara
FCS	Fetal calf serum
B2M	Beta-2-microglobulin
V2G	Volcano 2nd generation
PPE	Personal protective equipment
MSM	Men who have sex with men
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 15 January 2024
DOI 10.3389/fmicb.2023.1279632

OPEN ACCESS

EDITED BY
Shailendra Saxena,
King George's Medical University, India

REVIEWED BY
Van Thuan Hoang,
Thai Binh University of Medicine and Pharmacy, Vietnam
Kamal Chandima Jeewandara,
University of Sri Jayewardenepura, Sri Lanka

*CORRESPONDENCE
Yogendra Pratap Mathuria
✉ ypm.1702@yahoo.com
Shailesh Kumar Gupta
✉ shaileshgupta0603@gmail.com

RECEIVED 18 August 2023
ACCEPTED 26 December 2023
PUBLISHED 15 January 2024

CITATION
Negi S, Diksha, Kalita D, Ranakoti N, Negi A, Kandwal D, Gupta SK and Mathuria YP (2024) Trend of viral load during the first, second, and third wave of COVID-19 in the Indian Himalayan region: an observational study of the Uttarakhand state.
Front. Microbiol. 14:1279632.
doi: 10.3389/fmicb.2023.1279632

Trend of viral load during the first, second, and third wave of COVID-19 in the Indian Himalayan region: an observational study of the Uttarakhand state

Shailender Negi[1], Diksha[1], Deepjyoti Kalita[2], Neeraj Ranakoti[1], Ashish Negi[1], Diksha Kandwal[1], Shailesh Kumar Gupta[1]* and Yogendra Pratap Mathuria[1,3]*

[1]Viral Research and Diagnostic Laboratory (VRDL), All India Institute of Medical Sciences, Rishikesh, India, [2]Department of Microbiology, All India Institute of Medical Sciences, Guwahati, India, [3]Department of Microbiology, All India Institute of Medical Sciences, Rishikesh, India

India had faced three waves throughout the Coronavirus disease 2019 (COVID-19) pandemic, which had already impacted economic lives and affected the healthcare setting and infrastructure. The widespread impacts have inspired researchers to look for clinical indicators of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection prognosis. Cyclic threshold values have been used to correlate the viral load in COVID-19 patients and for viral transmission. In light of this correlation, a retrospective study was conducted to assess the trend of viral load in clinical and demographic profiles across the three waves. Data of a total of 11,125 COVID-19-positive patients were obtained, which had a Ct value of <35. We stratified Ct values as follows: under 25 (high viral load), 25–30 (moderate viral load), and over 30 (low viral load). We found a significantly high proportion of patients with high viral load during the second wave. A significantly high viral load across the symptomatic and vaccinated populations was found in all three waves, whereas a significantly high viral load across age groups was found only in the first wave. With the widespread availability of real-time PCR and the limited use of genomic surveillance, the Ct value and viral load could be a suitable tool for population-level monitoring and forecasting.

KEYWORDS

COVID-19, SARS-CoV-2, cycle threshold, viral load, trend

1 Introduction

Coronavirus disease 2019 (COVID-19) is one of the deadliest infectious diseases caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that has affected millions of lives worldwide (Seok et al., 2019). The WHO declared the novel coronavirus outbreak a public health emergency of international concern (PHEIC) on 30 January 2020 and a pandemic on 11 March 2020, raising it to its highest level of alarm. As of 5 July 2023, a total of 0.76 billion COVID-19 cases and 6.9 million deaths have been reported worldwide (WHO Coronavirus

(COVID-19), 2023). India has reported 44 million total cases of COVID-19 with half a million deaths (WHO Coronavirus Disease, 2023). Moreover, India has witnessed three massive waves: the first wave began in March 2020 and lasted till November 2020, the second wave ranged from March 2021 to May 2021, and the third wave ranged from January 2022 to March 2022 (Zirpe et al., 2021; Jayadevan et al., 2022). Among all three waves, the second wave appeared to be more deadly, with limited availability of vital therapies along with escalating cases and deaths accompanied by a shortage of hospital beds and oxygen (Asrani et al., 2021). The vast effects of the COVID-19 pandemic on healthcare systems have inspired researchers to look for clinical indicators of SARS-CoV-2 infection prognosis (Martínez et al., 2022). Cycle threshold (Ct) values in real-time reverse transcription–polymerase chain reaction (rRT-PCR) have been used to quantify the amount required for the target viral gene to pass the threshold and are inversely correlated with viral load (Tom and Mina, 2020). Several investigations have revealed a strong association between COVID-19 severity and disease progression and lower Ct values, which indicate increased viral load (Rao et al., 2020). The most common targets for RT-PCR coronavirus detection are conserved or highly expressed genes, such as structural spike glycoprotein (S) and nucleocapsid protein (N) genes, as well as non-structural RdRp and replicase open reading frame (ORF) 1ab genes (Cui et al., 2019). A correlation between viral load and Ct value was also assessed in influenza and respiratory syncytial (RSV) virus, but a significant relationship was observed only in the RSV (Reina et al., 2018). Similarly, the Ct value of SARS-CoV-2 was used to evaluate the viral load in the adult and elderly populations and to determine the role of these populations in virus transmission (Mishra et al., 2022). In view of this concept, the present study aimed to assess the viral load across the demographic and clinical profiles during the first, second, and third waves in India.

2 Materials and methods

2.1 Study design and study site

The present study used a retrospective cross-sectional design for evaluating Ct values of SARS-CoV-2-positive samples tested in the Viral Research and Diagnostic Laboratory (VRDL), Department of Microbiology, AIIMS, Rishikesh, during the first, second, and third waves.

2.2 Data source

We used the ICMR COVID-19 data portal for extracting SARS-CoV-2-positive data during the first, second, and third waves. The Ct values of the SARS-CoV-2-positive samples were obtained from the RT-PCR in the BIO-RAD CFX 96TM system (Bio-Rad Laboratories, Inc., USA) using COVIDsure Multiplex Realtime RT-PCR Kit (Trivitron Healthcare Pvt. Ltd.). The viral RNA isolation from nasopharyngeal/oropharyngeal swabs was performed by an automated nucleic acid extraction system (KingFisher™ Flex system) (Thermo Fisher Scientific, USA) using MagMAX Viral/Pathogen Nucleic Acid Isolation Kit (Thermo Fisher Scientific, USA).

The VRDL, AIIMS, Rishikesh, has been authorized by the ICMR for the testing and reporting of SARS-CoV-2 samples in the ICMR COVID-19 data portal. The extracted data had a unique identifier for each sample, SRF number, generated at the time of sample collection and must be required during reporting. In addition, the data comprised demographic, clinical, hospitalization, vaccination, and reporting parameters. In the present study, we stratified the Ct values as follows: under 25 (high viral load), 25–30 (moderate viral load), and over 30 (low viral load). In addition, we categorized age into three groups: under 18 years (young), 18–60 years (adult), and over 60 years (elderly).

2.3 Study sample and sample size

Samples tested positive ($Ct < 35$) for SARS-CoV-2 and had at least one Ct value reported among the two confirmatory genes, ORF or RdRp, from various districts of Uttarakhand and Uttar Pradesh during the three waves were included in the study. The mean Ct value was obtained if both genes were reported. A total of 11,125 patients were found eligible for the study.

2.4 Statistical analysis

Statistical analyses were performed using Microsoft Excel, R (The R foundation), and GraphPad Prism. Descriptive analyses were carried out to determine the patient characteristics across the different COVID waves and were presented using frequency and proportion. The median and interquartile range were reported for Ct value after testing for normality using the Shapiro–Wilk test. The statistical comparison of the Ct values and viral load among patients' demographic, hospitalization, and clinical and vaccination statuses were made using the Mann–Whitney U tests, Kruskal–Wallis test, chi-square tests, or Fisher exact tests, as appropriate. The confidence interval was set at 95% and the significance level at 5%.

3 Results

3.1 Patient characteristics across the first, second, and third waves of the COVID-19 pandemic

In total, we gathered data from 11,125 COVID-positive patients, of which 6,030 suffered during the first wave, whereas 3,398 and 1,697 suffered during the second and third waves, respectively. In all three waves, adults were an overwhelming majority (78–87%), and the women-to-men ratio was approximately 2:3. There was a rapid increase in the community cases (59 to 97%), whereas the hospitalization rate declined (41 to 3%) with the subsequent wave. Overall, few patients had received vaccination (3.48%) and, approximately, three of five patients were asymptomatic (Table 1).

3.2 The trend of viral load across the three waves of the COVID-19 pandemic

The Shapiro–Wilk test of normality determined the skewness of Ct values. The median Ct value in the second wave was significantly low, 24 (interquartile range [IQR], 7) as compared to the first and third

TABLE 1 Patient characteristics in the first (March–November, 2020), second (March–May, 2021), and third (January–March, 2022) waves of the COVID-19 pandemic.

Characteristics	Wave 1 (*N* = 6,030)	Wave 2 (*N* = 3,398)	Wave 3 (*N* = 1,697)
	n (%)	n (%)	n (%)
Age groups			
≤18 years (young)	339 (5.6)	183 (5.4)	131 (7.7)
19–60 years (adult)	4,711 (78.1)	2,885 (84.9)	1,474 (86.9)
>60 years (elderly)	980 (16.3)	330 (9.7)	92 (5.4)
Sex			
Male	4,227 (70.1)	2089 (61.5)	1,011 (59.6)
Female	1803 (29.9)	1,309 (38.5)	686 (40.4)
Patient status			
Community	3,555 (59)	3,260 (96.3)	1,647 (97.1)
Hospitalized	2,469 (41)	126 (3.7)	50 (2.9)
Clinical status			
Asymptomatic	4,074 (67.6)	2,213 (65.1)	1,219 (71.8)
Symptomatic	1956 (32.4)	1,185 (34.9)	478 (28.2)
Vaccination status			
Non-vaccinated	6,030 (100)	3,062 (90.1)	1,646 (97)
Vaccinated	0 (0)	336 (9.9)	51 (3)
All values rounded up to 1 decimal			

waves, which was 26 (IQR, 10) and 26 (IQR, 6), respectively. In terms of viral load, we observed a significant trend that increased from the first wave to the second wave and then decreased in the third wave (Figure 1). The Indian SARS-CoV-2 Genomics Consortium (INSACOG) data also revealed the circulation of a more severe Delta variant (B.1.617) of SARS-CoV-2 in the second wave compared to B.1.1.529 (Omicron variant) in the third wave (INSACOG, 2023).

3.3 Distribution of viral load in different patient populations

In the first wave, the elderly population had a significantly low median Ct value (25) as compared to the young (26) and adult population (26); however, there was no significant difference observed in the second and third waves. Similarly, the viral load data of the first wave indicated a significantly high viral load (50.6%) in the elderly population compared to adults (48%) and the young population (47.2%). Although there was no significant difference observed in viral load across the second and third waves, the young population in the third wave had a high viral load (45.8%) as compared to the elderly (44.6%) and adult populations (42.3%) (Figure 2). We did not find any statistically significant difference in the Ct value and viral load across sex, but the viral load among women is higher in all three waves (Figure 3).

There was a significantly low median Ct value in the community population in the first and second waves; however, it was significantly low in the hospitalized patients in the third wave. The viral load data echo the Ct findings, as shown in Figure 4. The present study found out more hospital admissions even with low viral load in the first wave, indicating the severity as well as lack of preparedness during the first wave of the pandemic.

We found a significantly low median Ct value among the symptomatic population compared with the non-symptomatic population across the three waves (Figure 5). Similar results were obtained with viral load. The viral load data showed more symptomatic cases with low viral load during the first wave (21.8%) compared to the second (4.6%) and third (6.5%) waves. Our study found a higher proportion of vaccinated patients with significantly low Ct value and high viral load compared with non-vaccinated in the second and third waves. In addition, the lower proportion of the non-vaccinated population with high viral load in the third wave (42%) compared to the first (48.4%) and second (59%) waves indicated a decrease in the severity of the pandemic as well as the development of herd immunity (Figure 6).

4 Discussion

The best way to diagnose SARS-CoV-2 infection in the laboratory is the detection of viral RNA using a real-time reverse transcriptase PCR (rRT-PCR) assay from respiratory tract specimens. The single data points derived from real-time PCR amplification plots are called threshold cycles or Ct values. The Ct value has an inverse relationship with the amount of virus present in a particular sample. A low Ct value indicates a high level of genetic material, which is often associated with an increased level of infection (Rao et al., 2020). The Ct value is a semi-quantitative tool for the broad characterization of viral load as low, moderate, and high (Indian Journal of Medical Specialities, 2023). Ct value being a proxy of viral load can serve as an indicator for infectivity and severity of disease. Our study found a high proportion of the population with low Ct value and high viral load in the second wave compared to the first and third waves. In addition, Mishra et.al observed a high proportion of low Ct value in the second wave in positive samples (Mishra et al., 2022). In India, the sudden increase in cases in the second and third waves possibly indicates the unlocking of social activities as well as the transmission of a more infectious variant, B.1.617 (Delta variant) and B.1.1.529 (Omicron), respectively (Choudhary et al., 2021). The high transmissibility and infection rate of SARS-CoV-2 during the second wave led to viral spread in the asymptomatic population, especially those who stayed inside, contributing to an increase in positivity and death (Choudhary et al., 2021). Besides the high viral load in the second wave, we did not observe substantial hospitalization. Similar findings were observed in a comparative study from a tertiary care hospital in India, determining high mortality due to tachypnoea, hypoxia, and lack of an aggressive management system (ICU) could be the possible reason for the reduction in hospitalization percentage (Tendulkar et al., 2023).

Among various predictors of the severity and outcomes of COVID-19, age plays a significant role associated with an increased risk of the disease, leading to morbidity as well as mortality (Statsenko et al., 2022). A study by Hasanoglu et al. (2021) showed a statistically significant correlation between age and viral load. Few studies concluded the direct association between old age and high viral load and disease severity (Qian et al., 2020; To et al., 2020). Our findings revealed the severity of the first wave predominantly on the elderly population compared to the second and third waves. The possible

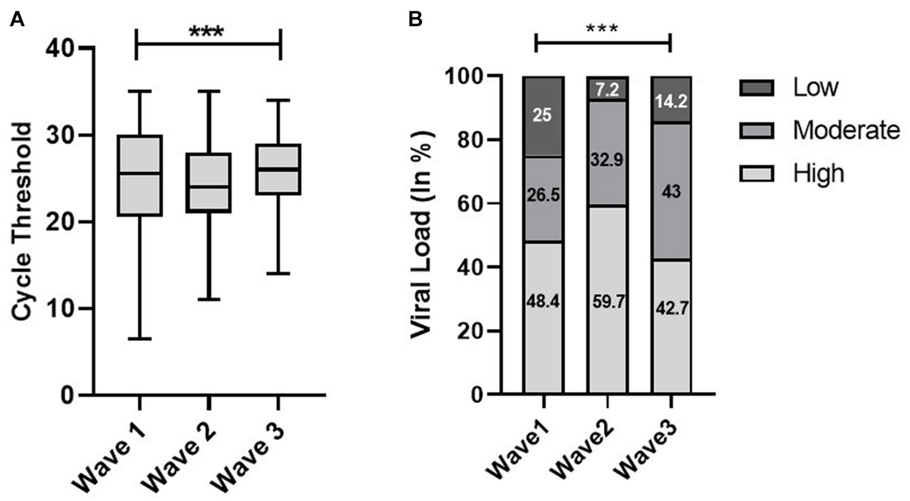


FIGURE 1
Distribution of cycle threshold (Ct) values **(A)** and viral load **(B)** across the first, second, and third waves of the COVID-19-positive cases: midlines indicate the median, boxes indicate interquartile ranges, and whiskers indicate the upper and lower adjacent values (within 1.5 times the interquartile range).

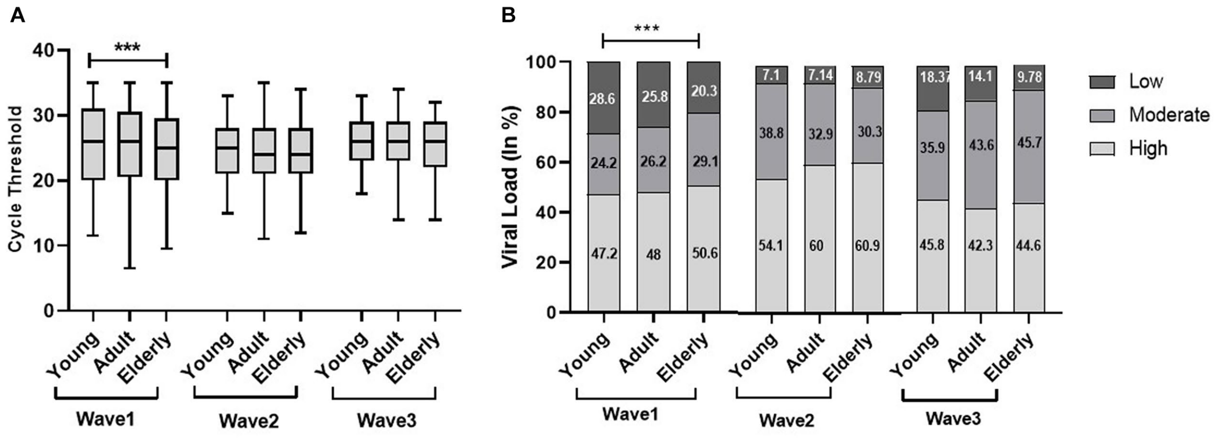


FIGURE 2
Age-specific comparisons of cycle threshold (Ct) values and viral load across the three COVID waves: midlines indicate the median, boxes indicate interquartile ranges, and whiskers indicate the upper and lower adjacent values (within 1.5 times the interquartile range).

reason could be age-related comorbidities, especially acute respiratory distress syndrome (ARDS) in the elderly population (Chen et al., 2020). The corresponding results obtained by Sun et al. demonstrate the catastrophic effects of COVID-19 in elderly patients compared to young patients, with a shorter average duration from the onset of symptoms to death. These findings suggested the quick advancement of the disease in older people than in younger ones (Sun et al., 2020). Despite the severity of the pandemic, the younger population had milder infections during the first and second waves. Good inflammatory response and delayed development of angiotensin-converting enzyme 2 led to better outcomes in children with COVID-19. However, concern had been raised primarily for non-vaccinated children during the third wave, and our findings revealed a high proportion of the young population with high viral load during the third wave (Sun et al., 2020).

The effect of the SARS-CoV-2 virus generates various clinical responses ranging from symptomatic to asymptomatic and more severe infections that require critical care (Gülbudak et al., 2021). According to a report, 40–45% of the SARS-CoV-2 cases do not exhibit any symptoms. Initially, there were different opinions regarding the infectiousness of asymptomatic patients for the COVID-19 spread. During the evaluation of viral load from different samples, Hasanoglu et al. demonstrated high viral load from nasopharyngeal/oropharyngeal samples of asymptomatic COVID-19 patients than symptomatic patients and considered these asymptomatic cases as an invisible part of the iceberg (Hasanoglu et al., 2021). In contrast, numerous studies, including the current study, demonstrated low Ct value and high viral load in symptomatic patients compared to asymptomatic patients (Gülbudak et al., 2021; Kociolek et al., 2021; Strutner et al., 2021). Despite these parameters,

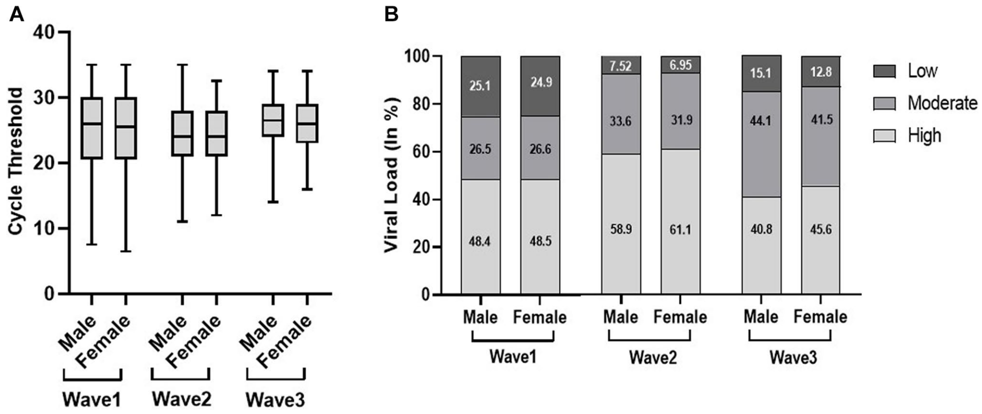


FIGURE 3
Distribution of cycle threshold (Ct) values **(A)** and viral load **(B)** in male and female patients across the three COVID waves: midlines indicate the median, boxes indicate interquartile ranges, and whiskers indicate the upper and lower adjacent values (within 1.5 times the interquartile range).

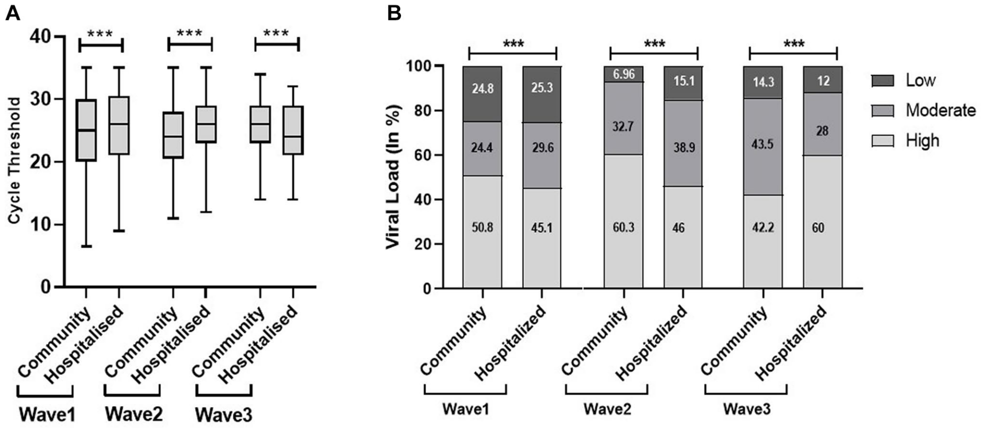


FIGURE 4
Cycle threshold (Ct) values **(A)** and viral load **(B)** of community and hospitalized patients across the three COVID waves: midlines indicate the median, boxes indicate interquartile ranges, and whiskers indicate the upper and lower adjacent values (within 1.5 times the interquartile range).

few researchers found no significant difference in Ct value among symptomatic and asymptomatic individuals (Lee et al., 2020; Louie et al., 2022). Therefore, there is a need for a longitudinal study for assessing the possible relationship between Ct value and symptoms.

One of the main reasons behind the successive waves, besides human influence, is the evolution of SARS-CoV-2. In India, the Delta and Omicron variants of SARS-CoV-2 were responsible for the second and third waves, respectively. It is a well-known fact that the Omicron variant has higher transmissibility and infectivity than Delta (ECDC) (European Centre for Disease Prevention and Control, 2023). Despite their higher transmissibility and infectivity, high mortality and less hospitalization had been observed during the second wave. The application of a low Ct value for the prediction of COVID-19 outcomes in hospitalized patients is considered a useful tool during the peak COVID wave (Kurzeder et al., 2022). One study compared low Ct value with poor disease outcomes in unvaccinated hospitalized patients, whereas another study utilized the Ct value (<26) as a risk score in predicting patient mortality along with other parameters in hospitalized patients of COVID-19 (Wright et al., 2021; Kurzeder et al., 2022). However, few studies limit its application to early warning indicators due to its marked variation in community patients, but one of the studies rejects the use of Ct value as a prognostic factor or a marker in community patients of COVID-19 (Martínez et al., 2022). The present study found a low Ct value among hospitalized COVID-19 patients during the third wave in our tertiary care hospital.

The best strategy to prevent from COVID-19 is the successful administration of vaccines as it minimizes diseases, acute illnesses, and deaths caused by the SARS-CoV-2. India started its COVID-19 vaccination program on 16 January 2021. Limited studies exploit the relationship of Ct value with the vaccination status of individuals. A study by Acharya et al. did not find any significant association between Ct value and vaccination status. Despite the vaccination program, we found a significantly high viral load in the vaccinated population compared with the non-vaccinated population in the second and third waves, which could be due to the fact that the newer strains most likely evade the immune

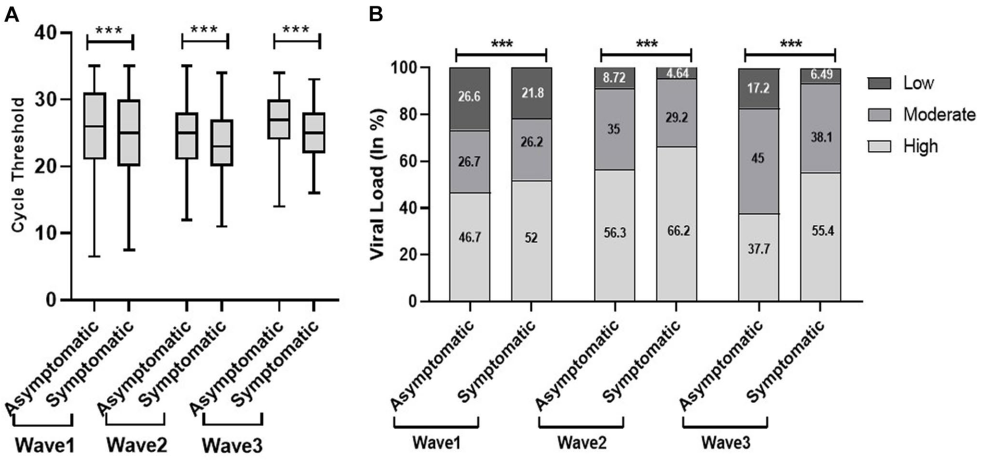


FIGURE 5
Comparison of cycle threshold (Ct) values **(A)** and viral load **(B)** among patients' clinical status during the three COVID waves: midlines indicate the median, boxes indicate interquartile ranges, and whiskers indicate the upper and lower adjacent values (within 1.5 times the interquartile range).

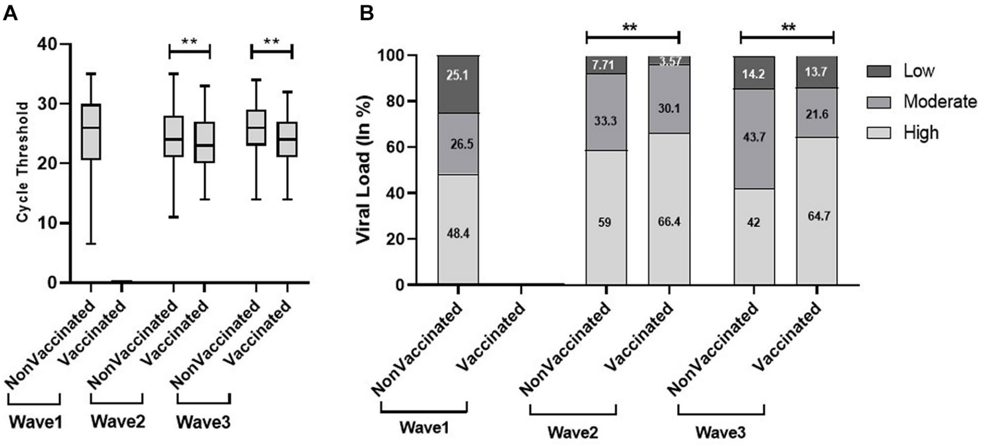


FIGURE 6
Comparison of cycle threshold (Ct) values **(A)** and viral load **(B)** in vaccinated and non-vaccinated patients across the three COVID waves: midlines indicate the median, boxes indicate interquartile ranges, and whiskers indicate the upper and lower adjacent values (within 1.5 times the interquartile range).

response that was initially triggered by a vaccine using an ancestral wild-type virus. The recurrent mutations and evolution in newer strains help in viral evasion from neutralizing antibodies in vaccinated individuals. We also found a high viral load in the non-vaccinated population in the second wave, and the reason could be the severity of the Delta variant, whereas the decrease in viral load in the non-vaccinated population in the third wave could be possibly explained by less severity of the Omicron variant as well as the development of herd immunity within the community (Mandal et al., 2021). More scientific studies are required for a better understanding of the role of Ct value in vaccinated patients.

The paucity of data related to the severity of the COVID-19 infection, mainly the specific symptoms, and co-infections, was the major limitation of this study. Though some researchers observed no or minimal role of co-infections in COVID-19 severity, the topic is still debatable and demands further research (Gupta et al., 2023).

5 Conclusion

Our study highlights the trend of viral load across the three waves in India. The increase in the proportion of high viral load during the second wave indicated the severity of the wave; however, the reduction in high viral load proportion suggested the decline in severity of the third wave. With the development of highly effective vaccines and an increase in immunity against the infection, the pandemic had shown a downward trend over a period, thereby allowing countries to return to normal life, which has led the WHO to announce COVID-19 as a no longer public health emergency on 5 May 2023. Despite this, the world is still facing 1.5 lakh new cases and 497 deaths weekly. Although the Ct value can potentially be used to improve diagnosis, management, and control, it is highly denounced in clinical settings for specific patient management due to its dependency on various factors such as kits, instruments, target genes, or biological material. With the widespread availability of real-time PCR and the limited use of genomic surveillance for predicting pandemic surge in resource-constrained and heavily populated countries, including India, the Ct value or viral load could be a suitable indicator for the early diagnosis of COVID-19 as well as population-level monitoring of COVID-19 dynamics and forecasting in subsequent waves, which can help in restricting the transmission of the virus.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Institutional Ethics Committee, AIIMS Rishikesh. The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were acquired from a by-product of routine care or industry. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

SN: Conceptualization, Data curation, Formal analysis, Investigation, Software, Writing – original draft, Writing – review & editing. Diksha: Visualization, Writing – original draft, Writing – review & editing. DeK: Formal analysis, Resources, Supervision, Writing – review & editing. NR: Data curation, Methodology, Writing – review & editing. AN: Data curation, Resources, Writing – original draft. DiK: Methodology, Writing – original draft. SG: Investigation, Supervision, Writing – original draft, Writing – review & editing. YM: Conceptualization, Data curation, Investigation, Project administration, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

We expressed our profound gratitude to the Department of Health Research (DHR) and the Indian Council of Medical Research (ICMR) for financial assistance.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Asrani, P., Eapen, M. S., Hassan, M. I., and Sohal, S. S. (2021). Implications of the second wave of COVID-19 in India. *Lancet Respir. Med.* 9, e93–e94. doi: 10.1016/S2213-2600(21)00312-X

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., et al. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 395, 507–513. doi: 10.1016/S0140-6736(20)30211-7

Choudhary, O. P., Priyanka, S. I., and Rodriguez-Morales, A. J. (2021). Second wave of COVID-19 in India: dissection of the causes and lessons learnt. *Travel Med. Infect. Dis.* 43:102126. doi: 10.1016/J.TMAID.2021.102126

Cui, J., Li, F., and Shi, Z.-L. (2019). Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* 17, 181–192. doi: 10.1038/s41579-018-0118-9

| European Centre for Disease Prevention and Control (2023) Acute respiratory infections in the EU/EEA. Available at:(https://www.ecdc.europa.eu/en).

Gülbudak, H., Karvar, S., Soydan, G., Tezcan Ülger, S., Kandemir, Ö., Tamer, L., et al. (2021). Comparison of real time pcr cycle threshold values in symptomatic and asymptomatic covid-19 patients. *Mikrobiyol. Bul.* 55, 435–444. doi: 10.5578/MB.20219812

Gupta, S., Negi, A., Negi, S., Diksha, D., Kandwal, D., Singh, A., et al. (2023). Incidence of SARS-CoV-2 co-infections during the second wave in sub-Himalayan region, India. *Cureus* 15:e36215. doi: 10.7759/cureus.36215

Hasanoglu, I., Korukluoglu, G., Asilturk, D., Cosgun, Y., Kalem, A. K., Altas, A. B., et al. (2021). Higher viral loads in asymptomatic COVID-19 patients might be the invisible part of the iceberg. *Infection* 49, 117–126. doi: 10.1007/S15010-020-01548-8/FIGURES/8

Indian Journal of Medical Specialities. COVID diagnostics: Do we have sufficient armamentarium for the present and the unforeseen? (2023). Available at:(https://www.ijms.in/downloadpdf.asp?issn=0976-2884;year=2020;volume=11;issue=3;spage=117;epage=123;aulast=Kashyap;type=2).

INSACOG. (2023). Tools for classification. Available at:(https://ibdc.rcb.res.in/insacog/statreportlineagegraph).

Jayadevan, R., Shenoy, R., and Anithadevi, T. (2022). COVID-19 third wave experience in India, a survey of 5971 adults. *MedRxiv* 493, 4921–4930. doi: 10.1101/2022.04.26.22274273

Kociolek, L. K., Muller, W. J., Yee, R., Bard, J. D., Brown, C. A., Revell, P. A., et al. (2021). Comparison of upper respiratory viral load distributions in asymptomatic and symptomatic children diagnosed with SARS-CoV-2 infection in pediatric hospital testing programs. *J. Clin. Microbiol.* 59:e02593-20. doi: 10.1128/JCM.02593-20

Kurzeder, L., Jörres, R. A., Unterweger, T., Essmann, J., Alter, P., Kahnert, K., et al. (2022). A simple risk score for mortality including the PCR Ct value upon admission in patients hospitalized due to COVID-19. *Infection* 50, 1155–1163. doi: 10.1007/s15010-022-01783-1

Lee, S., Kim, T., Lee, E., Lee, C., Kim, H., Rhee, H., et al. (2020). Clinical course and molecular viral shedding among asymptomatic and symptomatic patients with SARS-CoV-2 infection in a community treatment center in the Republic of Korea. *JAMA Intern. Med.* 180, 1447–1452. doi: 10.1001/JAMAINTERNMED.2020.3862

Louie, J. K., Stoltey, J. E., Scott, H. M., Trammell, S., Ememu, E., Samuel, M. C., et al. (2022). Comparison of symptomatic and asymptomatic infections due to severe acute respiratory coronavirus virus 2 (SARS-CoV-2) in San Francisco long-term care facilities. *Infect. Control Hosp. Epidemiol.* 43, 123–124. doi: 10.1017/ICE.2020.1371

Mandal, S., Arinaminpathy, N., Bhargava, B., and Panda, S. (2021). Plausibility of a third wave of COVID-19 in India: a mathematical modeling based analysis. *Indian J. Med. Res.* 153, 522–532. doi: 10.4103/ijmr.ijmr_1627_21

Martínez, M. J., Basile, L., Sisó-Almirall, A., Cristino, V., Cuesta, G., Hurtado, J. C., et al. (2022). Lack of prognostic value of SARS-CoV2 RT-PCR cycle threshold in the community. *Infect. Dis. Ther.* 11, 587–593. doi: 10.1007/S40121-021-00561-0

Mishra, B., Ranjan, J., Purushotham, P., Kar, P., Payal, P., Saha, S., et al. (2022). Comparison of cycle threshold and clinical status among different age groups of COVID-19 cases. *Cureus* 14:e24194. doi: 10.7759/cureus.24194

Qian, J., Zhao, L., Ye, R. Z., Li, X. J., and Liu, Y. L. (2020). Age-dependent gender differences of COVID-19 in mainland China: comparative study. *Clin. Infect. Dis.* 71, 2488–2494. doi: 10.1093/CID/CIAA683

Rao, S. N., Manissero, D., Steele, V. R., and Pareja, J. (2020). A narrative systematic review of the clinical utility of cycle threshold values in the context of COVID-19. *Infect. Dis. Ther.* 9, 573–586. doi: 10.1007/S40121-020-00324-3/TABLES/3

Reina, J., Morales, C., Busquets, M., and Norte, C. (2018). Usefulness of Ct value in acute respiratory infections caused by respiratory syncytial virus A and B and influenza virus a (H1N1) pdm09, a (H3N2) and B. *Enferm. Infecc. Microbiol. Clin.* 36, 332–335. doi: 10.1016/j.eimc.2017.04.008

Seok, H., Lim, S., Kim, J. Y., Park, C. H., Kim, J. H., Woo, M. L., et al. (2019). Infectivity of coronavirus disease 2019: a prospective cohort study in the Korean metropolitan area. *J. Korean Med. Sci.* 37:37. doi: 10.3346/JKMS.2022.37.E106

Statsenko, Y., Al Zahmi, F., Habuza, T., Almansoori, T. M., Smetanina, D., Simiyu, G. L., et al. (2022). Impact of age and sex on COVID-19 severity assessed from radiologic and clinical findings. *Front. Cell. Infect. Microbiol.* 11:777070. doi: 10.3389/FCIMB.2021.777070/FULL

Strutner, J., Ramchandar, N., Dubey, S., Gamboa, M., Vanderpool, M. K., Mueller, T., et al. (2021). Comparison of reverse-transcription polymerase chain reaction cycle threshold values from respiratory specimens in symptomatic and asymptomatic children with severe acute respiratory syndrome coronavirus 2 infection. *Clin. Infect. Dis.* 73, 1790–1794. doi: 10.1093/CID/CIAB120

Sun, P., Lu, X., Xu, C., Sun, W., and Pan, B. (2020). Understanding of COVID-19 based on current evidence. *J. Med. Virol.* 92, 548–551. doi: 10.1002/JMV.25722

Tendulkar, P., Pandey, P., Panda, P. K., Bhadoria, A. S., Kulshreshtha, P., Mishra, M., et al. (2023). Comparative study between the first and second wave of COVID-19 deaths in India: a single center study. *Cureus* 15:e37472. doi: 10.7759/cureus.37472

To, K. K., Tsang, O. T. Y., Leung, W. S., Tam, A. R., Wu, T. C., Lung, D. C., et al. (2020). Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. *Lancet Infect. Dis.* 20, 565–574. doi: 10.1016/S1473-3099(20)30196-1

Tom, M. R., and Mina, M. J. (2020). To interpret the SARS-CoV-2 test, consider the cycle threshold value. *Clin. Infect. Dis.* 71, 2252–2254. doi: 10.1093/CID/CIAA619

WHO Coronavirus (COVID-19). (2023). Dashboard With Vaccination Data. Available at:(https://covid19.who.int/).

WHO Coronavirus Disease. (2023). Dashboard with vaccination data. Available at:(https://covid19.who.int/region/searo/country/in).

Wright, J., Achana, F., Diwakar, L., Semple, M. G., Carroll, W. D., Baillie, K., et al. (2021). Cycle threshold values are inversely associated with poorer outcomes in hospitalized patients with COVID-19: a prospective, observational cohort study conducted at a UK tertiary hospital. *Int. J. Infect. Dis.* 111, 333–335. doi: 10.1016/J.IJID.2021.08.022

Zirpe, K. G., Dixit, S., Kulkarni, A., Pandit, R. A., Ranganathan, P., Prasad, S., et al. (2021). The second-vs first-wave COVID-19: more of the same or a lot worse? A comparison of mortality between the two waves in patients admitted to intensive care units in nine hospitals in Western Maharashtra. *Indian J. Crit. Care Med.* 25, 1343–1348. doi: 10.5005/JP-JOURNALS-10071-24042

Frontiers in Microbiology

TYPE Original Research
PUBLISHED 20 March 2024
DOI 10.3389/fmicb.2024.1303915

OPEN ACCESS

EDITED BY
Shailendra Saxena,
King George's Medical University, India

REVIEWED BY
Qigai He,
Huazhong Agricultural University, China
Baochao Fan,
Jiangsu Academy of Agricultural Sciences (JAAS), China

*CORRESPONDENCE
Ligen Xiong
✉ 250759970@qq.com
Huayuan Ji
✉ jhy478731@163.com

RECEIVED 28 September 2023
ACCEPTED 04 March 2024
PUBLISHED 20 March 2024

CITATION
Zhang F, Luo Y, Lin C, Tan M, Wan P, Xie B, Xiong L and Ji H (2024) Epidemiological monitoring and genetic variation analysis of pathogens associated with porcine viral diarrhea in southern China from 2021 to 2023.
Front. Microbiol. 15:1303915.
doi: 10.3389/fmicb.2024.1303915

Epidemiological monitoring and genetic variation analysis of pathogens associated with porcine viral diarrhea in southern China from 2021 to 2023

Fanfan Zhang[1], Yangyang Luo[2], Cui Lin[1], Meifang Tan[1], Peiwei Wan[3], Baobao Xie[4], Ligen Xiong[1]* and Huayuan Ji[1]*

[1]Institute of Animal Husbandry and Veterinary Medicine, Jiangxi Academy of Agricultural Sciences, Nanchang, Jiangxi, China, [2]Wen's Foodstuff Group Co., Ltd., Wen's Group Research Institute, Yunfu, Guangdong, China, [3]Jiangxi Biological Vocational College, Nanchang, Jiangxi, China, [4]Dabeinong Technology Co., Ltd. of Jiangxi, Nanchang, Jiangxi, China

Large-scale outbreaks of virus-associated severe diarrhea have occurred in pig populations since 2010. To investigate the prevalence and genetic evolution of the diarrhea-associated viruses responsible for the outbreaks, we tested 1,791 diarrhea samples collected from 213 pig farms in five provinces in southern China between 2021 and 2023. The test results showed that porcine epidemic diarrhea virus (PEDV) was the most frequently detected virus. The prevalence rates ranged from 47.40 to 52.22% in samples and 76.06% (162/213) in pig farms. Porcine rotavirus (PoRV) was the second common virus, with prevalence rates ranging from 25.81 to 50.81% in samples and 72.77%(155/213) in pig farms. Porcine delta coronavirus (PDCoV) was the third common virus, with prevalence rates ranging from 16.33 to 17.48% in samples and 38.50% (82/213) in pig farms. The detection rates of both transmissible gastroenteritis virus (TGEV) and porcine acute diarrheal syndrome coronavirus (SADS-CoV) were very low, less than 1.01% in samples and less than 3.76% in pig farms. In this study, we found SADS-CoV only in piglet diarrhea samples from Jiangxi, Guangdong, and Guangxi provinces in China, with a prevalence rate of 5.16% (11/213) in pig farms. Co-infection with these diarrhea-associated viruses is a common occurrence. The most common co-infections were PEDV and PoRV, with a prevalence rate of 6.64% (119/1,791), followed by PDCoV and PoRV, with a prevalence rate of 4.19% (75/1,791). Phylogenetic analyses showed that PEDV and PEDV variants prevalent in southern China during the past three years clustered into genotype GIIb and recombinant PEDV subtypes. Among the currently endemic PEDV, the most common mutations occurred in the collagenase equivalent (COE) and epitope regions of the spike gene. PoRV strains were mainly dominated by the G9 subtype, followed by the G5, G3 and G4 subtypes. Our results suggest that variant PEDV, PDCoV and PoRV are the main pathogens of swine diarrhea, and singular- or co-infection with pathogenic enteric CoV is common in pig herds in southern China. Therefore, prevention and control of porcine viral diarrhea should be given high attention.

KEYWORDS

porcine diarrhea, prevalence, porcine epidemic diarrhea virus, porcine deltacoronavirus, swine acute diarrhea syndrome coronavirus, porcine rotavirus

1 Introduction

Viral diarrhea is an important disease that seriously damages the pig industry, with porcine epidemic diarrhea virus (PEDV), porcine delta coronavirus (PDCoV), porcine transmissible gastroenteritis virus (TGEV), and porcine rotavirus (PoRV) being the four main pathogens (Liu and Wang, 2021; Zhang et al., 2022). PEDV can infect pigs of all ages, causing watery diarrhea and vomiting accompanied by anorexia and depression. Morbidity approaches 100% in piglets, but can vary in sows. The virus was first reported by British researchers in 1971, and the disease was first reported in China in 1980, and is currently circulating in all major pig producing countries in the world (Jang et al., 2023). PDCoV is a newly discovered porcine enteropathogenic coronavirus in 2014, which can cause vomiting, diarrhea, dehydration and even death in pigs of all ages, especially piglets, with clinical symptoms similar to those of PED (Zhang F. et al., 2017). TGEV was first isolated in the United States in 1946 and can cause acute gastrointestinal infections in pigs of all ages. Clinical symptoms are extremely similar to PDCoV, with mortality rates of up to 100 per cent in piglets and very few deaths in adult pigs (Zhang et al., 2019). Porcine rotavirus disease is an acute intestinal infectious disease caused by PoRV. In 1974, Woode and Bridge first isolated rotavirus from pigs in the UK and the clinical signs are similar to those of PEDV. Pigs of different ages can be infected, but clinically the most obvious symptoms are seen in piglets within 2 months of age and the mortality rate is as high as 95%. With increasing age, resistance gradually increases and medium and large pigs have strong resistance and clinically it is mostly a latent infection process and infection of pregnant pigs can be transmitted to piglets through the placenta (Ryu et al., 2021). Epidemics caused by PDCoV, PEDV, TGEV, and PoRV not only have similar clinical symptoms, but also often have co-infections and secondary infections that increase herd morbidity and mortality, making these pathogens important pathogens that are currently harming the pig industry (Hou et al., 2023). SADS-CoV was first detected in Guangdong in February 2017, causing a relatively large outbreak of piglet diarrhea with clinical symptoms similar to those of PED, and is a novel porcine coronavirus discovered before the emergence of the COVID-19 outbreak (Zhou et al., 2018). To date, SADS-CoV has only been a regional sporadic infection, and has only been found in a few pig farms in Guangdong, Fujian and Guangxi provinces. However, SADS-CoV has the potential to be broadly species-specific, especially infecting a wide range of human respiratory and intestinal progenitor cells, which should be highly regarded as a potential public health hazard, and the possibility of cross-species transmission of SADS-CoV to other species or to human beings in the future. Strengthening the epidemiological investigation of the above five pathogens is an important basis for good prevention and control of viral diarrhea in pigs.

In this study, we investigated five major porcine diarrhea-associated viruses including PEDV, PDCoV, TGEV, PoRV, and SADS-CoV in five provinces (Guangdong, Guangxi, Jiangxi, Fujian and Hunan) from porcine diarrhea samples collected from 2021 to 2023. Sequencing analysis of the PEDV S1 gene and PoRV VP7 was performed to elucidate the genetic characteristics of PEDV and PoRV in the southern provinces of China.

2 Materials and methods

2.1 Animal ethics statement

All samples were collected on commercial pig farms by pig veterinarians during routine diagnostic sampling after permission from the farm owner. No specific permits from an animal ethics committee were required.

2.2 Sample collection and viral RNA/DNA extraction

From 2021 to 2023, a total of 1,791 diarrheal samples from 213 pig farms in five provinces (Guangdong, Guangxi, Fujian, Jiangxi and Hunan) in southern China were submitted to the Key Laboratory for Animal Health at the Department of Preventive Veterinary Medicine in Jiangxi Agricultural University and Research Institute of Guangdong Wen's Food Group Co., Ltd. for diagnosis. These samples included small intestinal contents and tissues ($N=774$), feces ($N=1{,}017$) from domestic pigs (*Sus scrofa*) of different age groups with diarrhea: 223 samples from sows, 1,192 samples from suckling piglets, 219 samples from nursery pigs and 157 samples from finishing/adult pigs (Table 1). The small intestinal contents and tissues samples used in this study were obtained from the dead piglets and the fecal samples were non-invasively collected immediately after excretion from diarrheal pigs from premises by veterinarians in these farms and then submitted to laboratory. All samples were recorded and aliquoted into 5 mL eppendorf tubes and stored at −80°C until tested.

2.3 Detection of porcine diarrhea associated viruses

Total viral DNA/RNA was extracted using TaKaRa MiniBEST Viral RNA/DNA Extraction Kit Ver.5.0 (Takara, Dalian, China) following the manufacturer's instruction, and reverse transcribed into first-strand cDNA using HiScript II 1st Strand cDNA Synthesis Kit (Vazyme, Nanjing, China) by random hexamer primers. The steps of reverse transcription included denaturation of the RNA template and synthesis of the 1st strand cDNAs. A total volume of 20 μL of reaction solution was instantly centrifuged and placed in an applied biosystems PCR cycler (Applied Biosystems, Waltham, MA USA). The RNA template is converted to complementary (c)DNA and used as a template for triplex real-time PCR or commercial singleplex real-time PCR. The previously established PCR protocols were used to test four major diarrhea-associated viruses, PEDV, PDCoV, TGEV, and PoRV. The newly emerged SADS-CoV was tested by a method previously established in our laboratory (primer sequence information is presented in the Supplementary Table S2) (Zhang et al., 2019). Data were analyzed based on the discriminations of year, pig growing stage, and sampling area.

2.4 Analysis of PEDV S1 gene and PoRV VP7 gene

To elucidate the molecular characteristics of the S1 gene of PEDV and the VP7 gene of PoRV, representative PEDV- and PoRV-positive

TABLE 1 Categorization of detection results on porcine diarrhea associated viruses of samples collected from 2021 to 2023.

Classifications	Sample No.	Viruses [Positive rate % (Number)]				
		PEDV	PRoV	PDCoV	TGEV	SADS-CoV
Year						
2021	496	52.22 (259)	25.81 (128)	16.33 (81)	1.01 (5)	2.62 (13)
2022	738	49.73 (367)	34.01 (251)	17.48 (129)	0.95 (7)	3.66 (27)
2023	557	47.40 (264)	50.81 (283)	17.41 (97)	0.54 (3)	0.00 (0)
Total	1,791	49.69 (890)	36.96 (662)	17.14 (307)	0.84 (15)	2.23 (40)
Province						
Guangdong	793	47.54 (377)	36.44 (289)	15.89 (126)	0.88 (7)	1.01 (8)
Guangxi	501	50.70 (254)	3,713 (186)	18.16 (91)	1.20 (6)	4.19 (21)
Jiangxi	372	53.49 (199)	38.98 (145)	19.35 (72)	0.54 (2)	2.96 (11)
Fujian	82	47.56 (39)	30.49 (25)	17.07 (14)	0.00 (0)	0.00 (0)
Hunan	43	48.84 (21)	39.53 (17)	9.30 (4)	0.00 (0)	0.00 (0)
Sample type						
Intestine	774	58.79 (455)	36.30 (281)	25.58 (198)	0.00 (0)	3.75 (29)
Feces	1,017	42.77 (435)	37.46 (381)	10.72 (109)	1.47 (15)	1.08 (11)
Growing stage						
Sow	223	46.19 (103)	31.39 (70)	12.56 (28)	2.24 (5)	1.79 (4)
Suckling piglet	1,192	54.95 (655)	41.28 (492)	21.22 (253)	0.67 (8)	2.68 (32)
Nursery pig	219	38.81 (85)	29.22 (64)	6.85 (15)	0.46 (1)	0.91 (2)
Finishing pig	157	29.94 (47)	21.66 (34)	7.01 (11)	0.64 (1)	1.27 (2)

PEDV, Porcine epidemic diarrhea virus; PDCoV, Porcine deltacoronavirus; TGEV, Transmissible gastroenteritis virus; PoRV, Porcine rotavirus; SADS-CoV, Swine acute diarrhea syndrome coronavirus.

samples were amplified, cloned and then sequenced on the basis of previous studies. The sequence fragments of the PCR products were assembled and annotated using SeqMan software (DNASTAR, Madison, USA). Nt and aa sequences of the PEDV S1 and PoRV VP7 genes were aligned by using the Clustal W Method in DNAStar software (Version 7.10). Phylogenetic trees were generated based on the PEDV S1 and PoRV VP7 genes by using the neighbor-joining method of Molecular Evolutionary Genetics Analysis (MEGA version 7.0) with a bootstrap value of 1,000 replicate datasets. The obtained strains and 29 reference strains from the GenBank database were used for this purpose.

2.5 Statistical analysis

Statistical analyses were performed using SPSS software V20.0 (IBM Corporation, Chicago, USA). A p-value of < 0.05 was set as the statistically significant level.

3 Results

3.1 Prevalence of PEDV, PoRV, PDCoV, TGEV, and SADS-CoV

In this study, 1,791 samples from 213 pig farms in five southern provinces of China were detected from 2021 to 2023. The results showed that variant PEDV was the main virus associated with severe diarrhea, with prevalence rates ranging between 47.40 and 52.22%. With good biosecurity measures and widespread use of the PEDV variant vaccine, the PEDV detection rate shows a downward trend year by year during the period 2021–2023 (Table 1). In addition to PEDV, PoRV, PDCoV, TGEV, and SADS-CoV have also been observed among some diarrhea samples tested. Interestingly, the positive rate of PoRV has been increasing in recent years, surpassing PDCoV as the second pathogen causing piglet diarrhea, with a positive rate up to 50.81%. In 2023, the detection rate of PoRV has surpassed that of PEDV, and it has become the most important killer of piglet diarrhea. SADS-CoV was found only in a small number of pig farm samples collected in Guangdong, Guangxi and Jiangxi provinces in this study, but the SADS-CoV was not observed in samples collected in Fujian and Hunan provinces. The SADS-CoV positivity rate of diarrheic piglets in Guangdong, Guangxi and Jiangxi provinces was 1.01% (8/793), 4.19% (21/501) and 2.96% (11/372), respectively. The detection rate of TGEV was low (<1.5%).

In the context of the sample source, the small intestines of suckling piglets showed the highest detection rate of PEDV (63.31%), followed by feces (52.72%). Likewise, the detection rates of PoRV from small intestine and feces were 29.66 and 7.92%, respectively (Table 1). As to the growing stage of pigs, PEDV was frequently detectable in pigs of all ages, followed by PoRV. PEDV infection was more common in sows (46.19%) and suckling piglets (54.95%), and similar results were observed for PoRV.

3.2 Co-infections of diarrhea-associated viruses in pigs in southern China

In this study, Mono- and co-infection frequency was analyzed. Among the 1,791 clinical samples, the rates of mono-infection of PEDV, PDCoV, TGEV, SADS-CoV and PoRV in samples tested were

38.86% (696/1,791), 8.77% (157/1,791), 0.22% (4/1,791), 1.45% (26/1,791), and 24.23% (434/1,791), respectively (Table 2). The prevalence rate of co-infections due to two or more diarrhea-associated pathogens ranged from 0.11 to 9.87% in 2021–2023. The most common co-infection was PEDV with PoRV, with a mean positivity rate of 6.64% (119/1,791), with the highest positivity rate of 9.87% (55/557) in 2023.The mean co-infection rates of PEDV with PDCoV, PEDV with TGEV, PEDV with SADS-CoV, PDCoV with PoRV, TGEV with PoRV, and SADS-CoV with PoRV had mean dual infection rates of 2.29, 0.28, 0.22, 4.19, 0.22 and 0.39%, respectively. It should be noted that the co-infection rate of PEDV and PoRV is increasing year by year, with a positive rate of 4.64, 5.56, and 9.87% in 2021–2023, respectively. Meanwhile, triple infections were found in clinical samples, with PEDV, PDCoV, and PoRV infections being the most common, with an average positivity rate of 1.90%. In addition, two cases were found to be co-infected with PEDV, TGEV and PoRV, and three cases were co-infected with PEDV, SADS-CoV, and PoRV.

3.3 Molecular characterization and phylogeny of PEDVs circulating in southern China during 2021 to 2023

To elucidate the genetic characteristics of PEDVs circulating in southern China during 2021 to 2023, the S genes of 9 representative strains of PEDV were sequenced, and analyzed. Phylogenetically, the S regions (aa 1 ~ 794) of the 9 strains of PEDV identified in this study and other 34 selected reference PEDV strains were divided into two genotypes (genotype I: GI and genotype II: GII). All nine PEDV strains identified in this study were classified in GII (subgroup GIIb and Recombinant PEDV) (Figure 3A). Eight epidemic wild strains were located on the GIIb branch, and one belonged to the group GII Recombinant PEDV type, which was distantly related to the vaccine (CV777, Attenuated DR13, JS2008, etc.) strains located in group G1b. The strains in the group GIIb belonged to the epidemic strains, which were more similar to the strains of the original outbreaks in China (AJ1102, FL2013, GD1, etc.). In addition, the eight PEDV strains isolated in this study were distantly related to MEX/104/2013 (Mexico), KNU-1305 (South Korea) and IA1 (USA). JX3 belongs to the Recombinant PEDV type of group GII, and is closer to the LS-23 strain and the CH/HB/ZJK02 strain.

Homology analyses were performed between the nine PEDV strains identified in this study and other 34 selected reference PEDV strains based on the S gene. The results showed that the nucleotide (nt) and amino acid (aa) homologies of all field strains were 98.6–100% and 98.4–99.9%, respectively, and the homologies with the GI strains were 92.9–95.6% and 91.7–95.4%, respectively, and the homologies with the GII strains were 96.9–99.8% and 96.0–99.6% for nt and aa, respectively (Supplementary Table S3). The S protein epitopes of the nine PEDV strains were analyzed, and SS2 and 2C10 were relatively conserved, with the major variants located in the COE region and SS6. In comparison with the classical strain CV777, this isolate had up to 15 amino acid mutations in the S protein and in the COE segment of the antigenic epitope, including H500I/T, C501S, A521S, L525H/Y, S527G, V531I, F540L, T543S/L, G598S, A609E, L616F, F621F, P633H, E637V, and I639V, but no amino acid insertions or deletions were found (Figure 1). Point mutations in SS6 included Y775S (Figure 2).

3.4 Molecular characterization and phylogeny of PoRVs circulating in southern China during 2021 to 2023

VP7 gene sequences of PoRV strains ($N = 13$) were obtained from representative PoRV-positive samples. To analyze the molecular characteristics and phylogenetic relationships among different PoRV

TABLE 2 Mono- and co-infections of PEDV, PDCoV, TGEV, PoRV, and SADS-CoV in samples from southern China during 2021 to 2023.

Type of infection	2021		2022		2023		Total	
	Positive	Positive rate, %	Positive	Positive rate,%	Positive	Positive rate (%)	Positive	Positive rate, %
PEDV only	217	43.75	295	39.97	184	33.03	696	38.86
PDCoV only	43	8.67	68	9.21	46	8.26	157	8.77
TGEV only	2	0.40	1	0.14	1	0.18	4	0.22
SADS-CoV only	9	1.81	17	2.30	0	0.00	26	1.45
PRoV only	79	15.93	166	22.49	189	33.93	434	24.23
PEDV+PDCoV	12	2.42	17	2.30	12	2.15	41	2.29
PEDV+TGEV	1	0.20	2	0.27	2	0.36	5	0.28
PEDV+SADS-CoV	1	0.20	3	0.41	0	0.00	4	0.22
PEDV+PRoV	23	4.64	41	5.56	55	9.87	119	6.64
PDCoV+PRoV	19	3.83	30	4.07	26	4.67	75	4.19
TGEV+PRoV	1	0.20	3	0.41	0	0.00	4	0.22
SADS-CoV+PRoV	2	0.40	5	0.68	0	0.00	7	0.39
PEDV+PDCoV+PRoV	7	1.41	14	1.90	13	2.33	34	1.90
PEDV+TGEV+PRoV	1	0.20	1	0.14	0	0.00	2	0.11
PEDV+SADS-CoV+PRoV	1	0.20	2	0.27	0	0.00	3	0.17

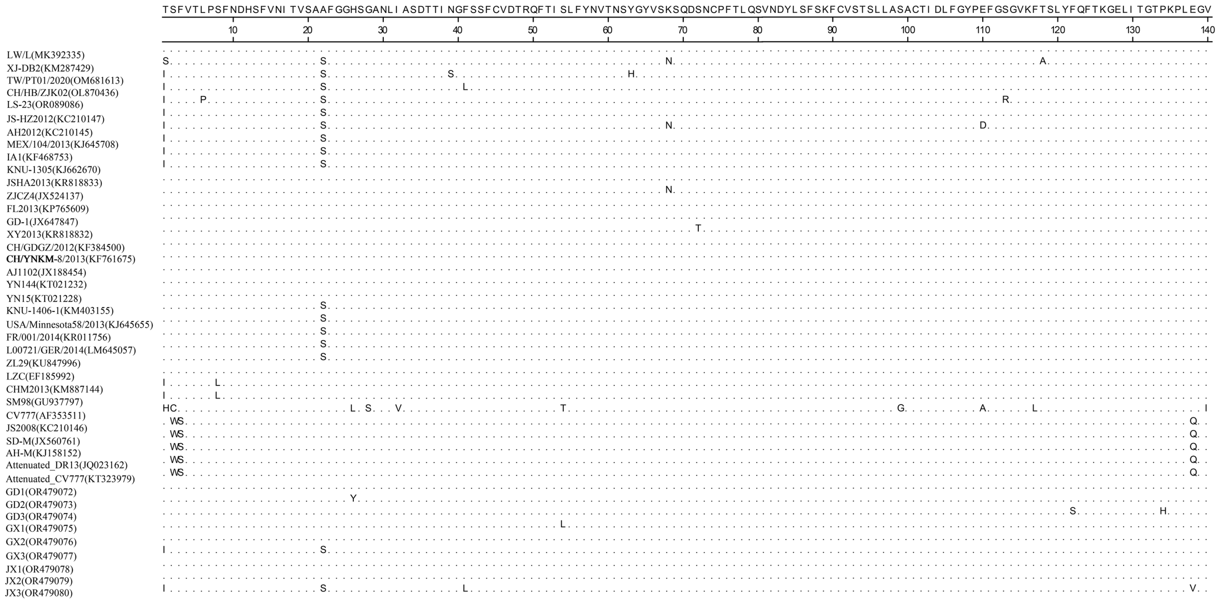


FIGURE 1
Amino acid alignment results based on the S protein sequences of PEDVs. The neutralizing epitopes COE are shown.

isolates, the 13 PoRV strains determined in this study and the 23 PoRV reference strains retrieved from GenBank were used. The nt identities among the 13 PoRV isolates determined in this study ranged from 95.6 to 98.2% and the aa identities ranged from 76.6 to 97.9%; the nt identities with the reference strains from China and the rest of the world ranged from 57.6 to 100% and the aa identities ranged from 50.9 to 100% (Supplementary Table S4). The genetic evolution tree showed that there were four genotypes of the VP7 gene in the 13 sample sequences, namely, one strain of G3 genotype, one strain of G4 subtype, two strains of G5 subtype, and nine strains of G9 genotype. In the evolutionary tree, we can see a separate branch of the G3-subtype sequence for the JX2 strain, which is close to the RVA/Human-wt/THA/MS2014-0134/2014/G3P strain; and GD1, GD2, GD3, GX2, FJ, GX3, JX1, JX3, and HN all belonged to the G9 subtype (Figure 3B).

4 Discussion

Porcine diarrheal disease is a major problem plaguing the pig industry, and among the many causes of porcine diarrhea, porcine viral diarrheal disease has always been the main influencing factor. Common viruses that cause swine diarrhea include PEDV, TGEV, PDCoV, SADS-CoV, and PoRV (Wang et al., 2019; Zhang et al., 2019; Jung et al., 2020). Infection of pigs with these viruses results in diarrhea that affects energy intake and growth and development of individual pigs, and in the case of piglets the fate of the disease is likely to be death. Relevant studies have shown that PEDV dominates the pathogens of porcine viral diarrhea in China (Lee, 2015; Wang et al., 2016, 2019).

In order to understand the main pathogens causing diarrhea in pigs in southern China and the degree of infection, a total of 1,791 clinical diarrhea samples from 213 pig farms in five provinces in southern China were collected from January 2021 to August 2023 for nucleic acid testing of the five pathogens mentioned above, and the results were subjected to data counting and analysis, and the results showed that PEDV infection was the most serious in diarrheic pig herds. The positive detection rate of the samples was 49.67%. These results were consistent with other studies conducted in other parts of China during the same period (Li et al., 2022). Zhang et al. investigated samples from 29 provinces in China between 2011 and 2014 and found that the detection rate of PEDV in diarrheal pigs ranged from 61.10 to 78.49%(Zhang et al., 2019). A study in 2016–2017 that included 116 diarrheal samples from six provinces in China showed that the PEDV prevalence rate was 52.60 per cent (Zhang Q. et al., 2017). Jia et al. detected 672 diarrhea samples collected in Northeast China from 2017 to 2018, and the results showed 19.05% (128/672), 4.32% (29/672), and 3.87% (26/672) positive rates for PEDV, PoRV, and PDCoV, respectively (Jia et al., 2019)。These data suggest that PEDV infections in pigs are common in China, the largest pork-producing country. However, according to the literature, before 2018, the viral diarrhea pathogen in Chinese pigs was mainly dominated by PEDV, and most of the PoRV infections in China were sporadic or local, with no large-scale outbreaks. According to the results of this survey, PEDV infections showed an overall decreasing trend, which may be closely related to the decline in pig inventory and the general strengthening of biosecurity management in pig farms after the occurrence of the African swine fever outbreak in the country in 2018. From 2021 to 2023, the detection rate of PoRV positivity increased each year. In 2023, the detection rate of PoRV will exceed that of PEDV for the first time, making it the most important pathogen causing viral diarrhea in pigs. At present, the prevalent serotypes of porcine rotavirus have changed significantly, evolving from the original G5 type to G9 type, with G9 type accounting for more than 80% of the total, while there are multiple serotypes such as G5 and G4. The serotype of the existing porcine rotavirus vaccine in the triple live vaccine for porcine infectious gastroenteritis, porcine epidemic diarrhea and porcine rotavirus is still G5, and there is almost no

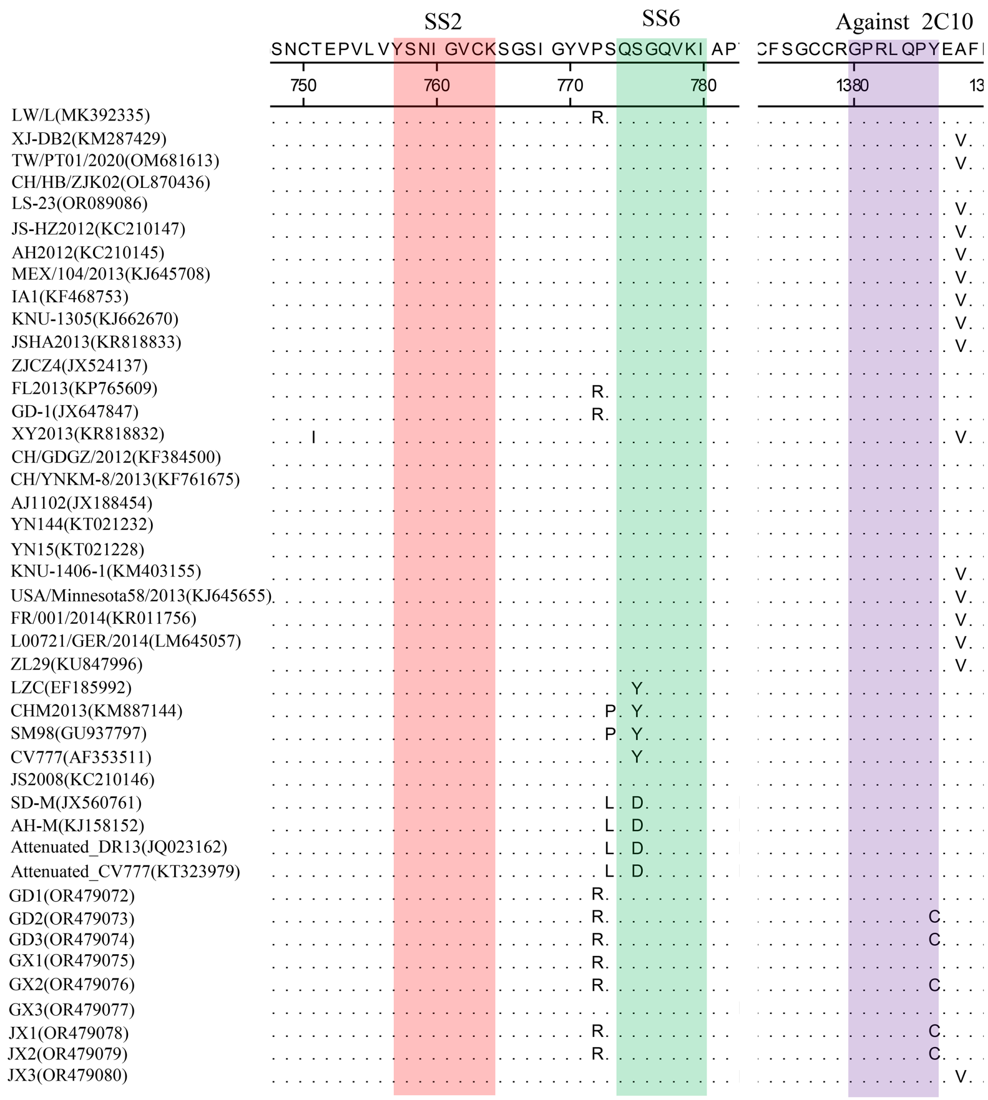


FIGURE 2

Amino acid alignment results based on the S protein sequences of PEDVs. The neutralizing epitopes SS2 (red), SS6 (green) and against 2C10 (purple) are shown.

cross-protection between different serotypes of rotavirus, which is not able to produce an effective preventive effect for the current porcine rotavirus epidemic strains.

PDCoV was first detected in Hong Kong, China, in 2012 and broke out in the United States in 2014 (Song et al., 2015). In our study, we found that PDCoV was the third most common prevalent virus, which is similar to our previous findings. Ding et al. reported that the prevalence rate of PDCoV was as high as 36.18% in faucal samples from nine provinces in China between 2015 and 2017 (Ding et al., 2020). Therefore, PDCoV is also common in pigs in China.

TGEV used to be an important pathogen associated with diarrhea in most pig-producing countries/regions of the world (Li et al., 2022; Sun et al., 2023). However, the clinical cases and impact of TGEV have been limited in recent decades. In this study, we found that TGEV was only sporadically detected during 2021–2023, which is consistent with previous reports (Zhang et al., 2019). SADS-CoV, an emerging coronavirus causing acute diarrhea in suckling piglets, was first detected in Guangdong Province in southern China in 2017 (Zhou et al., 2018). It was subsequently reported in Fujian and Guangxi. In this study, SADS-CoV was detected in some samples from Jiangxi, Guangdong and Guangxi provinces, and the positive rate of the virus was 2.23% (40/1,791). SADS-CoV infections showed an upward trend, and comprehensive prevention and control should be strengthened to avoid the emergence of a large-scale epidemic of SADS-CoV. Sows (46.19%) and lactating piglets (54.95%) were infected with PEDV at a higher rate, as were PDCoV and PoRV infections. These data suggest that the degree of diarrhea may be related to differences in pig resistance, with adult pigs being more likely to be cryptically infected, and lactating piglets having the highest morbidity and mortality rates and being the most damaging. Co-infections with diarrhea-associated coronaviruses are common. Thus, the diagnosis of porcine diarrhea has become increasingly complex, and accurate differential diagnosis can only be made by laboratory tests because diarrhea caused by different enteric pathogens has similar clinical features.

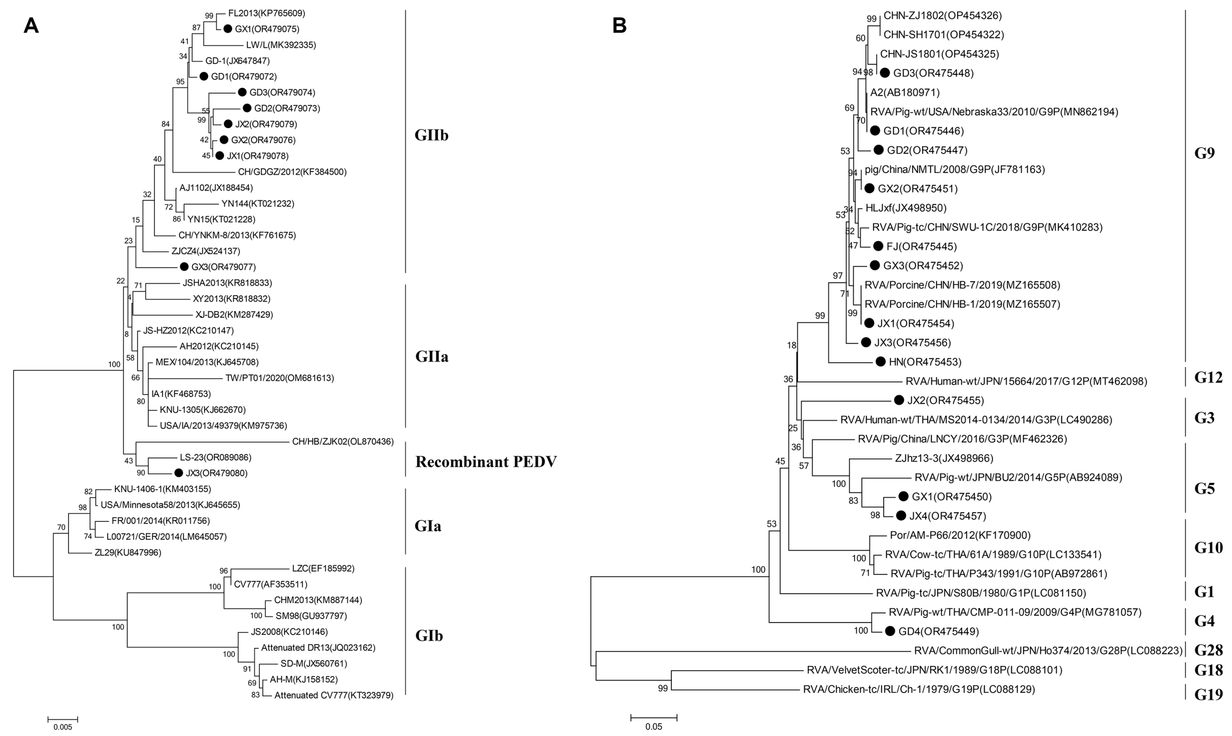


FIGURE 3
Phylogenetic analysis on the animo acid sequences of the S protein (full length sequence) of selected PEDV **(A)** and the VP7 protein (full length sequence) of selected PoRV **(B)** strains from different countries. Solid black circle indicates the strains determined in this study. The tree was constructed using the neighbor-joining method (bootstrap resampling = 1,000 replications) in the MEGA software package, version 7.0.

In this study, we conducted a statistical analysis of the five viruses' single and mixed infections, and found that the viral infections in diarrheal pig herds in southern China were dominated by single infections, mainly PEDV and PoRV single infections, with the infection rates of 38.86 and 24.23%, respectively; among the mixed infections, PDEV+PoRV was the most common, with the positivity rate of 2.29%, which was basically in line with the results of other studies (Wang et al., 2016; Zhang et al., 2019; Jung et al., 2020). In this study, it was found that most of the farms co-infected with PEDV and other pathogens (e.g., PoRV and PDCoV) had a history of epidemics of other pathogens (e.g., PoRV or PDCoV) before the emergence of PEDV. It has been demonstrated that mixed infections of PEDV+TGEV increase PEDV replication and lead to higher disease and death rates of piglets, but mixed infection of PEDV with other pathogens can promote virus replication or increase its pathogenicity needs further study (Sungsuwan et al., 2020). Overall, although PEDV is still the main pathogen causing diarrhea in piglets in different regions of China, the prevalence rate of other enteroviruses varies greatly in different regions, e.g., the detection rate of PEDV and PDCoV is higher than that of PoRV and TGEV in Xinjiang region, and the results of the present study showed that the detection rate of PEDV and PoRV is higher than that of PDCoV and TGEV, which may be due to the differences in geographic regions and pig farms' management. Management differences.

The PEDV S gene is highly variable and is often used as a molecular marker for the analysis of viral genotypes and their variability (Cui et al., 2019). The homology between the S gene sequences of the nine PEDV strains detected in this study ranged from 98.6 to 100%, and the homology of amino acid sequences ranged from 98.4 to 99.9%; the S gene sequences of the nine strains showed the highest homology compared with the corresponding gene sequences of the reference strain of PEDV of type GII, which suggests that all of the above strains belong to the type GII. Further genetic evolutionary analysis based on the S gene sequences showed that eight strains of PEDV belonged to the GII-b subtype, and one strain was the Recombinant PEDV subtype. Most commercially available PEDV vaccines are of GI (e.g., strain CV777) and GII-a subtypes (e.g., strain AJ1102), and a large number of studies have demonstrated that the PEDV CV777 vaccine strain is unable to provide sufficient immunoprotection against the attack of endemic strains of PEDV of GII type. In contrast, vaccines based on GII-a subtype PEDV provide only partial immunological protection against attack by GII-b subtype PEDV, but do not fully guarantee protection against infection or disease (Liu et al., 2019; Zhang et al., 2020). The PEDV S protein is involved in fusion of the virus with host cell receptors and stimulation of the body's production of neutralizing antibodies, etc. Based on the amino acid sequence characteristics of the S protein, it can be divided into the S1 and S2 proteins. The S1 protein contains three neutralizing epitopes: the collagenase equivalence (COE) region (aa499 ~ aa638), epitope SS2 (aa748 ~ aa755), and epitope SS6 (aa764 ~ aa771). Compared with the CV777 strain, the nine PEDV strains detected in this study had partial amino acid mutations in the above and epitope regions of the PEDV S amino acid sequence; furthermore, compared with the Gll-a subtype PEDV vaccine strain (AJ1102) developed in China, part of the PEDV (GD3) S1 protein, aa633, was mutated from P The amino acid of this site is located in the epitope of the COE region of the S1 protein, but whether its mutation (P → H) affects the epitope structure needs to be further explored.

The evolution of PEDV in the wild has rapidly accelerated over the last decade. Since the end of 2010, highly pathogenic strains of PEDV have appeared in China and have subsequently been detected in other countries (Chen et al., 2012; Pan et al., 2012; Sun et al., 2012). Although CV777-based PEDV vaccines have been widely used in China, high morbidity and mortality of neonatal piglets infected with mutated PEDV are still common. Genetic analyses based on circulating strains of PEDV showed that all PEDV can be classified into two genotypes, GI type and GII type: (1) GI type includes classical strains of PEDV, represented by CV777 and strains that appeared before 2010; and (2) GII type includes mutant strains found after 2010. Mutations such as insertions and deletions are frequently observed in the complete genome of variant PEDV, most of which are located in spiking genes, including the antigenic epitope region (Yu et al., 2023). The S-protein of PEDV is an important structural protein of coronaviruses, and is thought to encode antigenic determinants, particularly the viral neutralization epitope. Based on phylogenetic analyses of the S protein, all nine strains of viruses identified in this study belong to the GII variant of PEDV (Kristen-Burmann et al., 2023). Mutations in the neutralization epitopes on the S protein of PEDV have been reported in comparison to conventional PEDV strains, including the CO-26 K-equivalent epitopes COE (aa 499–638), SS2 (aa 748–755), SS6 (aa 764–771) and 2C10 (1368–1,374) (Kim et al., 2016; Lara-Romero et al., 2018; Tian et al., 2021). In this study, we found that the nine identified PEDV strains had multiple mutations in the neutralization epitope regions of SS2 and SS6 compared to CV777, which further validates that the currently prevalent PEDV strains are mutant PEDV, which may account for the poor protection of pigs against CV777-based vaccines.

PoRV is the second most common viral pathogen causing diarrhea in pigs. The 13 PoRV strains in this study were dominated by the G9 genotype, accounting for 61.54% of the total, followed by the G5 genotype, accounting for 15.38% of the total, which was consistent with the results of the previous study (Tao et al., 2023). Papp et al. collected and screened 77 original articles from various regions between 1976 and 2011 to investigate the distribution of porcine rotavirus group A genotypes. They found that G5 (45.8%) was by far the most common genotype among G typeable RVA strains, followed by G3 (11.2%) and G4 (9.6%) (Papp et al., 2013). 98 PoRV-positive samples were detected in 303 piglet diarrhea samples collected from 40 pig farms in 14 districts of Sichuan Province from 2017–2019, of which G9 was the dominant strain, accounting for 41%, while G4, G5, G26 and G3 accounted for 23 per cent, 28.2 per cent, 5.1 per cent and 2.7 per cent, respectively, (Zhou et al., 2021). In recent years, the detection rate of G9-type PoRV has been increasing, and it may be the latest epidemic G genotype PoRV in China. The prevalence of G9-type PoRV will undoubtedly increase the pressure of PoRV prevention and control, so it is necessary to further strengthen the isolation and detection of PoRV prevalent strains, and to screen for vaccine strains matching with the prevalent strains, with a view to providing a technological reserve for the control of PoRV epidemics (Hou et al., 2023).

5 Conclusion

To investigate the prevalence of major diarrhea-associated viruses in pigs, we examined 1,791 clinical samples from five provinces in southern China from 2021 to 2023. The results showed that PEDV was the most frequently detected virus, with prevalence rates ranging from 50.21 to 62.1%, PoRV was the second most prevalent virus detected in porcine diarrhea samples, with localized epidemics and increasing severity in the southern part of China, PDCoV had a tendency to become endemic, and TGEV and the newly emerged SADS-CoV remained sporadic. It is noteworthy that we reported for the first time the emergence of SADS-CoV in Jiangxi Province in southern China. The above pathogens were mainly mono-infections, and the dual mixed infections were dominated by PEDV+PoRV. Phylogenetic analyses showed that PEDV in South China over the past 3 years was mainly dominated by the GIIb variant of PEDV, and the PoRV detected was mainly dominated by G9.

Data availability statement

The data presented in the study are deposited in the NCBI repository, accession number OR479072-OR479080 and OR475445-OR475457.

Ethics statement

The Animal Ethics Committee of the Institute of Animal Husbandry and Veterinary, Jiangxi Academy of Agricultural Science for the studies involving animals because all samples were collected on commercial pig farms by pig veterinarians during routine diagnostic sampling after permission from the farm owner. No specific permits from an animal ethics committee were required. The studies were conducted in accordance with the local legislation and institutional requirements.

Author contributions

FZ: Data curation, Formal analysis, Writing – original draft, Funding acquisition, Methodology. YL: Investigation, Resources, Funding acquisition, Writing – original draft. CL: Data curation, Formal analysis, Methodology, Writing – original draft. MT: Data curation, Investigation, Project administration, Writing – original draft. PW: Writing – original draft. BX: Writing – original draft. LX: Project administration, Writing – review & editing. HJ: Funding acquisition, Methodology, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was funded by the National Natural Science Foundation of China

(32160827), the Natural Science Foundation of Jiangxi Province (20232BAB215054), the Jiangxi Pig Industry Technology System Project (JXARS-01), and the Jiangxi Academy of Agricultural Sciences Basic Research Project (JXSNKYJCRC202307).

Conflict of interest

YL was employed by Wen's Foodstuff Group Co., Ltd., Wen's Group Research Institute. BX was employed by Dabeinong Technology Co., Ltd. of Jiangxi.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2024.1303915/full#supplementary-material

References

Chen, J., Liu, X., Shi, D., Shi, H., Zhang, X., and Feng, L. (2012). Complete genome sequence of a porcine epidemic diarrhea virus variant. *J. Virol.* 86:3408. doi: 10.1128/JVI.07150-11

Cui, J., Li, F., and Shi, Z. L. (2019). Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* 17, 181–192. doi: 10.1038/s41579-018-0118-9

Ding, G., Fu, Y., Li, B., Chen, J., Wang, J., Yin, B., et al. (2020). Development of a multiplex RT-PCR for the detection of major diarrhoeal viruses in pig herds in China. *Transbound. Emerg. Dis.* 67, 678–685. doi: 10.1111/tbed.13385

Hou, W., Fan, M., Zhu, Z., and Li, X. (2023). Establishment and application of a triplex real-time RT-PCR assay for differentiation of PEDV, PoRV, and PDCoV. *Viruses Basel* 15:1238. doi: 10.3390/v15061238

Jang, G., Lee, D., Shin, S., Lim, J., Won, H., Eo, Y., et al. (2023). Porcine epidemic diarrhea virus: an update overview of virus epidemiology, vaccines, and control strategies in South Korea. *J. Vet. Sci.* 24:e58. doi: 10.4142/jvs.23090

Jia, S., Feng, B., Wang, Z., Ma, Y., Gao, X., Jiang, Y., et al. (2019). Dual priming oligonucleotide (DPO)-based real-time RT-PCR assay for accurate differentiation of four major viruses causing porcine viral diarrhea. *Mol. Cell. Probes* 47:101435. doi: 10.1016/j.mcp.2019.101435

Jung, K., Saif, L. J., and Wang, Q. (2020). Porcine epidemic diarrhea virus (PEDv): an update on etiology, transmission, pathogenesis, and prevention and control. *Virus Res.* 286:198045. doi: 10.1016/j.virusres.2020.198045

Kim, Y. K., Cho, Y. Y., An, B. H., Lim, S. I., Lim, J. A., Cho, I. S., et al. (2016). Molecular characterization of the spike and ORF3 genes of porcine epidemic diarrhea virus in the Philippines. *Arch. Virol.* 161, 1323–1328. doi: 10.1007/s00705-016-2758-2

Kristen-Burmann, C., Rogger, P., Veiga, I. B., Riebesehl, S., Rappe, J., Ebert, N., et al. (2023). Reverse genetic assessment of the roles played by the spike protein and ORF3 in porcine epidemic diarrhea virus pathogenicity. *J. Virol.* 97:e196422. doi: 10.1128/jvi.01964-22

Lara-Romero, R., Gomez-Nunez, L., Cerriteno-Sanchez, J. L., Marquez-Valdelamar, L., Mendoza-Elvira, S., Ramirez-Mendoza, H., et al. (2018). Molecular characterization of the spike gene of the porcine epidemic diarrhea virus in Mexico, 2013-2016. *Virus Genes* 54, 215–224. doi: 10.1007/s11262-017-1528-x

Lee, C. (2015). Porcine epidemic diarrhea virus: an emerging and re-emerging epizootic swine virus. *Virol. J.* 12:193. doi: 10.1186/s12985-015-0421-2

Li, C., Lu, H., Geng, C., Yang, K., Liu, W., Liu, Z., et al. (2022). Epidemic and evolutionary characteristics of swine enteric viruses in south-Central China from 2018 to 2021. *Viruses Basel* 14:1420. doi: 10.3390/v14071420

Liu, Q., and Wang, H. Y. (2021). Porcine enteric coronaviruses: an updated overview of the pathogenesis, prevalence, and diagnosis. *Vet. Res. Commun.* 45, 75–86. doi: 10.1007/s11259-021-09808-0

Liu, X., Zhang, L., Zhang, Q., Zhou, P., Fang, Y., Zhao, D., et al. (2019). Evaluation and comparison of immunogenicity and cross-protective efficacy of two inactivated cell culture-derived GIIa- and GIIb-genotype porcine epidemic diarrhea virus vaccines in suckling piglets. *Vet. Microbiol.* 230, 278–282. doi: 10.1016/j.vetmic.2019.02.018

Pan, Y., Tian, X., Li, W., Zhou, Q., Wang, D., Bi, Y., et al. (2012). Isolation and characterization of a variant porcine epidemic diarrhea virus in China. *Virol. J.* 9:195. doi: 10.1186/1743-422X-9-195

Papp, H., Laszlo, B., Jakab, F., Ganesh, B., De Grazia, S., Matthijnssens, J., et al. (2013). Review of group a rotavirus strains reported in swine and cattle. *Vet. Microbiol.* 165, 190–199. doi: 10.1016/j.vetmic.2013.03.020

Ryu, J., Kang, G. J., Kim, O., Park, J. Y., and Shin, H. J. (2021). Transplacental transmission of porcine epidemic diarrhea virus. *Front. Vet. Sci.* 8:802816. doi: 10.3389/fvets.2021.802816

Song, D., Zhou, X., Peng, Q., Chen, Y., Zhang, F., Huang, T., et al. (2015). Newly emerged porcine deltacoronavirus associated with diarrhoea in swine in China: identification, prevalence and full-length genome sequence analysis. *Transbound. Emerg. Dis.* 62, 575–580. doi: 10.1111/tbed.12399

Sun, R. Q., Cai, R. J., Chen, Y. Q., Liang, P. S., Chen, D. K., and Song, C. X. (2012). Outbreak of porcine epidemic diarrhea in suckling piglets, China. *Emerg. Infect. Dis.* 18, 161–163. doi: 10.3201/eid1801.111259

Sun, W., Shi, Z., Wang, P., Zhao, B., Li, J., Wei, X., et al. (2023). Metavirome analysis reveals a high prevalence of porcine hemagglutination encephalomyelitis virus in clinically healthy pigs in China. *Pathogens* 12:510. doi: 10.3390/pathogens12040510

Sungsuwan, S., Jongkaewwattana, A., and Jaru-Ampornpan, P. (2020). Nucleocapsid proteins from other swine enteric coronaviruses differentially modulate PEDV replication. *Virology* 540, 45–56. doi: 10.1016/j.virol.2019.11.007

Tao, R., Chang, X., Zhou, J., Zhu, X., Yang, S., Li, K., et al. (2023). Molecular epidemiological investigation of group a porcine rotavirus in East China. *Front. Vet. Sci.* 10:1138419. doi: 10.3389/fvets.2023.1138419

Tian, Y., Yang, X., Li, H., Ma, B., Guan, R., Yang, J., et al. (2021). Molecular characterization of porcine epidemic diarrhea virus associated with outbreaks in Southwest China during 2014-2018. *Transbound. Emerg. Dis.* 68, 3482–3497. doi: 10.1111/tbed.13953

Wang, D., Fang, L., and Xiao, S. (2016). Porcine epidemic diarrhea in china. *Virus Res.* 226, 7–13. doi: 10.1016/j.virusres.2016.05.026

Wang, Q., Vlasova, A. N., Kenney, S. P., and Saif, L. J. (2019). Emerging and re-emerging coronaviruses in pigs. *Curr. Opin. Virol.* 34, 39–49. doi: 10.1016/j.coviro.2018.12.001

Yu, K., Liu, X., Lu, Y., Long, M., Bai, J., Qin, Q., et al. (2023). Biological characteristics and pathogenicity analysis of a low virulence G2a porcine epidemic diarrhea virus. *Microbiol. Spectr.* 11:e453522. doi: 10.1128/spectrum.04535-22

Zhang, Y., Chen, Y., Yuan, W., Peng, Q., Zhang, F., Ye, Y., et al. (2020). Evaluation of cross-protection between G1a- and G2a-genotype porcine epidemic diarrhea viruses in suckling piglets. *Animals* 10:1674. doi: 10.3390/ani10091674

Zhang, Y., Chen, Y., Zhou, J., Wang, X., Ma, L., Li, J., et al. (2022). Porcine epidemic diarrhea virus: an updated overview of virus epidemiology, virulence variation patterns and virus-host interactions. *Viruses Basel* 14:2434. doi: 10.3390/v14112434

Zhang, Q., Liu, X., Fang, Y., Zhou, P., Wang, Y., and Zhang, Y. (2017). Detection and phylogenetic analyses of spike genes in porcine epidemic diarrhea virus strains circulating in China in 2016-2017. *Virol. J.* 14:194. doi: 10.1186/s12985-017-0860-z

Zhang, F., Luo, S., Gu, J., Li, Z., Li, K., Yuan, W., et al. (2019). Prevalence and phylogenetic analysis of porcine diarrhea associated viruses in southern China from 2012 to 2018. *BMC Vet. Res.* 15:470. doi: 10.1186/s12917-019-2212-2

Zhang, F., Ye, Y., Song, D., Guo, N., Peng, Q., Li, A., et al. (2017). A simple and rapid identification method for newly emerged porcine deltacoronavirus with loop-mediated isothermal amplification. *Biol. Res.* 50:30. doi: 10.1186/s40659-017-0135-6

Zhou, Q., Chen, X., Kan, R., Li, Y., Cao, H., Peng, L., et al. (2021). Molecular epidemiological investigation of porcine group A rotavirus in Sichuan from 2017 to 2019. *Sci. Agric. Sin.* 54, 1063–1072. doi: 10.3864/j.issn.0578-1752.2021.05.017

Zhou, P., Fan, H., Lan, T., Yang, X. L., Shi, W. F., Zhang, W., et al. (2018). Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin. *Nature* 556, 255–258. doi: 10.1038/s41586-018-0010-9

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact